

SAS® for Mixed Models

Second Edition



Ramon C. Littell, Ph.D.
George A. Milliken, Ph.D.
Walter W. Stroup, Ph.D.
Russell D. Wolfinger, Ph.D.
Oliver Schabenberger, Ph.D.



Praise from the Experts

“This is a revision of an already excellent text. The authors take time to explain and provide motivation for the calculations being done. The examples are information rich, and I can see them serving as templates for a wide variety of applications. Each is followed by an interpretation section that is most helpful. Nonlinear and generalized linear mixed models are addressed, as are Bayesian methods, and some helpful suggestions are presented for dealing with convergence problems. Those familiar with the previous release will be excited to learn about the new features in PROC MIXED.

“The MIXED procedure has had a great influence on how statistical analyses are performed. It has allowed us to do correct analyses where we have previously been hampered by computational limitations. It is hard to imagine anyone claiming to be a modern professional data analyst without knowledge of the methods presented in this book. The mixed model pulls into a common framework many analyses of experimental designs and observational studies that have traditionally been treated as being different from each other. By describing the three model components X, Z, and the error term e, one can reproduce and often improve on the analysis of any designed experiment.

“I am looking forward to getting my published copy of the book and am sure it will be well worn in no time.”

David A. Dickey
Professor of Statistics, North Carolina State University

“*SAS for Mixed Models, Second Edition* addresses the large class of statistical models with random and fixed effects. Mixed models occur across most areas of inquiry, including all designed experiments, for example.

“This book should be required reading for all statisticians, and will be extremely useful to scientists involved with data analysis. Most pages contain example output, with the capabilities of mixed models and SAS software clearly explained throughout. I have used the first edition of *SAS for Mixed Models* as a textbook for a second-year graduate-level course in linear models, and it has been well received by students. The second edition provides dramatic enhancement of all topics, including coverage of the new GLIMMIX and NLMIXED procedures, and a chapter devoted to power calculations for mixed models. The chapter of case studies will be interesting reading, as we watch the experts extract information from complex experimental data (including a microarray example).

“I look forward to using this superb compilation as a textbook.”

Arnold Saxton
Department of Animal Science, University of Tennessee

“With an abundance of new material and a thorough updating of material from the first edition, *SAS for Mixed Models, Second Edition* will be of inordinate interest to those of us engaged in the modeling of messy continuous and categorical data. It contains several new chapters, and its printed format makes this a much more readable version than its predecessor. We owe the authors a tip of the hat for providing such an invaluable compendium.”

**Timothy G. Gregoire
J. P. Weyerhaeuser Professor of Forest Management
School of Forestry and Environmental Studies, Yale University**

“Because of the pervasive need to model both fixed and random effects in most efficient experimental designs and observational studies, the *SAS System for Mixed Models* book has been our most frequently used resource for data analysis using statistical software. The second edition wonderfully updates the discussion on topics that were previously considered in the first edition, such as analysis of covariance, randomized block designs, repeated measures designs, split-plot and nested designs, spatial variability, heterogeneous variance models, and random coefficient models. If that isn’t enough, the new edition further enhances the mixed model toolbase of any serious data analyst. For example, it provides very useful and not otherwise generally available tools for diagnostic checks on potentially influential and outlying random and residual effects in mixed model analyses.

“Also, the new edition illustrates how to compute statistical power for many experimental designs, using tools that are not available with most other software, because of this book’s foundation in mixed models. Chapters discussing the relatively new GLIMMIX and NLMIXED procedures for generalized linear mixed model and nonlinear mixed model analyses will prove to be particularly profitable to the user requiring assistance with mixed model inference for cases involving discrete data, nonlinear functions, or multivariate specifications. For example, code based on those two procedures is provided for problems ranging from the analysis of count data in a split-plot design to the joint analysis of survival and repeated measures data; there is also an implementation for the increasingly popular zero-inflated Poisson models with random effects! The new chapter on Bayesian analysis of mixed models is also timely and highly readable for those researchers wishing to explore that increasingly important area of application for their own research.”

**Robert J. Tempelman
Michigan State University**

“We welcome the second edition of this book, given a multitude of scientific and software evolutions in the field of mixed models. Important new developments have been incorporated, including generalized linear mixed models, nonlinear mixed models, power calculations, Bayesian methodology, and extended information on spatial approaches.

“Since mixed models have been developing in a variety of fields (agriculture, medicine, psychology, etc.), notation and terminology encountered in the literature is unavoidably scattered and not as streamlined as one might hope. Faced with these challenges, the authors have chosen to serve the various applied segments. This is why one encounters randomized block designs, random effects models, random coefficients models, and multilevel models, one next to the other.

“Arguably, the book is most useful for readers with a good understanding of mixed models theory, and perhaps familiarity with simple implementations in SAS and/or alternative software tools. Such a reader will encounter a number of generic case studies taken from a variety of application areas and designs. Whereas this does not obviate the need for users to reflect on the peculiarities of their own design and study, the book serves as a useful starting point for their own implementation. In this sense, the book is ideal for readers familiar with the basic models, such as a mixed model for Poisson data, looking for extensions, such as zero-inflated Poisson data.

“Unavoidably, readers will want to deepen their understanding of modeling concepts alongside working on implementations. While the book focuses less on methodology, it does contain an extensive and up-to-date reference list.

“It may appear that for each of the main categories (linear, generalized linear, and nonlinear mixed models) there is one and only one SAS procedure available (MIXED, GLIMMIX, and NLMIXED, respectively), but the reader should be aware that this is a rough rule of thumb only. There are situations where fitting a particular model is easier in a procedure other than the one that seems the obvious choice. For example, when one wants to fit a mixed model to binary data, and one insists on using quadrature methods rather than quasi-likelihood, NLMIXED is the choice.”

Geert Verbeke
Biostatistical Centre, Katholieke Universiteit Leuven, Belgium

Geert Molenberghs
Center for Statistics, Hasselt University, Diepenbeek, Belgium

“Publication of this second edition couldn’t have come at a better time. Since the release of the first edition, a number of advances have been made in the field of mixed models, both computationally and theoretically, and the second edition captures many if not most of these key developments. To that end, the second edition has been substantially reorganized to better explain the general nature and theory of mixed models (e.g., Chapter 1 and Appendix 1) and to better illustrate, within dedicated chapters, the various types of mixed models that readers are most likely to encounter. This edition has been greatly expanded to include chapters on mixed model diagnostics (Chapter 10), power calculations for mixed models (Chapter 12), and Bayesian mixed models (Chapter 13).

“In addition, the authors have done a wonderful job of expanding their coverage of generalized linear mixed models (Chapter 14) and nonlinear mixed models (Chapter 15)—a key feature for those readers who are just getting acquainted with the recently released GLIMMIX and NLMIXED procedures. The inclusion of material related to these two procedures enables readers to apply any number of mixed modeling tools currently available in SAS. Indeed, the strength of this second edition is that it provides readers with a comprehensive overview of mixed model methodology ranging from analytically tractable methods for the traditional linear mixed model to more complex methods required for generalized linear and nonlinear mixed models. More importantly, the authors describe and illustrate the use of a wide variety of mixed modeling tools available in SAS—tools without which the analyst would have little hope of sorting through the complexities of many of today’s technology-driven applications. I highly recommend this book to anyone remotely interested in mixed models, and most especially to those who routinely find themselves fitting data to complex mixed models.”

Edward F. Vonesh, Ph.D.
Senior Baxter Research Scientist
Statistics, Epidemiology and Surveillance
Baxter Healthcare Corporation



SAS[®] for Mixed Models

Second Edition

*Ramon C. Littell, Ph.D.
George A Milliken, Ph.D.
Walter W. Stroup, Ph.D.
Russell D. Wolfinger, Ph.D.
Oliver Schabenberger, Ph.D.*

The correct bibliographic citation for this manual is as follows: Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. 2006. *SAS® for Mixed Models, Second Edition*. Cary, NC: SAS Institute Inc.

SAS® for Mixed Models, Second Edition

Copyright © 2006, SAS Institute Inc., Cary, NC, USA

ISBN-13: 978-1-59047-500-3

ISBN-10: 1-59047-500-3

All rights reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a Web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, February 2006

SAS Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/pubs or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Contents

Preface ix

Chapter 1 Introduction 1

- 1.1 Types of Models That Produce Data 1
- 1.2 Statistical Models 2
- 1.3 Fixed and Random Effects 4
- 1.4 Mixed Models 6
- 1.5 Typical Studies and the Modeling Issues They Raise 7
- 1.6 A Typology for Mixed Models 11
- 1.7 Flowcharts to Select SAS Software to Run Various Mixed Models 13

Chapter 2 Randomized Block Designs 17

- 2.1 Introduction 18
- 2.2 Mixed Model for a Randomized Complete Blocks Design 18
- 2.3 Using PROC MIXED to Analyze RCBD Data 22
- 2.4 Introduction to Theory of Mixed Models 42
- 2.5 Example of an Unbalanced Two-Way Mixed Model: Incomplete Block Design 44
- 2.6 Summary 56

Chapter 3 Random Effects Models 57

- 3.1 Introduction: Descriptions of Random Effects Models 58
- 3.2 Example: One-Way Random Effects Treatment Structure 64
- 3.3 Example: A Simple Conditional Hierarchical Linear Model 75
- 3.4 Example: Three-Level Nested Design Structure 81
- 3.5 Example: A Two-Way Random Effects Treatment Structure to Estimate Heritability 88
- 3.6 Summary 91

Chapter 4 Multi-factor Treatment Designs with Multiple Error Terms 93

- 4.1 Introduction 94
- 4.2 Treatment and Experiment Structure and Associated Models 94
- 4.3 Inference with Mixed Models for Factorial Treatment Designs 102
- 4.4 Example: A Split-Plot Semiconductor Experiment 113

4.5	Comparison with PROC GLM	130
4.6	Example: Type \times Dose Response	135
4.7	Example: Variance Component Estimates Equal to Zero	148
4.8	More on PROC GLM Compared to PROC MIXED: Incomplete Blocks, Missing Data, and Estimability	154
4.9	Summary	156

Chapter 5 Analysis of Repeated Measures Data 159

5.1	Introduction	160
5.2	Example: Mixed Model Analysis of Data from Basic Repeated Measures Design	163
5.3	Modeling Covariance Structure	174
5.4	Example: Unequally Spaced Repeated Measures	198
5.5	Summary	202

Chapter 6 Best Linear Unbiased Prediction 205

6.1	Introduction	206
6.2	Examples of BLUP	206
6.3	Basic Concepts of BLUP	210
6.4	Example: Obtaining BLUPs in a Random Effects Model	212
6.5	Example: Two-Factor Mixed Model	219
6.6	A Multilocation Example	226
6.7	Location-Specific Inference in Multicenter Example	234
6.8	Summary	241

Chapter 7 Analysis of Covariance 243

7.1	Introduction	244
7.2	One-Way Fixed Effects Treatment Structure with Simple Linear Regression Models	245
7.3	Example: One-Way Treatment Structure in a Randomized Complete Block Design Structure—Equal Slopes Model	251
7.4	Example: One-Way Treatment Structure in an Incomplete Block Design Structure—Time to Boil Water	263
7.5	Example: One-Way Treatment Structure in a Balanced Incomplete Block Design Structure	272
7.6	Example: One-Way Treatment Structure in an Unbalanced Incomplete Block Design Structure	281
7.7	Example: Split-Plot Design with the Covariate Measured on the Large-Size Experimental Unit or Whole Plot	286
7.8	Example: Split-Plot Design with the Covariate Measured on the Small-Size Experimental Unit or Subplot	297
7.9	Example: Complex Strip-Plot Design with the Covariate Measured on an Intermediate-Size Experimental Unit	308
7.10	Summary	315

Chapter 8	Random Coefficient Models 317
8.1	Introduction 317
8.2	Example: One-Way Random Effects Treatment Structure in a Completely Randomized Design Structure 320
8.3	Example: Random Student Effects 326
8.4	Example: Repeated Measures Growth Study 330
8.5	Summary 341
Chapter 9	Heterogeneous Variance Models 343
9.1	Introduction 344
9.2	Example: Two-Way Analysis of Variance with Unequal Variances 345
9.3	Example: Simple Linear Regression Model with Unequal Variances 354
9.4	Example: Nested Model with Unequal Variances for a Random Effect 366
9.5	Example: Within-Subject Variability 374
9.6	Example: Combining Between- and Within-Subject Heterogeneity 393
9.7	Example: Log-Linear Variance Models 402
9.8	Summary 411
Chapter 10	Mixed Model Diagnostics 413
10.1	Introduction 413
10.2	From Linear to Linear Mixed Models 415
10.3	The Influence Diagnostics 424
10.4	Example: Unequally Spaced Repeated Measures 426
10.5	Summary 435
Chapter 11	Spatial Variability 437
11.1	Introduction 438
11.2	Description 438
11.3	Spatial Correlation Models 440
11.4	Spatial Variability and Mixed Models 442
11.5	Example: Estimating Spatial Covariance 447
11.6	Using Spatial Covariance for Adjustment: Part 1, Regression 457
11.7	Using Spatial Covariance for Adjustment: Part 2, Analysis of Variance 460
11.8	Example: Spatial Prediction—Kriging 471
11.9	Summary 478
Chapter 12	Power Calculations for Mixed Models 479
12.1	Introduction 479
12.2	Power Analysis of a Pilot Study 480
12.3	Constructing Power Curves 483
12.4	Comparing Spatial Designs 486

12.5 Power via Simulation 489

12.6 Summary 495

Chapter 13 Some Bayesian Approaches to Mixed Models 497

13.1 Introduction and Background 497

13.2 P-Values and Some Alternatives 499

13.3 Bayes Factors and Posterior Probabilities of Null Hypotheses 502

13.4 Example: Teaching Methods 507

13.5 Generating a Sample from the Posterior Distribution with the PRIOR Statement 509

13.6 Example: Beetle Fecundity 511

13.7 Summary 524

Chapter 14 Generalized Linear Mixed Models 525

14.1 Introduction 526

14.2 Two Examples to Illustrate When Generalized Linear Mixed Models Are Needed 527

14.3 Generalized Linear Model Background 529

14.4 From GLMs to GLMMs 538

14.5 Example: Binomial Data in a Multi-center Clinical Trial 542

14.6 Example: Count Data in a Split-Plot Design 557

14.7 Summary 566

Chapter 15 Nonlinear Mixed Models 567

15.1 Introduction 568

15.2 Background on PROC NLMIXED 569

15.3 Example: Logistic Growth Curve Model 571

15.4 Example: Nested Nonlinear Random Effects Models 587

15.5 Example: Zero-Inflated Poisson and Hurdle Poisson Models 589

15.6 Example: Joint Survival and Longitudinal Model 595

15.7 Example: One-Compartment Pharmacokinetic Model 607

15.8 Comparison of PROC NLMIXED and the %NLINMIX Macro 623

15.9 Three General Fitting Methods Available in the %NLINMIX Macro 625

15.10 Troubleshooting Nonlinear Mixed Model Fitting 629

15.11 Summary 634

Chapter 16 Case Studies 637

16.1 Introduction 638

16.2 Response Surface Experiment in a Split-Plot Design 639

16.3 Response Surface Experiment with Repeated Measures 643

16.4	A Split-Plot Experiment with Correlated Whole Plots	650
16.5	A Complex Split Plot: Whole Plot Conducted as an Incomplete Latin Square	659
16.6	A Complex Strip-Split-Split-Plot Example	667
16.7	Unreplicated Split-Plot Design	674
16.8	2³ Treatment Structure in a Split-Plot Design with the Three-Way Interaction as the Whole-Plot Comparison	684
16.9	2³ Treatment Structure in an Incomplete Block Design Structure with Balanced Confounding	694
16.10	Product Acceptability Study with Crossover and Repeated Measures	699
16.11	Random Coefficients Modeling of an AIDS Trial	716
16.12	Microarray Example	727

Appendix 1 Linear Mixed Model Theory 733

A1.1	Introduction	734
A1.2	Matrix Notation	734
A1.3	Formulation of the Mixed Model	735
A1.4	Estimating Parameters, Predicting Random Effects	742
A1.5	Statistical Properties	751
A1.6	Model Selection	752
A1.7	Inference and Test Statistics	754

Appendix 2 Data Sets 757

A2.2	Randomized Block Designs	759
A2.3	Random Effects Models	759
A2.4	Analyzing Multi-level and Split-Plot Designs	761
A2.5	Analysis of Repeated Measures Data	762
A2.6	Best Linear Unbiased Prediction	764
A2.7	Analysis of Covariance	765
A2.8	Random Coefficient Models	768
A2.9	Heterogeneous Variance Models	769
A2.10	Mixed Model Diagnostics	771
A2.11	Spatial Variability	772
A2.13	Some Bayesian Approaches to Mixed Models	773
A2.14	Generalized Linear Mixed Models	774
A2.15	Nonlinear Mixed Models	775
A2.16	Case Studies	776

References 781

Index 795

Preface

The subject of mixed linear models is taught in graduate-level statistics courses and is familiar to most statisticians. During the past 10 years, use of mixed model methodology has expanded to nearly all areas of statistical applications. It is routinely taught and applied even in disciplines outside traditional statistics. Nonetheless, many persons who are engaged in analyzing mixed model data have questions about the appropriate implementation of the methodology. Also, even users who studied the topic 10 years ago may not be aware of the tremendous new capabilities available for applications of mixed models.

Like the first edition, this second edition presents mixed model methodology in a setting that is driven by applications. The scope is both broad and deep. Examples are included from numerous areas of application and range from introductory examples to technically advanced case studies. The book is intended to be useful to as diverse an audience as possible, although persons with some knowledge of analysis of variance and regression analysis will benefit most.

Since the first edition of this book appeared in 1996, mixed model technology and mixed model software have made tremendous leaps forward. Previously, most of the mixed model capabilities in the SAS System hinged on the MIXED procedure. Since the first edition, the capabilities of the MIXED procedure have expanded, and new procedures have been developed to implement mixed model methodology beyond classical linear models. The NLMIXED procedure for nonlinear mixed models was added in SAS 8, and recently the GLIMMIX procedure for generalized linear mixed models was added in SAS 9.1. In addition, ODS and ODS statistical graphics provide powerful tools to request and manage tabular and graphical output from SAS procedures. In response to these important advances we not only brought the SAS code in this edition up-to-date with SAS 9.1, but we also thoroughly re-examined the text and contents of the first edition. We rearranged some topics to provide a more logical flow, and introduced new examples to broaden the scope of application areas.

Note to SAS 8 users: Although the examples in this book were tested using SAS 9.1, you will find that the vast majority of the SAS code applies to SAS 8 as well. Exceptions are ODS statistical graphics, the RESIDUAL and INFLUENCE options in the MODEL statement of PROC MIXED, and the GLIMMIX procedure.

The second edition of *SAS for Mixed Models* will be useful to anyone wishing to use SAS for analysis of mixed model data. It will be a good supplementary text for a statistics course in mixed models, or a course in hierarchical modeling or applied Bayesian statistics. Many mixed model applications have emerged from agricultural research, but the same or similar methodology is useful in other subject areas, such as the pharmaceutical, natural resource, engineering, educational, and social science disciplines. We are of the belief that almost all data sets have features of mixed models, and sometimes are identified by other terminology, such as hierarchical models and latent variables.

Not everyone will want to read the book from cover to cover. Readers who have little or no exposure to mixed models will be interested in the early chapters and can progress through later chapters as their needs require. Readers with good basic skills may want to jump into the chapters on topics of specific interest and refer to earlier material to clarify basic concepts.

The introductory chapter provides important definitions and categorizations and delineates mixed models from other classes of statistical models. Chapters 2–9 cover specific forms of mixed models and the situations in which they arise. Randomized block designs with fixed treatment and random block effects (Chapter 2) are among the simplest mixed models; they allow us to discuss some of the elementary mixed model operations, such as best linear unbiased prediction and expected mean squares, and to demonstrate the use of SAS mixed model procedures in this simple setting. Chapter 3 considers models in which all effects are random. Situations with multiple random components also arise naturally when an experimental design gives rise to multiple error terms, such as in split-plot designs. The analysis of the associated models is discussed in Chapter 4. Repeated measures and longitudinal data give rise to mixed models in which the serial dependency among observations can be modeled directly; this is the topic of Chapter 5. A separate chapter is devoted to statistical inference based on best linear unbiased prediction of random effects (Chapter 6). Models from earlier chapters are revisited here. Chapter 7 deals with the situation where additional continuous covariates have been measured that need to be accommodated in the mixed model framework. This naturally leads us to random coefficient and multi-level linear models (Chapter 8). Mixed model technology and mixed model software find application in situations where the error structure does not comply with that of the standard linear model. A typical example is the correlated error model. Also of great importance to experimenters and analysts are models with independent but heteroscedastic errors. These models are discussed in Chapter 9. Models with correlated errors are standard devices to model spatial data (Chapter 11).

Chapters 10, 12, and 13 are new additions to this book. Diagnostics for mixed models based on residuals and influence analysis are discussed in Chapter 10. Calculating statistical power of tests is the focus of Chapter 12. Mixed modeling from a Bayesian perspective is discussed in Chapter 13.

Chapters 14 and 15 are dedicated to mixed models that exhibit nonlinearity. The first of these chapters deals with generalized linear mixed models where normally distributed random effects appear inside a link function. This chapter relies on the GLIMMIX procedure. Mixed models with general nonlinear conditional mean function are discussed in Chapter 15, which relies primarily on the NLMIXED procedure.

The main text ends with Chapter 16, which provides 12 case studies that cover a wide range of applications, from response surfaces to crossover designs and microarray analysis.

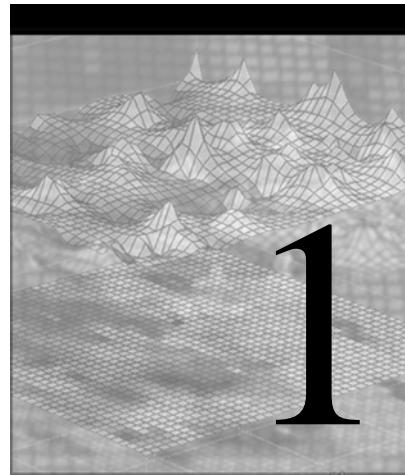
Good statistical applications require a certain amount of theoretical knowledge. The more advanced the application, the more theoretical skills will help. While this book certainly revolves around applications, theoretical developments are presented as well, to describe how mixed model methodology works and when it is useful. Appendix 1 contains some important details about mixed model theory.

Appendix 2 lists the data used for analyses in the book in abbreviated form so you can see the general structure of the data sets. The full data sets are available on the accompanying CD and on the companion Web site for this book (support.sas.com/companionsites). These sources also contain the SAS code to perform the analyses in the book, organized by chapter.

We would like to extend a special thanks to the editorial staff at SAS Press. Our editor, Stephenie Joyner, has shown a precious combination of persistence and patience that kept us on track. Our admiration goes out to our copy editor, Ed Huddleston, for applying his thorough and exacting style to our writing, adding perspicuity.

Writing a book of this scope is difficult and depends on the support, input, and energy of many individuals, groups, and organizations. Foremost, we need to thank our families for their patience, understanding, and support. Thanks to our respective employers—the University of Florida, Kansas State University, the University of Nebraska, and SAS Institute—for giving us degrees of freedom to undertake this project. Thanks to mixed model researchers and statistical colleagues everywhere for adjusting those degrees of freedom by shaping our thinking through their work. Thanks to the statisticians, analysts, and researchers who shared their data sets and data stories and allowed us to pass them along to you. Special thanks go to Andrew Hartley for his considerable and thoughtful commentary on Chapter 13, as well as for many of the references in that chapter. Thanks to the many SAS users who have provided feedback about the first edition. Providing the details of all those who have effectively contributed to this book and by what means would require another whole volume!

As mixed model methodology blazes ahead in the coming decades and continues to provide a wonderful and unifying framework for understanding statistical practice, we trust this volume will be a useful companion as you apply the techniques effectively. We wish you success in becoming a more proficient mixed modeler.



Introduction

1.1	Types of Models That Produce Data	1
1.2	Statistical Models	2
1.3	Fixed and Random Effects.....	4
1.4	Mixed Models.....	6
1.5	Typical Studies and the Modeling Issues They Raise.....	7
1.5.1	Random Effects Model.....	7
1.5.2	Multi-location Example.....	8
1.5.3	Repeated Measures and Split-Plot Experiments	9
1.5.4	Fixed Treatment, Random Block, Non-normal (Binomial) Data Example	9
1.5.5	Repeated Measures with Non-normal (Count) Data	10
1.5.6	Repeated Measures and Split Plots with Effects Modeled by Nonlinear Regression Model.....	10
1.6	A Typology for Mixed Models.....	11
1.7	Flowcharts to Select SAS Software to Run Various Mixed Models.....	13

1.1 Types of Models That Produce Data

Data sets presented in this book come from three types of sources: (1) designed experiments, (2) sample surveys, and (3) observational studies. Virtually all data sets are produced by one of these three sources.

In designed experiments, some form of treatment is applied to experimental units and responses are observed. For example, a researcher might want to compare two or more drug formulations to control high blood pressure. In a human clinical trial, the experimental units are volunteer patients who meet the criteria for participating in the study. The various drug formulations are randomly assigned to patients and their responses are subsequently observed and compared. In sample surveys, data are collected according to a plan, called a survey design, but treatments are

not applied to units. Instead, the units, typically people, already possess certain attributes such as age or occupation. It is often of interest to measure the effect of the attributes on, or their association with, other attributes. In observational studies, data are collected on units that are available, rather than on units chosen according to a plan. An example is a study at a veterinary clinic in which dogs entering the clinic are diagnosed according to their skin condition and blood samples are drawn for measurement of trace elements.

The objectives of a project, the types of resources that are available, and the constraints on what kind of data collection is possible all dictate your choice of whether to run a designed experiment, a sample survey, or an observational study. Even though the three have striking differences in the way they are carried out, they all have common features leading to a common terminology. For example, the terms **factor**, **level**, and **effect** are used alike in design experiments, sample surveys, and observational studies. In designed experiments, the treatment condition under study (e.g., from examples we decide to use) is the *factor* and the specific treatments are the *levels*. In the observational study, the dogs' diagnosis is the factor and the specific skin conditions are the levels. In all three types of studies, each level has an *effect*; that is, applying a different treatment in a designed experiment has an effect on the mean response, or the different skin conditions show differences in their respective mean blood trace amounts. These concepts are defined more precisely in subsequent sections.

In this book, the term **study** refers to whatever type of project is relevant: designed experiment, sample survey, or observational study.

1.2 Statistical Models

Statistical models for data are mathematical descriptions of how the data conceivably can be produced. Models consist of at least two parts: (1) a formula relating the response to all explanatory variables (e.g., effects), and (2) a description of the probability distribution assumed to characterize random variation affecting the observed response.

Consider the experiment with five drugs (say, A, B, C, D, and E) applied to subjects to control blood pressure. Let μ_A denote the mean blood pressure for subjects treated with drug A, and define μ_B , μ_C , μ_D , and μ_E similarly for the other drugs. The simplest model to describe how observations from this experiment were produced for drug A is $Y_A = \mu_A + e$. That is, a blood pressure observation (Y_A) on a given subject treated with drug A is equal to the mean of drug A plus random variation resulting from whatever is particular to a given subject other than drug A. The random variation, denoted by the term e , is called the **error** in Y . It follows that e is a random variable with a mean of zero and a variance of σ^2 . This is the simplest version of a **linear statistical model**—that is, a model where the observation is the sum of terms on the right-hand side of the model that arise from treatment or other explanatory factors plus random error.

The model $Y_A = \mu_A + e$ is called a **means model** because the only term on the right-hand side of the model other than random variation is a treatment mean. Note that the mean is also the expected value of Y_A . The mean can be further modeled in various ways. The first approach leads to an effects model. You can define the effect of drug A as α_A such that $\mu_A = \mu + \alpha_A$, where μ is defined as the intercept. This leads to the one-way **analysis of variance** (ANOVA) model $Y_A = \mu + \alpha_A + e$, the simplest form of an **effects model**. Note that the effects model has more parameters (in this case 6, μ and the α_i) than factor levels (in this case 5). Such models are said to be **over-parameterized** because there are more parameters to estimate than there are unique items of information. Such models require some constraint on the solution to estimate

the parameters. Often, in this kind of model, the constraint involves defining μ as the overall mean implying $\alpha_A = \mu_A - \mu$ and thus

$$\sum_{i=A}^E \alpha_i = 0$$

This is called a sum-to-zero constraint. Its advantage is that if the number of observations per treatment is equal, it is easy to interpret. However, for complex designs with unequal observations per treatment, the sum-to-zero constraint becomes intractable, whereas alternative constraints are more generally applicable. SAS procedures use the constraint that the last factor level, in this case α_E , is set to zero. In general, for effects models, the estimate of the mean $\mu_A = \mu + \alpha_A$ is unique and interpretable, but the individual components μ and the α_i may not be.

Another approach to modeling μ_A , which would be appropriate if levels A through E represented doses, or amounts, of a drug given to patients, is to use linear regression. Specifically, let X_A be the drug dose corresponding to treatment A, X_B be the drug dose corresponding to treatment B, and so forth. Then the regression model, $\mu_A = \beta_0 + \beta_1 X_A$, could be used to describe a linear increase (or decrease) in the mean blood pressure as a function of changing dose. This gives rise to the statistical **linear regression** model $Y_A = \beta_0 + \beta_1 X_A + e$.

Now suppose that each drug (or drug dose) is applied to several subjects, say, n of them for each drug. Also, assume that the subjects are assigned to each drug completely at random. Then the experiment is a **completely randomized design**. The blood pressures are determined for each subject. Then Y_{A1} stands for the blood pressure observed on the first subject treated with drug A. In general, Y_{ij} stands for the observation on the j^{th} subject treated with drug i . Then you can write the model equation $Y_{ij} = \mu + e_{ij}$, where e_{ij} is a random variable with mean zero and variance σ^2 . This means that the blood pressures for different subjects receiving the same treatment are not all the same. The error, e_{ij} , represents this variation. Notice that this model uses the simplifying assumption that the variance of e_{ij} is the same, σ^2 , for each drug. This assumption may or may not be valid in a given situation; more complex models allow for unequal variances among observations within different treatments. Also, note that the model can be elaborated by additional description of μ_i —e.g., as an effects model $\mu_i = \mu + \alpha_i$ or as a regression model $\mu_i = \beta_0 + \beta_1 X_i$. Later in this section, more complicated versions of modeling μ_i are considered.

An alternative way of representing the models above describes them through an assumed probability distribution. For example, the usual linear statistical model for data arising from completely randomized designs assumes that the errors have a normal distribution. Thus, you can write the model $Y_{ij} = \mu_i + e_{ij}$ equivalently as $Y_{ij} \sim N(\mu_i, \sigma^2)$ if the e_{ij} are assumed *iid* $N(0, \sigma^2)$. Similarly, the one-way ANOVA model can be written as $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$ and the linear regression model as $Y_{ij} \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$. This is important because it allows you to move easily to models other than linear statistical models, which are becoming increasingly important in a variety of studies.

One important extension beyond linear statistical models involves cases in which the response variable does not have a normal distribution. For example, suppose in the drug experiment that c_i clinics are assigned at random to each drug, n_{ij} subjects are observed at the j^{th} clinic assigned to drug i , and each subject is classified according to whether a medical event such as a stroke or heart attack has occurred or not. The resulting response variable Y_{ij} can be defined as the number of subjects having the event of interest at the i^{th} clinic, and $Y_{ij} \sim \text{Binomial}(\pi_i, n_{ij})$, where π_i is the probability of a subject showing improvement when treated with drug i . While it

is possible to fit a linear model such as $p_{ij} = \mu_i + e_{ij}$, where $p_{ij} = y_{ij}/n_{ij}$ is the sample proportion and $\mu_i = \pi_i$, a better model might be $\pi_i = 1/(1 + e^{-\mu_i})$ and $\mu_i = \mu + \alpha_i$ or $\mu_i = \beta_0 + \beta_1 X_i$ depending on whether the effects-model or regression framework discussed above is more appropriate. In other contexts, modeling $\pi_i = \Phi(\mu_i)$, where $\mu_i = \mu + \alpha_i$ or $\mu_i = \beta_0 + \beta_1 X_i$, may be preferable, e.g., because interpretation is better connected to subject matter under investigation. The former are simple versions of logistic ANOVA and logistic regression models, and the latter are simple versions of probit ANOVA and regression. Both are important examples of **generalized linear models**.

Generalized linear models use a general function of a linear model to describe the expected value of the observations. The linear model is suggested by the design and the nature of the explanatory variables, similar to the rationale for ANOVA or regression models. The general function (which can be linear or nonlinear) is suggested by the probability distribution of the response variable. Note that the general function can be the linear model itself and the distribution can be normal; thus, “standard” ANOVA and regression models are in fact special cases of generalized linear models. Chapter 14 discusses mixed model forms of generalized linear models.

In addition to generalized linear models, another important extension involves nonlinear statistical models. These occur when the relationship between the expected value of the random variable and the treatment, explanatory, or predictor variables is nonlinear. Generalized linear models are a special case, but they require a linear model embedded within a nonlinear function of the mean. **Nonlinear models** may use any function, and may occur when the response variable has a normal distribution. For example, increasing amounts of fertilizer nitrogen (N) are applied to a crop. The observed yield can be modeled using a normal distribution—that is, $Y_{ij} \sim N(\mu_i, \sigma^2)$. The expected value of Y_{ij} in turn is modeled by $\mu_i = \alpha_i \exp\{-\exp(\beta_i - \gamma X_i)\}$, where X_i is the i^{th} level or amount of fertilizer N, α_i is the asymptote for the i^{th} level of N, γ is the slope, and β_i / γ is the inflection point. This is a Gompertz function that models a nonlinear increase in yield as a function of N: the response is small to low N, then increases rapidly at higher N, then reaches a point of diminishing returns and finally an asymptote at even higher N. Chapter 15 discusses mixed model forms of nonlinear models.

1.3 Fixed and Random Effects

The previous section considered models of the mean involving only an assumed distribution of the response variable and a function of the mean involving only factor effects that are treated as known constants. These are called **fixed effects**. An effect is called fixed if the levels in the study represent all possible levels of the factor, or at least all levels about which inference is to be made. Note that this includes regression models where the observed values of the explanatory variable cover the entire region of interest. In the blood pressure drug experiment, the effects of the drugs are fixed if the five specific drugs are the only candidates for use and if conclusions about the experiment are restricted to those five drugs. You can examine the differences among the drugs to see which are essentially equivalent and which are better or worse than others. In terms of the model $Y_{ij} = \mu + \alpha_i + e_{ij}$, the effects α_A through α_E represent the effects of a particular drug relative to the intercept μ . The parameters $\alpha_A, \alpha_B, \dots, \alpha_E$ represent fixed, unknown quantities.

Data from the study provide estimates about the five drug means and differences among them. For example, the sample mean from drug A, $\bar{y}_{A\Box}$ is an estimate of the population mean μ_A .

Notation note: When data values are summed over a subscript, that subscript is replaced by a period. For example, $y_{A\cdot}$ stands for $y_{A1} + y_{A2} + \dots + y_{An}$. A bar over the summed value denotes the sample average. For example, $\bar{y}_{A\cdot} = n^{-1}y_{A\cdot}$.

The difference between two sample means, such as $\bar{y}_{A\cdot} - \bar{y}_{B\cdot}$, is an estimate of the difference between two population means $\mu_A - \mu_B$. The variance of the estimate $\bar{y}_{A\cdot}$ is $n^{-1}\sigma^2$ and the variance of the estimate $\bar{y}_{A\cdot} - \bar{y}_{B\cdot}$ is $2\sigma^2/n$. In reality, σ^2 is unknown and must be estimated. Denote the sample variance for drug A by s_A^2 , the sample variance for drug B by s_B^2 , and similarly for drugs C, D, and E. Each of these sample variances is an estimate of σ^2 with $n-1$ degrees of freedom. Therefore, the average of the sample variances, $s^2 = (s_A^2 + s_B^2 + \dots + s_E^2)/5$, is also an estimate of σ^2 with $5(n-1)$ degrees of freedom. You can use this estimate to calculate standard errors of the drug sample means, which can in turn be used to make inferences about the drug population means. For example, the standard error of the estimate $\bar{y}_{A\cdot} - \bar{y}_{B\cdot}$ is $\sqrt{2s^2/n}$.

The confidence interval is $(\bar{y}_{A\cdot} - \bar{y}_{B\cdot}) \pm t_\alpha \sqrt{2s^2/n}$, where t_α is the α -level, two-sided critical value of the t -distribution with $5(n-1)$ degrees of freedom.

Factor effects are **random** if they are used in the study to represent only a sample (ideally, a *random sample*) of a larger set of potential levels. The factor effects corresponding to the larger set of levels constitute a population with a probability distribution. The last statement bears repeating because it goes to the heart of a great deal of confusion about the difference between fixed and random effects: *a factor is considered random if its levels plausibly represent a larger population with a probability distribution*. In the blood pressure drug experiment, the drugs would be considered random if there are actually a large number of such drugs and only five were sampled to represent the population for the study. Note that this is different from a regression or response surface design, where doses or amounts are selected deliberately to optimize estimation of fixed regression parameters of the experimental region. Random effects represent true sampling and are assumed to have probability distributions.

Deciding whether a factor is random or fixed is not always easy and can be controversial. Blocking factors and locations illustrate this point. In agricultural experiments blocking often reflects variation in a field, such as on a slope with one block in a strip at the top of the slope, one block on a strip below it, and so forth, to the bottom of the slope. One might argue that there is nothing random about these blocks. However, an additional feature of random effects is **exchangeability**. Are the blocks used in this experiment the only blocks that could have been used, or could any set of blocks from the target population be substituted? Treatment levels are not exchangeable: you cannot estimate the effects of drugs A through E unless you observe drugs A through E. But you could observe them on any valid subset of the target population. Similar arguments can be made with respect to locations. Chapter 2 considers the issue of random versus fixed blocks in greater detail. Chapter 6 considers the multi-location problem.

When the effect is random, we typically assume that the distribution of the random effect has mean zero and variance σ_a^2 , where the subscript a refers to the variance of the treatment effects; if the drugs were random, it would denote the variance among drug effects in the population of drugs. The linear statistical model can be written $Y_{ij} = \mu + a_i + e_{ij}$, where μ represents the mean of all drugs in the population, not just those observed in the study. Note that the drug effect is denoted a_i rather than α_i as in the previous model. A frequently used convention, which this book follows, is to denote fixed effects with Greek letters and random effects with Latin letters. Because the drugs in this study are a sample, the effects a_i are random variables with mean 0 and variance σ_a^2 . The variance of Y_{ij} is $\text{Var}[Y_{ij}] = \text{Var}[\mu + a_i + e_{ij}] = \sigma_a^2 + \sigma^2$.

1.4 Mixed Models

Fixed and random effects were described in the preceding section. A **mixed model** contains both fixed and random effects. Consider the blood pressure drug experiment from the previous sections, but suppose that we are given new information about how the experiment was conducted. The n subjects assigned to each drug treatment were actually identified for the study in carefully matched groups of five. They were matched for criteria such that they would be expected to have similar blood pressure history and response. Within each group of five, drugs were assigned so that each of the drugs A, B, C, D, and E was assigned to exactly one subject. Further assume that the n groups of five matched subjects each was drawn from a larger population of subjects who potentially could have been selected for the experiment. The design is a randomized blocks with fixed treatment effects and random block effects.

The model is $Y_{ij} = \mu + \alpha_i + b_j + e_{ij}$, where μ , α_A , ..., α_E represent unknown fixed parameters—intercept and the five drug treatment effects, respectively—and the b_j and e_{ij} are random variables representing blocks (matched groups of five) and error, respectively. Assume that the random variables b_j and e_{ij} have mean zero and variances σ_b^2 and σ^2 , respectively. The variance of Y_{ij} of the randomly chosen matched set j assigned to drug treatment i is $\text{Var}[Y_{ij}] = \sigma_a^2 + \sigma^2$. The difference between two drug treatment means (say, drugs A and B) within the same matched group is $Y_{Aj} - Y_{Bj}$. It is noteworthy that the difference expressed in terms of the model equation is $Y_{Aj} - Y_{Bj} = \alpha_A - \alpha_B + e_{Aj} - e_{Bj}$, which contains no matched group effect. The term b_j drops out of the equation. Thus, the variance of this difference is $2\sigma^2/n$. The difference between drug treatments can be estimated free from matched group effects. On the other hand, the mean of a single drug treatment, \bar{y}_A , has variance $(\sigma_b^2 + \sigma^2)/n$, which *does* involve the variance among matched groups.

The randomized block design is just the beginning with mixed models. Numerous other experimental and survey designs and observational study protocols produce data for which mixed models are appropriate. Some examples are nested (or hierarchical) designs, split-plot designs, clustered designs, and repeated measures designs. Each of these designs has its own model structure depending on how treatments or explanatory factors are associated with experimental or observational units and how the data are recorded. In nested and split-plot designs there are typically two or more sizes of experimental units. Variances and differences between means must be correctly assessed in order to make valid inferences.

Modeling the variance structure is arguably the most powerful and important single feature of mixed models, and what sets it apart from conventional linear models. This extends beyond variance structure to include correlation among observations. In repeated measures designs, discussed in Chapter 5, measurements taken on the same unit close together in time are often more highly correlated than measurements taken further apart in time. The same principle occurs in two dimensions with spatial data (Chapter 11). Care must be taken to build an appropriate covariance structure into the model. Otherwise, tests of hypotheses, confidence intervals, and possibly even the estimates of treatment means themselves may not be valid. The next section surveys typical mixed model issues that are addressed in this book.

1.5 Typical Studies and the Modeling Issues They Raise

Mixed model issues are best illustrated by way of examples of studies in which they arise. This section previews six examples of studies that call for increasingly complex models.

1.5.1 Random Effects Model

In the first example, 20 packages of ground beef are sampled from a larger population. Three samples are taken at random from within each package. From each sample, two microbial counts are taken. Suppose you can reasonably assume that the log microbial counts follow a normal distribution. Then you can describe the data with the following linear statistical model:

$$Y_{ijk} = \mu + p_i + s(p)_{ij} + e_{ijk}$$

where Y_{ijk} denotes the k^{th} log microbial count for the j^{th} sample of the i^{th} package. Because packages represent a larger population with a plausible probability distribution, you can reasonably assume that package effects, p_i , are random. Similarly, sample within package effects, $s(p)_{ij}$, and count, or error, effects, e_{ijk} , are assumed random. Thus, the p_i , $s(p)_{ij}$, and e_{ijk} effects are all random variables with mean zero and variances σ_p^2 , σ_s^2 , and σ^2 , respectively. This is an example of a **random effects model**. Note that only the overall mean is a fixed effects parameter; all other model effects are random.

The modeling issues are as follows:

1. How should you estimate the variance components σ_p^2 , σ_s^2 , and σ^2 ?
2. How should you estimate the standard error of the estimated overall mean, $\hat{\mu}$?
3. How should you estimate random model effects p_i , or $s(p)_{ij}$ if these are needed?

Mixed model methods primarily use three approaches to variance component estimation: (1) procedures based on expected mean squares from the analysis of variance (ANOVA); (2) maximum likelihood (ML); and (3) restricted maximum likelihood (REML), also known as residual maximum likelihood. Of these, ML is usually discouraged, because the variance component estimates are biased downward, and hence so are the standard errors computed from them. This results in excessively narrow confidence intervals whose coverage rates are below the nominal $1-\alpha$ level, and upwardly biased test statistics whose Type I error rates tend to be well above the nominal α level. The REML procedure is the most versatile, but there are situations for which ANOVA procedures are preferable. PROC MIXED in SAS uses the REML approach by default, but provides optional use of ANOVA and other methods when needed. Chapter 4 presents examples where you would want to use ANOVA rather than REML estimation.

The estimate of the overall mean in the random effects model for packages, samples, and counts is $\hat{\mu} = \bar{y}_{\text{all}} = \sum y_{ijk} / IJK$, where I denotes the number of packages (20), J is the number of samples per package (3), and K is the number of counts per sample (2). Substituting the model equations yields $\sum (\mu + p_i + s(p)_{ij} + e_{ijk}) / IJK$, and taking the variance yields

$$\text{Var}[\hat{\mu}] = \text{Var}\left[\sum (p_i + s(p)_{ij} + e_{ijk})\right] / (IJK)^2 = (JK\sigma_p^2 + K\sigma_s^2 + \sigma^2) / IJK$$

8 SAS for Mixed Models

If you write out the ANOVA table for this model, you can show that you can estimate $\text{Var}[\hat{\mu}]$ by $\text{MS(package)}/(IJK)$. Using this, you can compute the standard error of $\hat{\mu}$ by $\sqrt{\text{MS(package)}/(IJK)}$, and hence the confidence interval for μ becomes

$$\bar{y}_{\text{all}} \pm t_{\alpha, df(\text{package})} \sqrt{\text{MS(package)}/(IJK)}$$

where α is the two-sided critical value from the t distribution and $df(\text{package})$ are the degrees of freedom associated with the package source of variation in the ANOVA table.

If we regard package effects as fixed, you would estimate its effect as $\hat{p}_i = \bar{y}_{i\text{all}} - \bar{y}_{\text{all}}$. However, because the package effects are random variables, the **best linear unbiased predictor (BLUP)**

$$E[p_i | y] = E[p_i] + \text{Cov}[\hat{p}_i, \bar{y}_{i\text{all}}] (\text{Var}[\bar{y}_{i\text{all}}])^{-1} (\bar{y}_{i\text{all}} - \bar{y}_{\text{all}})$$

is more efficient. This leads to the “BLUP”

$$\hat{p}_i = \left(\frac{\sigma_p^2}{(JK\sigma_p^2 + K\sigma_s^2 + \sigma^2)/JK} \right) (\bar{y}_{i\text{all}} - \bar{y}_{\text{all}})$$

When estimates of the variance components are used, the above is not a true BLUP, but an estimated BLUP, often called an **EBLUP**. Best linear unbiased predictors are used extensively in mixed models and are discussed in detail in Chapters 6 and 8.

1.5.2 Multi-location Example

The second example appeared in Output 3.7 of *SAS System for Linear Models, Fourth Edition* (Littell et al. 2002). The example is a designed experiment with three treatments observed at each of eight locations. At the various locations, each treatment is assigned to between three and 12 randomized complete blocks. A possible linear statistical model is

$$Y_{ijk} = \mu + L_i + b(L)_{ij} + \tau_k + (\tau L)_{ik} + e_{ijk}$$

where L_i is the i^{th} location effect, $b(L)_{ij}$ is the ij^{th} block within location effect, τ_k is the k^{th} treatment effect, and $(\tau L)_{ik}$ is the ik^{th} location by treatment interaction effect. The modeling issues are as follows:

1. Should location be a random or fixed effect?
2. Depending on issue 1, the F-test for treatment depends on MS(error) if location effects are fixed or $\text{MS(location} \times \text{treatment)}$ if location effects are random.
3. Also depending on issue 1, the standard error of treatment means and differences are affected.

The primary issue is one of **inference space**—that is, the population to which the inference applies. If location effects are fixed, then inference applies *only to those locations* actually involved in the study. If location effects are random, then inference applies to the *population represented by the observed locations*. Another way to look at this is to consider issues 2 and 3. The expected mean square for error is σ^2 , whereas the expected mean square for location \times treatment is $\sigma^2 + k\sigma_{TL}^2$, where σ_{TL}^2 is the variance of the location \times treatment effects and k is a

constant determined by a somewhat complicated function of the number of blocks at each location. The variance of a treatment mean is $\sigma^2 / (\text{number of observations per treatment})$ if location effects are fixed, but it is $[\sigma^2 + K(\sigma_{TL}^2 + \sigma_L^2)] / (\text{obs/trt})$ if location effects are random. The inference space question, then, depends on what sources you believe contribute to uncertainty. If you believe all uncertainty comes from variation among blocks and experimental units within locations, you believe locations are fixed. If, on the other hand, you believe that variation among locations contributes additional uncertainty, then you believe locations are random. Issues of this sort first appear in Chapter 2, and reappear in various forms throughout the rest of the book (e.g., Chapters 4 and 6).

1.5.3 Repeated Measures and Split-Plot Experiments

Because repeated measures and split-plot experiments share some characteristics, they have some modeling issues in common. Suppose that three drug treatments are randomly assigned to subjects, n_i , to the i^{th} treatment. Each subject is observed at 1, 2, ..., 7, and 8 hours post-treatment. A possible model for this study is

$$Y_{ijk} = \mu + \alpha_i + s(\alpha)_{ij} + \tau_k + (a\tau)_{ik} + e_{ijk}$$

where α represents treatment effects, τ represents time (or hour) effects, and $s(\alpha)$ represent the random subject within treatment effects. The main modeling issues here are as follows:

1. The experimental unit for the treatment effect (subject) and for time and time \times treatment effects (subject \times time) are different sizes, and hence these effects require different error terms for statistical inference. *This is a feature common to split-plot and repeated measures experiments.*
2. The errors, e_{ijk} , are correlated within each subject. How best to model correlation and estimate the relevant variance and covariance parameters? This is usually a question specific to repeated measures experiments.
3. How are the degrees of freedom for confidence intervals and hypothesis tests affected?
4. How are standard errors affected when estimated variance and covariance components are used?

Chapter 4 discusses the various forms of split-plot experiments and appropriate analysis using PROC MIXED. Repeated measures use similar strategies for comparing means. Chapter 5 builds on Chapter 4 by adding material specific to repeated measures data. Chapter 5 discusses procedures for identifying and estimating appropriate covariance matrices. Degree of freedom issues are first discussed in Chapter 2 and appear throughout the book. Repeated measures, and correlated error models in general, present special problems to obtain unbiased standard errors and test statistics. These issues are discussed in detail in Chapter 5. Spatial models are also correlated error models and require similar procedures (Chapter 11).

1.5.4 Fixed Treatment, Random Block, Non-normal (Binomial) Data Example

The fourth example is a clinical trial with two treatments conducted at eight locations. At each location, subjects are assigned at random to treatments; n_{ij} subjects are assigned to treatment i at location j . Subjects are observed to have either favorable or unfavorable reactions to the treatments. For the ij^{th} treatment-location combination, Y_{ij} subjects have favorable reactions, or, in other words, $p_{ij} = Y_{ij}/n_{ij}$ is the proportion of favorable reactions to treatment i at location j .

This study raises the following modeling issues:

1. Clinic effects may be random or fixed, raising inference space questions similar to those just discussed.
2. The response variable is binomial, not normal.
3. Because of issue 2, the response may not be linear in the parameters, and the errors may not be additive, casting doubt on the appropriateness of a linear statistical model.
4. Also as a consequence of issue 2, the errors are a function of the mean, and are therefore not homogeneous.

A possible model for this study is a generalized linear mixed model. Denote the probability of favorable reaction to treatment i at location j by π_{ij} . Then $Y_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij})$. The generalized linear model is

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mu + c_i + \tau_j + (c\tau)_{ij}$$

or alternatively

$$\pi_{ij} = \frac{e^{\mu+c_i+\tau_j+(c\tau)_{ij}}}{1+e^{\mu+c_i+\tau_j+(c\tau)_{ij}}} = \frac{1}{1+e^{-(\mu+c_i+\tau_j+(c\tau)_{ij})}}$$

where c_i are random clinic effects, τ_j are fixed treatment effects, and $(c\tau)_{ij}$ are random clinic \times treatment interaction effects. Generalized linear mixed models are discussed in Chapter 14.

1.5.5 Repeated Measures with Non-normal (Count) Data

The fifth example appears in Output 10.39 of *SAS System for Linear Models, Fourth Edition* (Littell et al. 2002). Two treatments are assigned at random to subjects. Each subject is then observed at four times. In addition, there is a baseline measurement and the subject's age. At each time of measurement, the number of epileptic seizures is counted. The modeling issues here are as follows:

1. Counts are not normally distributed.
2. Repeated measures raise correlated error issues similar to those discussed previously.
3. The model involves both factor effects (treatments) and covariates (regression) in the same model, i.e., analysis of covariance.

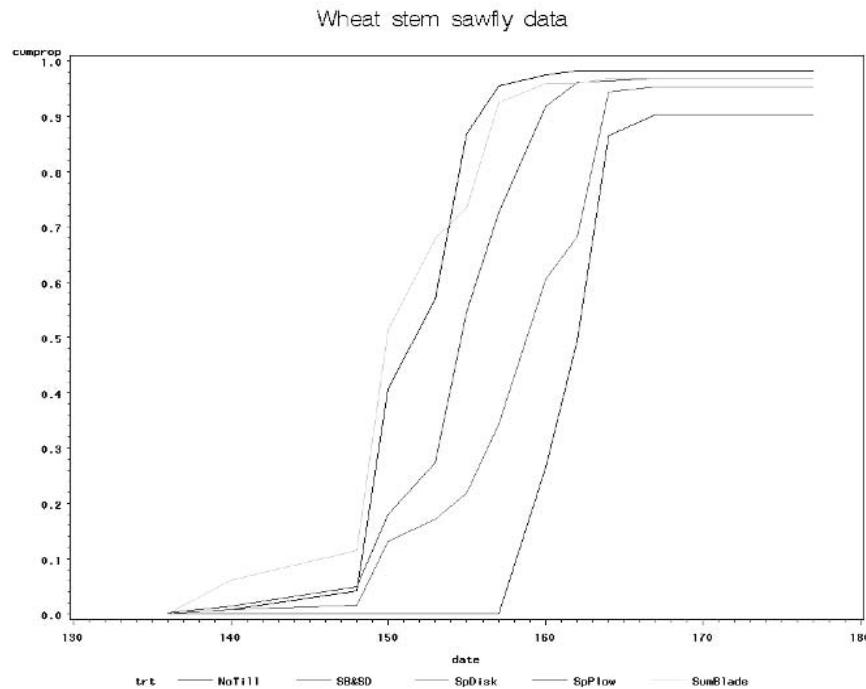
Chapter 7 introduces analysis of covariance in mixed models. Count data in conjunction with repeated measures lead to generalized linear mixed models discussed in Chapter 14.

1.5.6 Repeated Measures and Split Plots with Effects Modeled by Nonlinear Regression Model

The final example involves five treatments observed in a randomized block experiment. Each experimental unit is observed at several times over the growing season and percent emergence is recorded. Figure 1.1 shows a plot of the percent emergence by treatment over the growing season. Like Example 1.5.3, this is a repeated measures experiment, but the structure and model

equation are similar to split-plot experiments, so similar principles apply to mixed model analysis of these data.

Figure 1.1 Treatment Means of Sawfly Data over Time



The modeling issues are as follows:

1. The “usual” mixed model and repeated measures issues discussed in previous examples; plus
2. The obvious nonlinear function required to describe percent emergence as a function of date.

A possible model for this experiment is

$$Y_{ijk} = \mu_{ij} + w_{ij} + e_{ijk}$$

where μ_{ij} is the ij^{th} treatment \times date mean, w_{ij} is the random whole-plot error effect, and e_{ijk} are the repeated measures errors, possibly correlated. The Gompertz model described earlier is a suitable candidate to model μ_{ij} as a function of date j for treatment i . The model described here is an example of a nonlinear mixed model. These are discussed in Chapter 15.

1.6 A Typology for Mixed Models

From the examples in the previous section, you can see that contemporary mixed models cover a very wide range of possibilities. In fact, models that many tend to think of as distinct are, in reality, variations on a unified theme. Indeed, the model that only a generation ago was universally referred to as the “general linear model”—fixed effects only, normal and independent errors, homogeneous variance—is now understood to be one of the more restrictive special cases among commonly used statistical models. This section provides a framework to

view the unifying themes, as well as the distinctive features, of the various modeling options under the general heading of “mixed models” that can be implemented with SAS.

As seen in the previous example, the two main features of a statistical model are (1) a **characterization of the mean**, or expected value of the observations, as a function of model parameters and constants that describe the study design, and (2) a **characterization of the probability distribution** of the observations. The simplest example is a one-factor means model where the expected value of the observations on treatment i is μ_i and the distribution is $N(\mu_i, \sigma^2)$, which leads to the linear statistical model $Y_{ij} = \mu_i + e_{ij}$. The generalized linear mixed model from the fifth example of Section 1.5 provides a more complex example: the mean model is

$$\pi_{ij} = 1 / \left(1 + e^{-(\mu + c_j + (\tau_{ij})_j)} \right)$$

and the distribution has two parts—that of the random effects c_j and $(\tau_{ij})_j$, and that of the observations given the random effects, i.e., $Y_{ij} | c_j, (\tau_{ij})_j \sim \text{Binomial}(n_{ij}, \pi_{ij})$. But each model follows from the same general framework.

Appendix 1 provides a more detailed presentation of mixed model theory. In what follows we present an admittedly simplistic overview that uses matrix notation which is developed more fully at appropriate points throughout the book and in the appendix.

Models have two sets of random variables whose distributions we need to characterize: \mathbf{Y} , the vector of observations, and \mathbf{u} , the vector of random model effects. The models considered in this book assume that the random model effects follow a normal distribution, so that in general we assume $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{G})$ —that is, \mathbf{u} has a multivariate normal distribution with mean zero variance-covariance matrix \mathbf{G} . In a simple variance components model, such as the randomized block model given in Section 1.4, $\mathbf{G} = \sigma_b^2 \mathbf{I}$.

By “mean” of the observations we can refer to one of two concepts: either the **unconditional mean**, $E[\mathbf{Y}]$ or the **conditional mean** of the observations given the random model effects, $E[\mathbf{Y}|\mathbf{u}]$. In a fixed effects model, the distinction does not matter, but for mixed models it clearly does. Mixed models are mathematical descriptions of the **conditional mean** in terms of fixed effect parameters, random model effects, and various constants that describe the study design. The general notation is as follows:

$\boldsymbol{\beta}$ is the vector of fixed effect parameters.

\mathbf{X} is the matrix of constants that describe the structure of the study with respect to the fixed effects. This includes the treatment design, regression explanatory or predictor variables, etc.

\mathbf{Z} is the matrix of constants that describe the study’s structure with regard to random effects. This includes the blocking design, explanatory variables in random coefficient designs (see Chapter 8), etc.

The mixed model introduced in Section 1.4, where observations are normally distributed, models the conditional mean as $E[\mathbf{Y}|\mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, and assumes that the conditional distribution of the observations given the random effects is $\mathbf{Y}|\mathbf{u} \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R})$, where \mathbf{R} is the

variance-covariance matrix of the errors. In simple linear models where errors are independent with homogeneous variances, $\mathbf{R} = \sigma^2 \mathbf{I}$. However, in heterogeneous error models (presented in Chapter 9) and correlated error models such as repeated measures or spatial models, the structure of \mathbf{R} becomes very important.

In the most general mixed model included in SAS, the **nonlinear mixed model** (NLMM), the conditional mean is modeled as a function of \mathbf{X} , \mathbf{Z} , $\boldsymbol{\beta}$, and \mathbf{u} with no restrictions; i.e., $h(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, \mathbf{u})$ models $E[\mathbf{Y}|\mathbf{u}]$. Each successive model is more restrictive. The class of **generalized linear mixed models** (GLMM) has a linear model embedded within a nonlinear function—that is, $g(E[\mathbf{Y}|\mathbf{u}])$ is modeled by $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$. In NLMMs and GLMMs, the observations are not necessarily assumed to be normally distributed. The **linear mixed model** (LMM) does assume normally distributed observations and models the conditional mean directly—that is, you assume $E[\mathbf{Y}|\mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$. Each mixed model has a fixed effects model analog, which means that there are no random model effects and hence \mathbf{Z} and \mathbf{u} no longer appear in the model, and the model now applies to $E[\mathbf{Y}]$. The term “mixed model” is often associated with the LMM—it is the “standard” mixed model that is implemented in PROC MIXED. However, the LMM is a special case. The next section presents a flowchart to associate the various models with appropriate SAS software.

Table 1.1 shows the various models and their features in terms of the model equation used for the conditional mean and the assumed distribution of the observations.

Table 1.1 Summary of Models, Characteristics, and Related Book Chapters

Type of Model	Model of Mean	Distribution	Chapter
NLMM	$h(\mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{u})$	$\mathbf{u}, \mathbf{Y} \mathbf{u}$ general	15
GLMM	$g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$	$\mathbf{Y} \mathbf{u}$ general, \mathbf{u} normal	14
LMM	$\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$	$\mathbf{u}, \mathbf{Y} \mathbf{u}$ normal	2–11
NLM	$h(\mathbf{X}, \boldsymbol{\beta})$	\mathbf{Y} normal	15
GLM	$g^{-1}(\mathbf{X}\boldsymbol{\beta})$	\mathbf{Y} general	12
LM	$\mathbf{X}\boldsymbol{\beta}$	\mathbf{Y} normal	2, 4

1.7 Flowcharts to Select SAS Software to Run Various Mixed Models

SAS offers several procedures (PROCs) designed to implement the various mixed models introduced in the previous sections. PROC MIXED is probably the best known mixed model procedure. It is designed to implement LMMs. SAS has several fixed effects model procedures: PROC GLM implements LMs, PROC NLIN implements NLMs, and PROC GENMOD implements GLMs. There are also several procedures, e.g., LOGISTIC and LIFEREG, that implement special types of GLMs; PROC REG, which implements special types of LMs; and so forth. These special-purpose procedures are not discussed in this book, but they are discussed in detail in other SAS publications as noted throughout this book. Note that PROC GLM was

named before generalized linear models appeared, and was named for “general linear models”; these are now understood not to be general at all, but the most restrictive special case among the models described in Section 1.6, and are now known simply as linear models (LM).

For GLMMs and NLMMs, SAS offers PROC GLIMMIX,¹ PROC NLMIXED, and the %NLINMIX macro. PROC GLIMMIX is the latest addition to the mixed model tools in SAS/STAT. The GLIMMIX procedure fits mixed models with normal random effects where the conditional distribution of the data is a member of the exponential family. Because the normal distribution is also a member of this family, the GLIMMIX procedure can fit LMMs. And because you do not have to specify random effects in the SAS mixed model procedures, PROC MIXED can fit LMs, and PROC GLIMMIX can fit GLMs and LMs. Whereas the GLIMMIX procedure supersedes the %GLIMMIX macro, the %NLINMIX macro continues to have uses distinct and supplementary to the NLMIXED procedure.

Figures 1.2 and 1.3 provide flowcharts to help you select the appropriate model and software for your mixed model project. The basic questions you need to ask are as follows:

- Can you assume a normal distribution for your observations? If the model contains random effects, then this question refers to the conditional distribution of the data, given the random effects.
- Can you assume that the mean or a transformation of the mean is linearly related to the model effects? Note that “linear relation” does not mean the absence of curvature. A quadratic (in X) regression model $\beta_0 + \beta_1 X + \beta_2 X^2$ is a linear model in the β ’s because all the terms in the model are additive. The linear component is termed the linear predictor. Generalized linear (mixed) models imply such linearity on a certain scale (the transformation $g()$). On the other hand, the Gompertz regression equation (see Sections 1.4 and 1.5) is a nonlinear equation.
- Are all effects (except errors) fixed? Or are there random model effects?
- Can you assume the errors are independent? Or, as in repeated measures or spatial data, are errors possibly correlated?
- A corollary to the previous question is, Are the variances among the errors homogeneous? If the answer is no, then the same modeling strategies for correlated errors are also needed for heterogeneous errors.

Once you answer these questions you can follow the flowchart to see what kind of model you have and what SAS procedure is appropriate. Then you can refer to the relevant chapter in this book for more information about the model and procedures.

¹ The GLIMMIX procedure is an add-on in SAS 9.1 to SAS/STAT for the (32-bit) Windows platform. It does not ship with SAS 9.1. You can obtain the GLIMMIX procedure for SAS 9.1 as a download from www.sas.com/statistics. This site also contains the documentation for the GLIMMIX procedure.

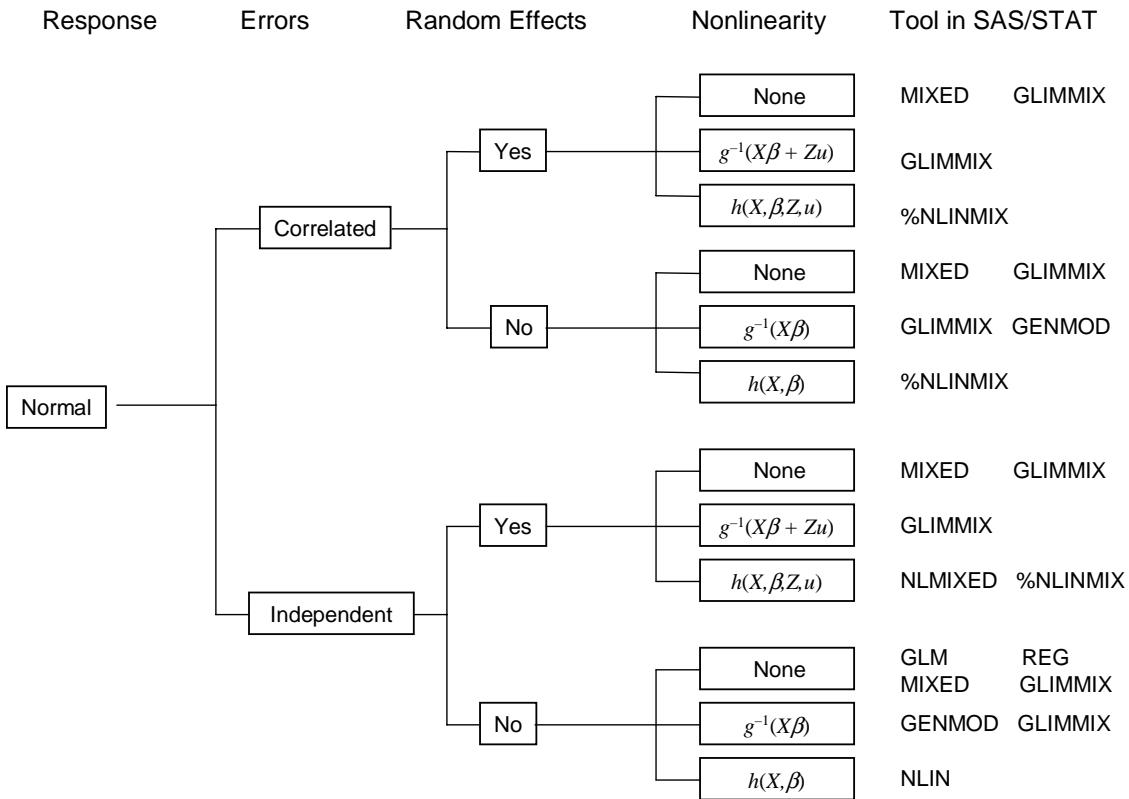
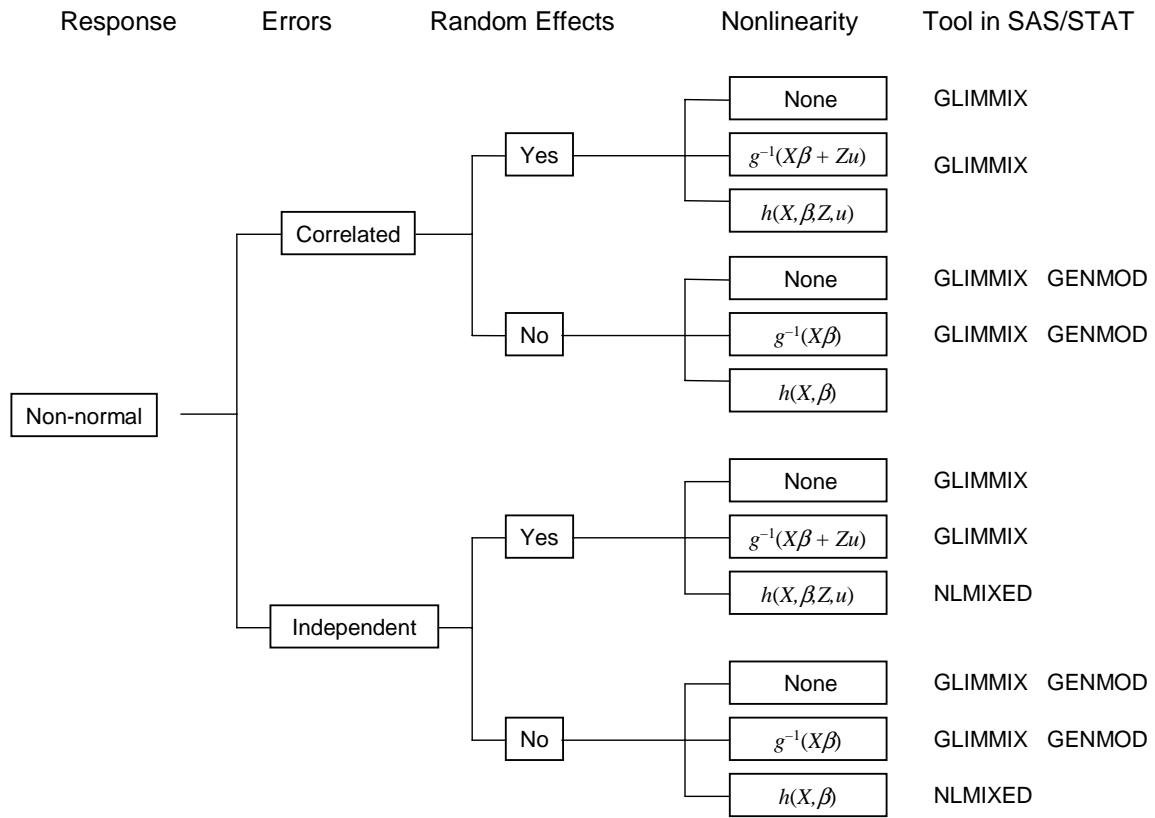
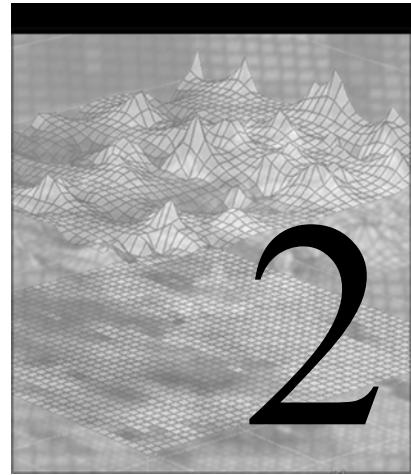
Figure 1.2 Flowchart Indicating Tools in SAS/STAT for Normal Distributed Response

Figure 1.3 Flowchart Indicating Tools in SAS/STAT for Non-normal Distributed Response





Randomized Block Designs

2.1 Introduction	18
2.2 Mixed Model for a Randomized Complete Blocks Design.....	18
2.2.1 Means and Variances from Randomized Blocks Design.....	19
2.2.2 The Traditional Method: Analysis of Variance.....	20
2.2.3 Using Expected Mean Squares	20
2.2.4 Example: A Randomized Complete Blocks Design	21
2.3 Using PROC MIXED to Analyze RCBD Data.....	22
2.3.1 Basic PROC MIXED Analysis Based on Sums of Squares.....	22
2.3.2 Basic PROC MIXED Analysis Based on Likelihood	26
2.3.3 Estimating and Comparing Means: LSMEANS, ESTIMATE, and CONTRAST Statements	28
2.3.4 Multiple Comparisons and Multiple Tests about Means	31
2.3.4 Confidence Intervals for Variance Components	35
2.3.5 Comparison of PROC MIXED with PROC GLM for the RCBD Data	37
2.4 Introduction to Theory of Mixed Models.....	42
2.4.1 Review of Regression Model in Matrix Notation.....	42
2.4.2 The RCBD Model in Matrix Notation	43
2.5 Example of an Unbalanced Two-Way Mixed Model: Incomplete Block Design	44
2.5.1 The Usual Intra-block Analysis of PBIB Data Using PROC GLM.....	46
2.5.2 The Combined Intra- and Inter-block Analysis of PBIB Data Using PROC MIXED	49
2.6 Summary	56

2.1 Introduction

Blocking is a research technique that is used to diminish the effects of variation among experimental units. The units can be people, plants, animals, manufactured mechanical parts, or numerous other objects that are used in experimentation. **Blocks** are groups of units that are formed so that units within the blocks are as nearly homogeneous as possible. Then levels of the factor being investigated, called **treatments**, are **randomly assigned to units within the blocks**. An experiment conducted in this manner is called a **randomized blocks design**. Usually, the primary objectives are to estimate and compare treatment means. In most cases, the **treatment effects** are considered **fixed** because the treatments in the experiment are the only ones to which inference is to be made. That is, no conclusions will be drawn about treatments that were not employed in the experiment. **Block effects** are usually considered **random** because the blocks in the experiment constitute only a small subset of the larger set of blocks over which inferences about treatment means are to be made. In other words, the investigator wants to estimate and compare treatment means with statements of precision (confidence intervals) and levels of statistical significance (from tests of hypothesis) that are valid in reference to the entire population of blocks, not just those blocks of experimental units in the experiment. To do so requires proper specification of random effects in model equations. In turn, computations for statistical methods must properly accommodate the random effects. The model for data from a randomized blocks design usually contains fixed effects for treatment contributions or factors and random effects for blocking factors contributions, making it a **mixed model**.

Section 2.2 presents the randomized blocks model as it is usually found in a basic statistical methods textbook. The standard analysis of variance methods are given, followed by an example to illustrate the standard methods. Section 2.3 illustrates using the MIXED procedure to obtain the results for the example, followed by results using the GLM procedure for comparison. Then, basic mixed model theory for the randomized blocks design is given in Section 2.4, including a presentation of the model in matrix notation. Section 2.5 presents an analysis of data from an incomplete blocks design to illustrate similarities and differences between analyses using PROC MIXED and PROC GLM with unbalanced data.

2.2 Mixed Model for a Randomized Complete Blocks Design

A randomized blocks design that has each treatment applied to an experimental unit in each block is called a **randomized complete blocks design** (RCBD). In the most common situation each treatment appears once in each block. Assume there are t treatments and r blocks of t experimental units and there will be one observation per experimental unit. Randomly assign each treatment to one experimental unit per block. Because each of the t treatments is assigned to an experimental unit in each of the r blocks, there are tr experimental units altogether. Letting Y_{ij} denote the response from the experimental unit that received treatment i in block j , the equation for the model is

$$Y_{ij} = \mu + \tau_i + b_j + e_{ij} \quad (2.1)$$

where

$$i = 1, 2, \dots, t$$

$$j = 1, 2, \dots, r$$

μ and τ_i are fixed parameters such that the mean for the i^{th} treatment is $\mu_i = \mu + \tau_i$

b_j is the random effect associated with the j^{th} block

e_{ij} is random error associated with the experimental unit in block j that received treatment i

Assumptions for random effects are as follows:

- Block effects are distributed **normally and independently** with mean 0 and variance σ_b^2 ; that is, the b_j ($j = 1, 2, \dots, r$) are distributed *iid* $N(0, \sigma_b^2)$.
- Errors e_{ij} are distributed **normally and independently** with mean 0 and variance σ^2 ; that is, the e_{ij} ($i = 1, 2, \dots, b_t; j = 1, 2, \dots, t$) are distributed *iid* $N(0, \sigma^2)$. The e_{ij} are also distributed independently of the b_j .

These are the conventional assumptions for a randomized blocks model.

2.2.1 Means and Variances from Randomized Blocks Design

The usual objectives of a randomized blocks design are to estimate and compare treatment means using statistical inference. Mathematical expressions are needed for the variances of means and differences between means in order to construct confidence intervals and conduct tests of hypotheses. It follows from equation (2.1) that a treatment mean, such as $\bar{Y}_{1\Box}$, can be written as

$$\bar{Y}_{1\Box} = \mu_1 + \bar{b}_{\Box} + \bar{e}_{1\Box} \quad (2.2)$$

Likewise, the difference between two means, such as $\bar{Y}_{1\Box} - \bar{Y}_{2\Box}$, can be written

$$\bar{Y}_{1\Box} - \bar{Y}_{2\Box} = \mu_1 - \mu_2 + \bar{e}_{1\Box} - \bar{e}_{2\Box} \quad (2.3)$$

From these expressions, the variances of $\bar{Y}_{1\Box}$ and $\bar{Y}_{1\Box} - \bar{Y}_{2\Box}$ are computed as

$$\text{Var}[\bar{Y}_{1\Box}] = (\sigma^2 + \sigma_b^2)/r \quad (2.4)$$

and

$$\text{Var}[\bar{Y}_{1\Box} - \bar{Y}_{2\Box}] = 2\sigma^2/r \quad (2.5)$$

Notice that the variance of a treatment mean $\text{Var}[\bar{Y}_{1\Box}]$ contains the block variance component σ_b^2 , but the variance of the difference between two means $\text{Var}[\bar{Y}_{1\Box} - \bar{Y}_{2\Box}]$ does *not* involve σ_b^2 .

This is the manifestation of the RCBD controlling block variation; the variance of differences between treatments are estimated free of block variation.

2.2.2 The Traditional Method: Analysis of Variance

Almost all statistical methods textbooks present analysis of variance (ANOVA) as a key component in analysis of data from a randomized blocks design. Our assumption is that readers are familiar with fundamental concepts for analysis of variance, such as degrees of freedom, sums of squares (SS), mean squares (MS), and expected mean squares (E[MS]). Readers needing more information concerning analysis of variance may consult Littell, Stroup, and Freund (2002), Milliken and Johnson (1992), or Winer (1971). Table 2.1 is a standard ANOVA table for the RCBD, showing sources of variation, degrees of freedom, mean squares, and expected mean squares.

Table 2.1 ANOVA Table for Randomized Complete Blocks Design

Source of Variation	df	MS	E[MS]
Blocks	$r - 1$	MS(Blks)	$\sigma^2 + t\phi^2$
Treatments	$t - 1$	MS(Trts)	$\sigma^2 + r\phi^2$
Error	$(r - 1)(t - 1)$	MS(Error)	σ^2

2.2.3 Using Expected Mean Squares

As the term implies, **expected mean squares** are the expectations of means squares. As such, they are the quantities that are estimated by mean squares in an analysis of variance. The expected mean squares can be used to motivate test statistics, to compute standard errors for means and comparisons of means, and to provide a way to estimate the variance components. The basic idea is to examine the expected mean square for a factor and see how it differs under null and alternative hypotheses. For example, the expected mean square for treatments, $E[MS(Trts)] = \sigma^2 + r\phi^2$, can be used to determine how to set up a test statistic for treatment differences. The null hypothesis is $H_0: \mu_1 = \mu_2 = \dots = \mu_t$. The expression ϕ^2 in $E[MS(Trts)]$ is

$$\phi^2 = (t-1)^{-1} \sum_{i=1}^t (\mu_i - \bar{\mu}_\square)^2$$

where $\bar{\mu}_\square$ is the mean of the μ_i . Thus, $\phi^2 = 0$ is equivalent to $\mu_1 = \mu_2 = \dots = \mu_t$. So, if the null hypothesis is true, $MS(Trts)$ simply estimates σ^2 . On the other hand, if $H_0: \mu_1 = \mu_2 = \dots = \mu_t$ is false, then $E[MS(Trts)]$ estimates a quantity larger than σ^2 . Now, $MS(Error)$ estimates σ^2 regardless of whether H_0 is true or false. Therefore, $MS(Trts)$ and $MS(Error)$ tend to be approximately the same magnitude if H_0 is true, and $MS(Trts)$ tends to be larger than $MS(Error)$ if $H_0: \mu_1 = \mu_2 = \dots = \mu_t$ is false. So a comparison of $MS(Trts)$ with $MS(Error)$ is an indicator of whether $H_0: \mu_1 = \mu_2 = \dots = \mu_t$ is true or false. In this way the expected mean squares show that a valid test statistic is the ratio $F = MS(Trts)/MS(Error)$.

Expected mean squares also can be used to estimate variance components, variances of treatment means, and differences between treatment means. Equating the observed mean squares to the expected mean squares provides the following system of equations:

$$MS(Blks) = \hat{\sigma}^2 + t\hat{\sigma}^2$$

$$MS(Error) = \hat{\sigma}^2$$

The solution for the variance components is

$$\hat{\sigma}^2 = \text{MS}(\text{Error}) \quad (2.6)$$

and

$$\hat{\sigma}_b^2 = \frac{1}{t} [\text{MS}(\text{Blks}) - \text{MS}(\text{Error})] \quad (2.7)$$

These are called **analysis of variance** estimates of the variance components. Using these estimates of the variance components, it follows that estimates of $\text{Var}[\bar{Y}_{1\cdot}]$ and $\text{Var}[\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}]$ are

$$\begin{aligned} \hat{\text{Var}}[\bar{Y}_{1\cdot}] &= (\hat{\sigma}^2 + \hat{\sigma}_b^2)/r \\ &= \frac{1}{rt} \text{MS}(\text{Blks}) + \frac{t-1}{rt} \text{MS}(\text{Error}) \end{aligned} \quad (2.8)$$

and

$$\hat{\text{Var}}[\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}] = \frac{2}{r} \text{MS}(\text{Error}) \quad (2.9)$$

The expression for $\text{Var}[\bar{Y}_{1\cdot}]$ points out a common misconception that the estimate of the variance of a treatment mean from a randomized blocks design is simply $\text{MS}(\text{Error})/r$. This misconception prevails in some textbooks and results in incorrect calculation of standard errors by computer software packages.

2.2.4 Example: A Randomized Complete Blocks Design

An example from Mendenhall, Wackerly, and Scheaffer (1996, p. 601) is used to illustrate analysis of data from a randomized blocks design.

Data for an RCB designed experiment are presented as Data Set 2.2, “BOND,” in Appendix 2, “Data Sets.” Blocks are ingots of a composition material, and treatments are metals (nickel, iron, or copper). Pieces of material from the same ingot are bonded using one of the metals as a bonding agent. The response is the amount of pressure required to break a bond of two pieces of material that used one of the metals as the bonding agent. Table 2.2 contains the analysis of variance table for the BOND data where the ingots form the blocks.

Table 2.2 ANOVA Table for BOND Data

Source of Variation	df	SS	MS	F	p-value
Ingots	6	268.29	44.72	4.31	0.0151
Metal	2	131.90	65.95	6.36	0.0131
Error	12	124.46	10.37		

The ANOVA table and the metal means provide the essential computations for statistical inference about the population means.

The ANOVA $F = 6.36$ for metal provides a statistic to test the null hypothesis $H_0: \mu_c = \mu_i = \mu_n$. The significance probability for the F -test is $p = 0.0131$, indicating strong evidence that the metal means are different. Estimates of the variance components are $\hat{\sigma}^2 = 10.37$ and $\hat{\sigma}_b^2 = (44.72 - 10.37)/3 = 11.45$. Thus, an estimate of the variance of a metal mean is $(\hat{\sigma}^2 + \hat{\sigma}_b^2)/7 = 3.11$, and the estimated standard error is $\sqrt{3.11} = 1.77$. An estimate of the variance of a difference between two metal means is $2\hat{\sigma}^2/7 = 2 \times 10.37/7 = 2.96$, and the standard error is $\sqrt{2.96} = 1.72$.

2.3 Using PROC MIXED to Analyze RCBD Data

PROC MIXED is a procedure with several capabilities for different methods of analysis. The estimates of the variance components can be obtained using sums of squares and expected means squares as described in the previous section or by using likelihood methods. Many of the estimation and inferential methods are implemented on the basis of the likelihood function and associated principles and theory (see Appendix 1, “Linear Mixed Model Theory,” for details). Readers may be more familiar with the analysis of variance approach described in the previous section; those results are obtained and presented in Section 2.3.1. The likelihood method results are presented in Section 2.3.2. The results of the analysis of variance and likelihood methods are compared and are shown to duplicate many of the results of the previous section.

2.3.1 Basic PROC MIXED Analysis Based on Sums of Squares

This section contains the code to provide the analysis of the RCBD with PROC MIXED using the sums of squares approach as described in Section 2.2.4. The METHOD=TYPE3 option is used to request that Type 3 sums of squares be computed along with their expected mean squares. Those mean squares and expected mean squares are used to provide estimates of the variance components and estimates of the standard errors associated with the means and comparisons of the means.

Program

The basic PROC MIXED statements for the RCBD data analysis are as follows:

```
proc mixed data=bond covtest cl method=type3;
  class ingot metal;
  model pres = metal / ddfm=kr;
  random ingot;
  lsmeans metal / diff adjust=simulate(report
                                             seed=4943838 cvadjust);
  estimate 'nickel mean' intercept 1 metal 0 0 1;
  estimate 'copper vs iron'           metal 1 -1 0;
  contrast 'nickel mean' intercept 1 metal 0 0 1;
  contrast 'copper vs iron'         metal 1 -1 0;
run;
```

The PROC MIXED statement calls the procedure. The METHOD=TYPE3 option requests that the Type 3 sums of squares method be used in estimating the variance components (Type 1, 2, or 3 sums of squares can be requested). The options COVTEST and CL provide estimates of the standard errors of the estimated variance components and 95% confidence intervals.

The CLASS statement specifies that INGOT and METAL are classification variables, not continuous variables.

The MODEL statement is an equation whose left-hand side contains the name of the response variable to be analyzed, in this case PRES. The right-hand side of the MODEL statement contains a list of the fixed effect variables, in this case the variable METAL. In terms of the statistical model, this specifies the τ_i parameters. (The intercept parameter μ is implicitly contained in all models unless otherwise declared by using the NOINT option.)

The RANDOM statement contains a list of the random effects, in this case the blocking factor INGOT, and represents the b_j terms in the statistical model.

The MODEL and RANDOM statements are the core essential statements for many mixed model applications, and the classification effects in the MODEL statement usually do not appear in the RANDOM statement and vice versa. Results from these statements appear in Output 2.1 and Output 2.2. The LSMEANS, ESTIMATE, and CONTRAST statements are discussed in subsequent sections.

Results

Output 2.1 Results of RCBD Data Analysis from PROC MIXED Using Type 3 Sums of Squares

Model Information	
Data Set	WORK.BOND
Dependent Variable	pres
Covariance Structure	Variance Components
Estimation Method	Type 3
Residual Variance Method	Factor
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Class Level Information		
Class	Levels	Values
ingot	7	1 2 3 4 5 6 7
metal	3	c i n

Dimensions	
Covariance Parameters	2
Columns in X	4
Columns in Z	7
Subjects	1
Max Obs Per Subject	21

Number of Observations	
Number of Observations Read	21
Number of Observations Used	21
Number of Observations Not Used	0

Interpretation

The “Model Information” table contains the model specifications for the data set used, the response variable, the methods used to estimate the variance components, the approximate degrees of freedom, and the standard errors for the fixed effects.

The “Class Level Information” table lists the levels for each of the variables declared in the class statement. You should be sure that these levels are specified consistently with how the study was conducted.

The “Dimensions” table shows how many columns are in the fixed effects matrix (\mathbf{X}) and in the random effects matrix (\mathbf{Z}) parts of the model, where the mixed model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ (see Section 1.6). For this study there are three levels of the treatment factor (metal) plus an intercept, which accounts for four columns in the \mathbf{X} matrix. There are seven ingots (blocks), thus there are seven columns in the \mathbf{Z} matrix. The inclusion of the RANDOM statement means there is one variance component for the ingot effects, plus the residual variance, providing two parameters in the covariance structure of the model. There is no SUBJECT= option used in this RANDOM statement, so PROC MIXED assumes that all observations are from the same subject, a quantity that can be ignored here.

The “Number of Observations” table indicates how many observations are in the data set and how many of those observations had valid data values for all variables used in the analysis. The difference between the number in the data set and the number used is the number of observations not used in the analysis. The information in these dimension specifications must match the information that is expected from the design being analyzed. Checking these values can help determine if there are data errors that need to be addressed, because they can cause the analysis to fail.

Results

Output 2.2 Results of the RCBD Data Analysis from PROC MIXED Using Type 3 Sums of Squares to Estimate the Variance Components

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
metal	2	131.900952	65.950476	Var(Residual) + Q(metal)	MS(Residual)	12	6.36	0.0131
ingot	6	268.289524	44.714921	Var(Residual) + 3 Var(ingot)	MS(Residual)	12	4.31	0.0151
Residual	12	124.459048	10.371587	Var(Residual)

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
ingot	11.4478	8.7204	1.31	0.1893	0.05	-5.6438	28.5394
Residual	10.3716	4.2342	2.45	0.0072	0.05	5.3332	28.2618

Fit Statistics	
-2 Res Log Likelihood	107.8
AIC (smaller is better)	111.8
AICC (smaller is better)	112.6
BIC (smaller is better)	111.7

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
metal	2	12	6.36	0.0131

Interpretation

The “Type 3 Analysis of Variance” table is the usual analysis of variance table with degrees of freedom, sums of squares, mean squares, expected mean squares, error terms for effects other than the residual, F -tests, and significance levels for these tests. The terms $\text{Var}(\text{Residual})$ and $\text{Var}(\text{ingot})$ denote the variance components σ^2 and σ_b^2 , respectively. See the discussion of the “Tests of Fixed Effects” table below for more detail.

The “Covariance Parameter Estimates” table gives estimates of the variance component parameters obtained by solving the set of equations from equating the observed mean squares to the expected mean squares. The estimate of σ_b^2 , the block variance component, is 11.4478 (labeled “ingot”), and the estimate of σ^2 , the error variance component, is 10.3716 (labeled “Residual”). The COVTEST option in the MODEL statement requests that the estimated standard errors of the estimated variance components be computed, and the CL option requests that 95% confidence intervals be provided where the interval about σ_b^2 is based on the Wald method and the interval about σ^2 is based on the Satterthwaite approximation.

The “Fit Statistics” table provides the value of the restricted log-likelihood function and three information measures (see Appendix 1, “Linear Mixed Model Theory,” for details).

The “Tests of Fixed Effects” table is like an abbreviated ANOVA table, showing a line of computations for each term in the MODEL statement. In this example, only METAL is included in the MODEL statement. The F -statistic is used to test the null hypothesis $H_0: \mu_c = \mu_i = \mu_n$ vs. H_a (not H_0). With 2 numerator and 12 denominator degrees of freedom, the F -value of 6.36 is significant at the 5% level (p -value is 0.0131). If the true METAL means are equal, then an F -value as large as 6.36 would occur less than 131 times in 10,000 by chance. This is the same F -test that was obtained from the analysis of variance.

In summary, these basic PROC MIXED computations are based on sums of squares and provide the same statistical computations obtained from analysis of variance methods for a balanced data set.

2.3.2 Basic PROC MIXED Analysis Based on Likelihood

A fundamental strength of PROC MIXED is its ability to apply likelihood methods to complex mixed models. The next section of code uses the default method for estimating the variance components, REML, standing for *REstricted (or REsidual) Maximum Likelihood* (Patterson and Thompson 1971). One could exclude METHOD=REML in the PROC MIXED statement and achieve the same results. The assumptions of normality of the various terms in the model equation (2.1) are required in order to construct the appropriate likelihood function that is maximized using PROC MIXED. The code to provide the likelihood based analysis is identical to that of the sums of squares method, except for the method specification.

Program

```
proc mixed data=bond method=reml;
  class ingot metal;
  model pres=metal / ddfm=kr;
  random ingot;
run;
```

The PROC MIXED statement invokes the procedure for the default method of estimation, REML. The CLASS, MODEL, and RANDOM statements are identical to those in Section 2.3.1. The results appear in Output 2.3.

Results

Output 2.3 Results of RCB Data Analysis from PROC MIXED METHOD=REML

Model Information	
Data Set	WORK.BOND
Dependent Variable	pres
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Class Level Information		
Class	Levels	Values
ingot	7	1 2 3 4 5 6 7
metal	3	c i n

Dimensions	
Covariance Parameters	2
Columns in X	4
Columns in Z	7
Subjects	1
Max Obs Per Subject	21

Number of Observations	
Number of Observations Read	21
Number of Observations Used	21
Number of Observations Not Used	0

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	112.40987952	
1	1	107.79020201	0.00000000

Convergence criteria met.

Covariance Parameter Estimates	
Cov Parm	Estimate
ingot	11.4478
Residual	10.3716

Fit Statistics	
-2 Res Log Likelihood	107.8
AIC (smaller is better)	111.8
AICC (smaller is better)	112.6
BIC (smaller is better)	111.7

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
metal	2	12	6.36	0.0131

Differences between results in Output 2.3 and Output 2.2 include the following:

- The “Model Information” table shows that REML is the specified method of estimating the variance components.
- The “Iteration History” table shows the sequence of evaluations to obtain (restricted) maximum likelihood estimates of the variance components. This portion of the output is not critical to most applications, such as the present RCBD analysis.
- The “Covariance Parameter Estimates” table gives estimates of the variance component parameters. The REML estimate of σ_b^2 , the block variance component, is 11.4478 (labeled “ingot”), and the estimate of σ^2 , the error variance component, is 10.3716 (labeled “Residual”). For this example of a balanced data set, these variance component estimates are identical to the estimates obtained from the analysis of variance method.

In summary, the default PROC MIXED computations are based on likelihood principles, but many of the statistical computations are the same as those obtained from analysis of variance methods for a balanced data set.

2.3.3 Estimating and Comparing Means: LSMEANS, ESTIMATE, and CONTRAST Statements

You can obtain treatment means from the LSMEANS (Least-Squares means) statement. Linear combinations of means can be estimated and compared using the ESTIMATE and CONTRAST statements. These statements all compute linear combinations of estimates of parameters defined in the MODEL statement.

The statement

```
model pres=metal;
```

specifies that expected value of Y_{ij} in equation (2.1) for a given metal can be represented as

$$E[Y_{ij}] = \mu + \tau_i = \mu_i$$

The **least-squares means** are model-based estimates of this mean. In other words, they are calculated from estimates of the parameters in the expressions:

$$\text{lsmean for metal } i = \hat{\mu} + \hat{\tau}_i$$

The SOLUTION option in the MODEL statement (you can abbreviate it as S) presents estimates of the fixed effect parameters.

Program

```
proc mixed data=bond;
  class ingot metal;
  model pres = metal / solution;
  random ingot;
  lsmeans metal / diff cl;
  estimate 'nickel mean' intercept 1 metal 0 0 1;
  estimate 'copper vs iron'           metal 1 -1 0;
  contrast 'copper vs iron'         metal 1 -1 0;
run;
```

Results

Results from the SOLUTION option and the LSMEANS statement appear in Output 2.4. Results from the ESTIMATE and CONTRAST statements appear in Output 2.5.

Output 2.4 Results from the SOLUTION Option and the LSMEANS Statement

Solution for Fixed Effects						
Effect	metal	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		71.1000	1.7655	6	40.27	<.0001
metal	c	-0.9143	1.7214	12	-0.53	0.6050
metal	i	4.8000	1.7214	12	2.79	0.0164
metal	n	0

Least Squares Means									
Effect	metal	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
metal	c	70.1857	1.7655	12	39.75	<.0001	0.05	66.3390	74.0324
metal	i	75.9000	1.7655	12	42.99	<.0001	0.05	72.0533	79.7467
metal	n	71.1000	1.7655	12	40.27	<.0001	0.05	67.2533	74.9467

Differences of Least Squares Means										
Effect	metal	_metal	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
metal	c	i	-5.7143	1.7214	12	-3.32	0.0061	0.05	-9.4650	-1.9636
metal	c	n	-0.9143	1.7214	12	-0.53	0.6050	0.05	-4.6650	2.8364
metal	i	n	4.8000	1.7214	12	2.79	0.0164	0.05	1.0493	8.5507

Interpretation

The “Solution for Fixed Effects” table contains estimates of the fixed effects parameters. Because the \mathbf{X} matrix is not of full rank, a consequence of having a classification variable in the MODEL statement and of including an intercept column in \mathbf{X} , the estimates labeled “Intercept,” “Metal c,” “Metal i,” and “Metal n” are not unbiased estimates of μ , τ_c , τ_i , and τ_n . The problem of a singular \mathbf{X} matrix is solved by setting to zero the fixed effects solution corresponding to the last column of the classification effect METAL in the MODEL statement. The other solutions are unbiased estimates of deviations from the reference level. The row labeled “Intercept” estimates $\mu + \tau_n$, the row labeled “Metal c” estimates $\tau_c - \tau_n$, the row labeled “Metal i” estimates $\tau_i - \tau_n$, and the row labeled “Metal n” estimates $\tau_n - \tau_n = 0$. Consequently,

$$\hat{\mu} + \hat{\tau}_n = 71.100$$

$$\hat{\tau}_c - \hat{\tau}_n = -0.9143$$

$$\hat{\tau}_i - \hat{\tau}_n = 4.800$$

Because there is no unique way of resolving the singularity in \mathbf{X} , there is no unique way of deriving solutions. For example, changing the names of the metals in such a way that their order changes affects the choice of the reference level. However, certain linear combinations of the estimates are unique. These linear combinations correspond to **estimable functions** (See Littell, Stroup, and Freund 2002). Least-squares means are such estimable functions.

The “Least Squares Means” table provides the model-based estimates of the least-squares means. Since, for example, $\mu_c = \mu + \tau_c = \mu + \tau_n + (\tau_c - \tau_n)$, the least-squares means in this example can be related easily to the estimates displayed in the “Solution for Fixed Effects” table:

$$\hat{\mu}_c = \hat{\mu} + \hat{\tau}_c = \hat{\mu} + \hat{\tau}_n + (\hat{\tau}_c - \hat{\tau}_n) = 71.1 + (-0.9142) = 70.1857$$

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \hat{\mu} + \hat{\tau}_n + (\hat{\tau}_i - \hat{\tau}_n) = 71.1 + (4.8) = 75.9$$

$$\hat{\mu}_n = \hat{\mu} + \hat{\tau}_n = \hat{\mu} + \hat{\tau}_n + (\hat{\tau}_n - \hat{\tau}_n) = 71.1 + (0.0) = 71.1$$

For these “balanced data,” the least-squares means are simply the arithmetic averages of the data values for each of the treatments.

Also printed are estimated standard errors for the least-squares means; they all equal

$$\sqrt{(\hat{\sigma}^2 + \hat{\sigma}_b^2)/7} = 1.7655$$

This is a valid estimate of the true standard error $\sqrt{(\sigma^2 + \sigma_b^2)/7}$, because $(\sigma^2 + \sigma_b^2)/7$ is the variance of a metal mean, and this variance is estimated unbiasedly by

$$(\hat{\sigma}^2 + \hat{\sigma}_b^2)/7$$

The “Differences of Least Squares Means” table shows all possible pairwise comparisons among the three treatments. The interpretation of the results is that each row is a comparison of the mean in the column with heading **metal** minus the mean in the column with heading **_metal**. The estimated standard errors of the differences are all equal to

$$\sqrt{2\hat{\sigma}^2/7} = 1.7214$$

T-values are computed to test that the differences between the respective means are zero. The results declare both copper and nickel as different from iron but copper and nickel as not different from each other.

Linear combinations of means can be estimated with the ESTIMATE statement. For illustration, consider estimating the linear combination equal to the nickel mean, μ_n . First, express the nickel mean as a linear combination of the model parameters, $\mu_n = \mu + \tau_n$. More explicitly, $\mu_n = 1 \times \mu + 0 \times \tau_c + 0 \times \tau_i + 1 \times \tau_n$. These coefficients of the fixed effects solutions are reflected in the ESTIMATE statement:

```
estimate 'nickel mean' intercept 1 metal 0 0 1;
```

Similarly, the difference between the means for copper and iron is $\mu_c - \mu_i = \tau_c - \tau_i$. The ESTIMATE and CONTRAST statements for this comparison are as follows:

```
estimate 'copper vs iron'          metal 1 -1 0;
contrast 'copper vs iron'         metal 1 -1 0;
```

Output 2.5 displays the results.

Output 2.5 Inference about Linear Combinations of Means

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
nickel mean	71.1000	1.7655	12	40.27	<.0001
copper vs iron	-5.7143	1.7214	12	-3.32	0.0061

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
copper vs iron	1	12	11.02	0.0061

Interpretation

The “Estimates” table collects the results of the ESTIMATE statements. The nickel mean and the differences between the copper and iron means are the same as those obtained from the LSMEANS statement in Output 2.4.

The “Contrasts” table collects the results of CONTRAST statements. The contrast of ‘copper vs iron’ tests the null hypothesis $H_0: \mu_c - \mu_i = 0$. It is associated with an F -statistic of 11.02. This equals the square of the t -value from the corresponding ESTIMATE statement (-3.32^2).

The ESTIMATE and CONTRAST statements are used to estimate and test hypotheses about linear combinations of terms in the mixed model, including random effects. The ESTIMATE statement consists of a single linear combination, while the CONTRAST statement can consist of several linear combinations where the resulting F -statistic provides a simultaneous test of the hypotheses that the linear combinations of the parameters are equal to zero. Note that the ESTIMATE statements in the GLIMMIX procedure allow multiple-row estimates and can adjust the corresponding t -tests for multiplicity.

2.3.4 Multiple Comparisons and Multiple Tests about Means

Frequently, multiple comparisons among the set of treatment means are among the objectives of an experiment. Whenever multiple comparisons are being carried out on the same data set, some strategy should be used to control the Type I error rates due to multiple testing (see, for example, Westfall et al. 1999 and Westfall 2002).

Pairwise comparisons of least-squares means are obtained with the DIFF option in the LSMEANS statement. Output 2.4 displays the pairwise differences and their estimated standard errors along with the results of t -tests for the statistical significance of the difference between the means in each pair. The results declare both copper and nickel different from iron but not different from each other. When performing multiple comparisons, one often wishes to use a procedure that controls the family-wise error rate. The ADJUST= option of the LSMEANS statement provides various adjustments of p -values and confidence limits for multiplicity. The ADJUST=SIMULATE option of the LSMEANS statement provides a family-wise error rate protection. Using the SIMULATE option to carry out the multiple comparisons does not require that the F -test for the effect be significant as is the case for using Fisher’s Protected LSD.

Program

Add the following LSMEANS statement to the previous PROC MIXED run (or replace the previous LSMEANS statement):

```
lsmeans metal / diff cl
    adjust=simulate(report seed=4943838 cvadjust);
```

The DIFF and CL options request pairwise differences and their confidence limits. The ADJUST=SIMULATE option is used with several sub-options. The SEED= sub-option sets the random number seed for the simulations, for reproducibility of the results. The REPORT

option requests a table detailing the quantile simulations. The CVADJUST option specifies that the quantile be estimated by the control variate adjustment method of Hsu and Nelson (1998).

Results

Output 2.6 Results from Pairwise Comparisons with the Simulate Option in the LSMEANS Statement

Details for Quantile Simulation	
Random number seed	4943838
Comparison type	All
Sample size	12605
Target alpha	0.05
Accuracy radius (target)	0.005
Accuracy radius (actual)	0.0034
Accuracy confidence	99%

Simulation Results				
Method	95% Quantile	Estimated Alpha	99% Confidence Limits	
Simulated	2.665074	0.0500	0.0466	0.0534
Tukey-Kramer	2.667757	0.0498	0.0463	0.0532
Bonferroni	2.779473	0.0402	0.0371	0.0433
Sidak	2.770301	0.0410	0.0379	0.0442
GT-2	2.747299	0.0425	0.0393	0.0457
Scheffe	2.787577	0.0395	0.0364	0.0427
T	2.178813	0.1154	0.1106	0.1202

Least Squares Means									
Effect	metal	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
metal	c	70.1857	1.7655	11.6	39.75	<.0001	0.05	66.3246	74.0469
metal	i	75.9000	1.7655	11.6	42.99	<.0001	0.05	72.0388	79.7612
metal	n	71.1000	1.7655	11.6	40.27	<.0001	0.05	67.2388	74.9612

Differences of Least Squares Means												
Effect	metal	_metal	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P	Alpha	Lower	Upper
metal	c	i	-5.7143	1.7214	12	-3.32	0.0061	Simulate	0.0145	0.05	-9.4650	-1.9636
metal	c	n	-0.9143	1.7214	12	-0.53	0.6050	Simulate	0.8570	0.05	-4.6650	2.8364
metal	i	n	4.8000	1.7214	12	2.79	0.0164	Simulate	0.0394	0.05	1.0493	8.5507

Differences of Least Squares Means				
Effect	metal	_metal	Adj Lower	Adj Upper
metal	c	i	-10.3020	-1.1266
metal	c	n	-5.5020	3.6734
metal	i	n	0.2123	9.3877

Interpretation

The “Details for Quantile Simulation” table reports the seed value at the beginning of the quantile simulations. It equals the value specified in the SEED= sub-option of ADJUST=SIMULATE. The “All” comparison type indicates that the multiplicity adjustment is performed for all pairwise comparisons, as compared to, for example, adjustments against a control level. PROC MIXED determines the necessary sample size for the simulation as the number of samples to achieve a 99% confidence interval for the target significance level (here, $\alpha = 0.05$). This sample size equals 12605. The “Simulation Results” table shows that this confidence level was achieved ($0.0466 - 0.0534$).

The “Differences of Least Squares Means” table in Output 2.6 contains confidence intervals about the differences without adjusting for multiple comparisons (Lower and Upper) and with the SIMULATE adjustment (Adj_Lower and Adj_Upper). For the individual comparison of iron to nickel, the 95% confidence interval is 1.049 to 8.551. But if you take into account that there are three confidence intervals, then the 95% simultaneous confidence interval about the difference between the means of iron and nickel is 0.212 to 9.388. As expected, the simultaneous confidence intervals are wider than the individual or unadjusted confidence intervals. Similarly, the adjusted p -values (“Adj P” column) exceed the unadjusted p -values (“Pr > |t|” column).

You can also examine the results of least-squares means comparisons in graphical form with the GLIMMIX¹ procedure. The following program fits the mixed model randomized block design and requests plots of least-squares means differences with and without simulated multiplicity adjustments.

Program

```

ods html;
ods graphics on;
proc glimmix data=bond plots=diffplot;
  class ingot metal;
  model press=metal;
  random ingot;
  lsmeans metal;
  lsmeans metal / adjust=simulate(seed=4943838);
run;
ods graphics off;
ods html close;

```

The ODS HTML statement specifies the destination for output, including ODS statistical graphics. The ODS GRAPHICS ON statement indicates that ODS statistical graphics are to be produced. The CLASS, MODEL, and RANDOM statements in GLIMMIX are the same as the

¹ The GLIMMIX procedure is an add-on in SAS 9.1 to SAS/STAT for the (32-bit) Windows platform. It does not ship with SAS 9.1. You can obtain the GLIMMIX procedure for SAS 9.1 as a download from www.sas.com/statistics. This site also contains the documentation for the GLIMMIX procedure.

earlier PROC MIXED statements. The PLOTS=DIFFPLOT option requests plots of least-squares means differences.

Results

The results are displayed as Diffograms in Figures 2.1 and 2.2.

Figure 2.1 Diffogram for METAL Effect without Multiplicity Adjustment

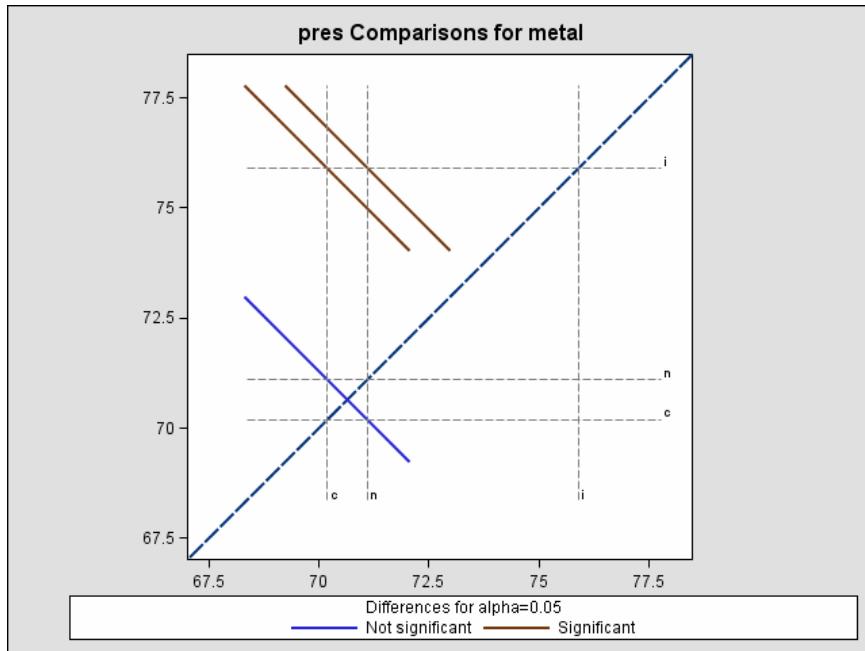
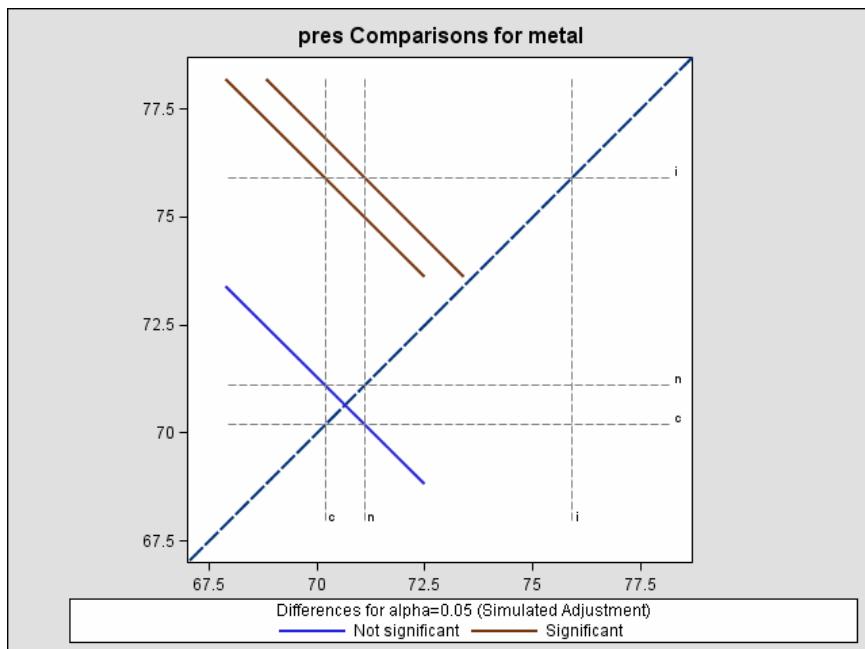


Figure 2.2 Diffogram for METAL Effect with Multiplicity Adjustment



Interpretation

The Diffogram, also known as a mean-mean scatter plot (Hsu 1996, Hsu and Peruggia 1994), is a graphical display of all pairwise differences. The 45° reference line indicates whether two least-squares means are significantly different at a given significance level. Vertical and horizontal reference (grid) lines are drawn at the values of the least-squares means. A line is drawn at the intersection of the grid lines that corresponds to the $(1-\alpha) \times 100\%$ confidence interval of the difference of the two least-squares means in the comparison. For example, the solid line closest to the vertical axis in Figures 2.1 and 2.2 corresponds to the comparison of nickel and copper. Since the line crosses the dashed 45° reference line, it is not significant at the 5% level. The other two comparisons are significant. Notice that the multiplicity-adjusted results in Figure 2.2 are reflected in longer solid lines, and the adjusted confidence intervals for the least-squares means differences are wider.

2.3.4 Confidence Intervals for Variance Components

Confidence intervals can be used when it is of interest to access the uncertainty about the variance components in the model. A $(1-\alpha) \times 100\%$ confidence interval about σ^2 can be constructed using the chi-square distribution, as

$$\frac{(b-1)(t-1)\hat{\sigma}^2}{\chi_{(1-\alpha/2),(b-1)(t-1)}^2} \leq \sigma^2 \leq \frac{(b-1)(t-1)\hat{\sigma}^2}{\chi_{\alpha/2,(b-1)(t-1)}^2}$$

where $\chi_{(1-\alpha/2),(b-1)(t-1)}^2$ and $\chi_{\alpha/2,(b-1)(t-1)}^2$ are the lower and upper $\alpha/2$ percentage points of a central chi-square distribution with $(b-1) \times (t-1)$ degrees of freedom, respectively. When the estimate of σ_b^2 is positive, an approximate $(1-\alpha) \times 100\%$ confidence interval about σ_b^2 can be constructed using a Satterthwaite (1946) approximation. The estimate of σ_b^2 is a linear combination of mean squares, which in general can be expressed as

$$\hat{\sigma}_b^2 = \sum_{i=1}^s q_i MS_i$$

where the i^{th} mean square is based on f_i degrees of freedom. The approximate number of Satterthwaite degrees of freedom associated with $\hat{\sigma}_b^2$ is

$$v = \frac{(\hat{\sigma}_b^2)^2}{\sum_{i=1}^s [(q_i MS_i)^2] / f_i}$$

For the randomized complete block,

$$\hat{\sigma}_b^2 = \frac{1}{t} (\text{MS(Blks)} - \text{MS(Error)})$$

and the approximate number of degrees of freedom is

$$v = \frac{\left(\hat{\sigma}_b^2\right)^2}{\frac{\left(t^{-1}\text{MS(Blks)}\right)^2}{b-1} + \frac{\left(t^{-1}\text{MS(Error)}\right)^2}{(b-1)(t-1)}}$$

A $(1-\alpha) \times 100\%$ confidence interval about σ_b^2 can be constructed using the chi-square distribution, as

$$\frac{v\hat{\sigma}_b^2}{\chi_{(1-\alpha/2),v}^2} \leq \sigma_b^2 \leq \frac{v\hat{\sigma}_b^2}{\chi_{\alpha/2,v}^2}$$

where $\chi_{(1-\alpha/2),v}^2$ and $\chi_{\alpha/2,v}^2$ are the lower and upper $\alpha/2$ percentage points with v degrees of freedom, respectively.

Program

The following includes the PROC MIXED code to fit the model using the Type 3 sums of squares and the ODS statement to generate an output data set with the Type 3 analysis of variance table. The information can be extracted from that data set to construct a confidence interval about σ_b^2 .

```
proc mixed data=bond method=type3 covtest cl;
  class ingot metal;
  model pres=metal / ddfm=kr;
  random ingot;
  ods select Type3 CovParms;
  ods output type3=type3;
run;
data est_var_ingot;
  merge type3(where =(source='ingot')
               rename=(ms=msingot
                       df=dfingot))
        type3(where =(source='Residual')
               rename=(ms=mseerror
                       df=dferror));
  var_Ingot    = (msingot-mserror)/3;
  df_var_ingot = (var_ingroup**2)/(((msingot/3)**2)/dfingot)+((mserror/3)**2)/dferror));
  low_var_ingroup= df_var_ingroup*var_Ingot/cinv(.975,df_var_ingroup);
  up_var_ingroup = df_var_ingroup*var_Ingot/cinv(.025,df_var_ingroup);
run;
proc print data=est_var_ingroup;
  var msingot dfingot mseerror dferror
        var_ingroup df_var_ingroup
        low_var_ingroup up_var_ingroup;
run;
```

Results

The results of computing the estimate of the variance components and using the Satterthwaite approximation to construct the confidence interval about σ_b^2 are given in Output 2.7.

Output 2.7 Type 3 Analysis of Variance Table and Computations for Confidence Interval about σ_b^2

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
metal	2	131.900952	65.950476	Var(Residual) + Q(metal)	MS(Residual)	12	6.36	0.0131
ingot	6	268.289524	44.714921	Var(Residual) + 3 Var(ingot)	MS(Residual)	12	4.31	0.0151
Residual	12	124.459048	10.371587	Var(Residual)

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
ingot	11.4478	8.7204	1.31	0.1893	0.05	-5.6438	28.5394
Residual	10.3716	4.2342	2.45	0.0072	0.05	5.3332	28.2618

Obs	msingot	dfingot	mseerror	dferror	var_Ingot	df_var_ingot	low_var_ingot	up_var_ingot
1	44.714921	6	10.371587	12	11.4478	3.44670	3.88112	121.547

Interpretation

The “Type 3 Analysis of Variance” table in Output 2.7 contains the result for the analysis of variance including the expected means squares.

The “Covariance Parameter Estimates” table contains the estimates of the variance components. The confidence interval about σ_b^2 is computed using the Wald method, while the confidence interval about σ^2 is computed using the chi-square distribution. The third table (from PROC PRINT) contains the results for the Satterthwaite approximation to construct the confidence interval about σ_b^2 . The approximate number of degrees of freedom is 3.446 and the 95% confidence interval is $3.881 < \sigma_b^2 < 121.547$.

2.3.5 Comparison of PROC MIXED with PROC GLM for the RCBD Data

PROC GLM was the principal SAS procedure for analyzing mixed models data prior to the advent of PROC MIXED, even though the basic computations of PROC GLM are for fixed effects models. The GLM procedure uses statements similar to those used by PROC MIXED. In this section you will see differences and similarities in the statements and output. However, you will not see complete coverage of PROC GLM capabilities. Refer to Littell, Stroup, and Freund (2002) for more detailed PROC GLM coverage.

Program

Statements for PROC GLM to obtain the ANOVA table, mean estimates, and comparisons analogous to those discussed in Sections 1.3.1 and 1.3.2 are as follows:

```
proc glm data=bond;
  class ingot metal;
  model pres = metal ingot / ss3;
  random ingot;
  lsmeans metal / pdiff
    adjust=simulate(report seed=4943838 cvadjust);
  estimate 'nickel mean' intercept 1 metal 0 0 1;
  estimate 'copper vs iron'               metal 1 -1 0;
  contrast 'nickel mean' intercept 1 metal 0 0 1;
  contrast 'copper vs iron'              metal 1 -1 0;
run;
```

Results

Results of these statements appear in Output 2.8.

Output 2.8 Randomized Blocks Analysis with PROC GLM

Class Level Information		
Class	Levels	Values
ingot	7	1 2 3 4 5 6 7
metal	3	c i n

Number of Observations Read	21
Number of Observations Used	21

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	400.1904762	50.0238095	4.82	0.0076
Error	12	124.4590476	10.3715873		
Corrected Total	20	524.6495238			

R-Square	Coeff Var	Root MSE	pres Mean
0.762777	4.448490	3.220495	72.39524

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metal	2	131.9009524	65.9504762	6.36	0.0131
ingot	6	268.2895238	44.7149206	4.31	0.0151

Least Squares Means

Details for Quantile Simulation	
Random number seed	4943838
Comparison type	All
Sample size	12605

Details for Quantile Simulation	
Target alpha	0.05
Accuracy radius (target)	0.005
Accuracy radius (actual)	0
Accuracy confidence	99%

Simulation Results		
Method	95% Quantile	Exact Alpha
Simulated	2.667757	0.0500
Tukey	2.667757	0.0500
Bonferroni	2.779473	0.0411
Sidak	2.770301	0.0417
GT-2	2.747299	0.0435
Scheffe	2.787577	0.0405
T	2.178813	0.1156

Adjustment for Multiple Comparisons: Simulated

metal	pres LSMEAN	Standard Error	Pr > t	LSMEAN Number
c	70.1857143	1.2172327	<.0001	1
i	75.9000000	1.2172327	<.0001	2
n	71.1000000	1.2172327	<.0001	3

east Squares Means for effect metal Pr > t for H0: LSMean(i)=LSMean(j)			
Dependent Variable: pres			
i/j	1	2	3
1		0.0156	0.8578
2	0.0156		0.0404
3	0.8578	0.0404	

metal	pres LSMEAN	95% Confidence Limits	
c	70.185714	67.533592	72.837836
i	75.900000	73.247878	78.552122
n	71.100000	68.447878	73.752122

Least Squares Means for Effect metal				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-5.714286	-10.306634	-1.121937
1	3	-0.914286	-5.506634	3.678063
2	3	4.800000	0.207652	9.392348

Source	Type III Expected Mean Square
metal	Var(Error) + Q(metal)
ingot	Var(Error) + 3 Var(ingot)

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metal	2	131.900952	65.950476	6.36	0.0131
Ingot	6	268.289524	44.714921	4.31	0.0151
Error: MS(Error)	12	124.459048	10.371587		

Contrast	Contrast Expected Mean Square
nickel mean	Var(Error) + Var(ingot) + Q(Intercept,metal)
copper vs iron	Var(Error) + Q(metal)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
nickel mean	1	35386.47000	35386.47000	3411.87	<.0001
copper vs iron	1	114.28571	114.28571	11.02	0.0061

Parameter	Estimate	Standard Error	t Value	Pr > t
nickel mean	71.1000000	1.21723265	58.41	<.0001
copper vs iron	-5.7142857	1.72142692	-3.32	0.0061

Interpretation

Following is a comparison of syntax and output for PROC GLM and PROC MIXED statements:

- First, both procedures use the same CLASS statements; if a variable is a classification variable in PROC GLM, then so it is in PROC MIXED.
- The MODEL statements of PROC MIXED and PROC GLM are *not* exactly the same. Here is a very important distinction between the procedures: in MIXED, you list only the *fixed* effects in the right-hand side of the MODEL statement. But in the GLM procedure you list all effects, fixed *and* random, in the MODEL statement, although PROC GLM does not really treat the declared random effects as random. The options in

PROC GLM for inference accommodating random effects are adaptations of the fixed effect computations for this procedure. PROC MIXED, on the other hand, was conceived from the outset for mixed models. The distinction between MODEL statements in PROC MIXED and PROC GLM carries over to the output from the two procedures. In Output 2.8, from PROC GLM, you see an ANOVA table listing all the terms in the MODEL statement, with no distinction between fixed and random. But in Output 2.2, from PROC MIXED, the “Tests of Fixed Effects” table contains only terms in the MODEL statement.

- The LSMEANS statements for the two procedures are essentially the same, except that you need the STDERR option in the LSMEANS statement for PROC GLM to print the estimated standard errors. Here is another important distinction between the procedures: In comparing Output 2.4 with Output 2.8, you see that the LSMEANS estimates are the same for the two procedures, but their standard errors are *not* the same. This is due to the inherent fixed effect nature of PROC GLM. The standard errors for the LSMEANS printed by PROC GLM are computed as $\sqrt{\sigma^2/7}$, which would be appropriate if INGOT (the blocks) were fixed. Recall from Section 2.3.1 that PROC MIXED performed correct computations for the standard errors of least-squares means when the INGOT effect was random.
- Syntax for ESTIMATE statements is the same in PROC GLM as in PROC MIXED for estimating linear combinations of fixed effects. The remarks comparing standard errors of least-squares means estimates from PROC MIXED and PROC GLM also apply to ESTIMATE statements. You see in Output 2.8 from PROC GLM that the standard error of the estimate of the nickel mean is the same as for the nickel least-squares mean, which previously was stated to be incorrect.
- Estimates, their standard errors, and tests of the difference between the COPPER and IRON means are the same for PROC GLM (Output 2.8) and PROC MIXED (Output 2.6). This is true for the present example, but not for all mixed model data sets, as you will see in the case of an incomplete block design in Section 2.5.2.
- The RANDOM statements for PROC MIXED and PROC GLM represent another major distinction between the two procedures although they have the same appearance for the present example. In PROC MIXED, listing INGOT in the RANDOM statement causes all standard errors and test statistics to incorporate the information that the effect is random. This is not true in PROC GLM. The RANDOM statement in PROC GLM (as used here) merely computes expected mean squares for terms in the MODEL statement and for linear combinations in the CONTRAST statement. You must then digest the information in the expected means squares table and formulate appropriate tests. (The TEST option in the PROC GLM RANDOM statement will do this automatically for terms in the MODEL statement, but not for CONTRAST statements.) In the RCBD example, the default tests computed by PROC GLM are correct, so no modification is needed for the test of differences from the MODEL effects and CONTRAST statements.

These comparisons of PROC MIXED and PROC GLM are summarized in Table 2.3.

Table 2.3 Summary Comparison of Syntax and Output for PROC GLM and PROC MIXED

Statement	PROC MIXED	PROC GLM
CLASS	List classification variables	Same
MODEL	Specify dependent variable and list fixed effect	Specify dependent variable and list all terms in model
RANDOM	Specify random effects	Obtain table of expected mean squares
LSMEANS	Estimate means for fixed effect factors	Estimate means for model terms
ESTIMATE	Estimate linear combination of model terms	Estimate linear combination of model terms
CONTRAST	Test set of linear combinations of model terms	Test set of linear combination of model terms

2.4 Introduction to Theory of Mixed Models

The randomized complete blocks design presents one of the simplest applications of mixed models. It has one fixed effect (treatments) and one random effect (blocks). In this section, the RCBD is used to introduce the representation of the mixed model as it appears throughout this book. Refer to Appendix 1, “Linear Mixed Model Theory,” for the general setting and for additional details.

2.4.1 Review of Regression Model in Matrix Notation

The equation for the standard equation for the linear regression model is

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e$$

In an application there would be n observed values of Y and the corresponding values of x_1, \dots, x_k . Often the values of Y are considered to be independent realizations with equal variance. These can be represented in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where \mathbf{Y} is the n vector of observations, \mathbf{X} is an $n \times (k+1)$ matrix containing a column of 1's and columns of observed values of x_1, \dots, x_k , and \mathbf{e} is a vector of realizations of the errors e . At this point we assume only that the error vector \mathbf{e} has mean $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}$, denoted $\mathbf{e} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$. This covariance matrix reflects the assumption of uncorrelated errors. In mixed models, especially linear mixed models, we typically add the assumptions that the errors are normally distributed, denoted $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Note that when the normality assumption is added, lack of correlation among the errors is tantamount to independence of the errors. In mixed models, the covariance structure of \mathbf{Y} often does not meet the independence assumption. But the covariance is usually structured so that it can be accommodated indirectly by random effects in the model or directly by parameterizing $\text{Var}[\mathbf{e}]$.

2.4.2 The RCBD Model in Matrix Notation

The RCBD model in equation (2.1) can be written in matrix notation. In explicit detail, the model equation is

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{t1} \\ \vdots \\ Y_{1r} \\ \vdots \\ Y_{tr} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 1 & 0 & 1 & \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_t \end{bmatrix} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_r \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{t1} \\ \vdots \\ e_{1r} \\ \vdots \\ e_{tr} \end{bmatrix} \quad (2.10)$$

with terms defined following equation (2.1). In more compact matrix notation the equation is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2.11)$$

where

- \mathbf{Y} is the vector of observations
- \mathbf{X} is the treatment design matrix
- $\boldsymbol{\beta}$ is the vector of treatment fixed effect parameters
- \mathbf{Z} is the block design matrix
- \mathbf{u} is the vector of random block effects
- \mathbf{e} is the vector of experimental errors

The model equation (2.11) states that the vector \mathbf{Y} of observations can be expressed as a sum of fixed treatment effects $\mathbf{X}\boldsymbol{\beta}$, random block effects $\mathbf{Z}\mathbf{u}$, and random experimental errors \mathbf{e} . The $\mathbf{X}\boldsymbol{\beta}$ portion is defined by the MODEL statement, and the $\mathbf{Z}\mathbf{u}$ portion is defined by the RANDOM statement. It is not necessary in this example to define the experimental errors \mathbf{e} .

For the RCBD model in matrix notation, the random vector \mathbf{u} has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma_b^2 \mathbf{I}$, $\mathbf{u} \sim N(\mathbf{0}, \sigma_b^2 \mathbf{I}_t)$, and the random vector \mathbf{e} is distributed $N(\mathbf{0}, \sigma_e^2 \mathbf{I}_{tr})$.

The variance of the observation vector \mathbf{Y} is

$$\text{Var}[\mathbf{Y}] = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

$$\begin{aligned} &= \begin{bmatrix} \sigma_b^2 \mathbf{J}_t & \mathbf{0}_{t \times t} & \cdots & \mathbf{0}_{t \times t} \\ \mathbf{0}_{t \times t} & \sigma_b^2 \mathbf{J}_t & \cdots & \mathbf{0}_{t \times t} \\ \mathbf{0}_{t \times t} & \mathbf{0}_{t \times t} & \ddots & \vdots \\ \mathbf{0}_{t \times t} & \mathbf{0}_{t \times t} & \cdots & \sigma_b^2 \mathbf{J}_t \end{bmatrix} + \sigma^2 \mathbf{I}_tr \\ &= \begin{bmatrix} \sigma_b^2 \mathbf{J}_t + \sigma^2 \mathbf{I}_t & \mathbf{0}_{t \times t} & \cdots & \mathbf{0}_{t \times t} \\ \mathbf{0}_{t \times t} & \sigma_b^2 \mathbf{J}_t + \sigma^2 \mathbf{I}_t & \cdots & \mathbf{0}_{t \times t} \\ \mathbf{0}_{t \times t} & \mathbf{0}_{t \times t} & \ddots & \vdots \\ \mathbf{0}_{t \times t} & \mathbf{0}_{t \times t} & \cdots & \sigma_b^2 \mathbf{J}_t + \sigma^2 \mathbf{I}_t \end{bmatrix} \end{aligned}$$

where $\sigma_b^2 \mathbf{J}_t + \sigma^2 \mathbf{I}_t$ is the covariance matrix of the observations in a particular block, $\mathbf{0}_{t \times t}$ is a $t \times t$ matrix of zeros, and \mathbf{J}_t is a $t \times t$ matrix of ones.

The matrix \mathbf{X} is not of full column rank and so $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ is singular; thus a generalized inverse must be used to obtain a GLS estimate of the fixed effect parameter vector $\boldsymbol{\beta}$. But the treatment means and differences between treatment means are estimable parameters. Thus, no matter what generalized inverse is used, there will be a vector \mathbf{K} for which $\mathbf{K}'\boldsymbol{\beta}$ is equal to a mean or a difference between means. For example, choosing $\mathbf{K}' = [1, 1, 0, \dots, 0]$ gives $\mathbf{K}'\boldsymbol{\beta} = \mu + \tau_1 = \mu_1$. Then the general theory gives

$$\text{Var}[\mathbf{K}'\hat{\boldsymbol{\beta}}] = (\sigma_b^2 + \sigma^2)/r$$

where $\hat{\boldsymbol{\beta}}$ is the generalized least-squares estimate. Likewise, $\mathbf{K}' = [0, 1, -1, 0, \dots, 0]$ gives $\mathbf{K}'\boldsymbol{\beta} = \mu_1 - \mu_2$, and $\text{Var}[\mathbf{K}'\hat{\boldsymbol{\beta}}] = 2\sigma^2/r$. These are the expressions presented in Section 2.2.1.

In the case of a relatively simple, balanced design such as an RCBD, the variance expressions can be derived directly from the model. This was the approach in Section 2.2.1. But in more complicated unbalanced situations, the general theoretical results must be invoked. In this subsection, we have illustrated the general results in the RCBD setting to confirm their validity and to assist you in becoming more comfortable in using the general linear mixed model.

2.5 Example of an Unbalanced Two-Way Mixed Model: Incomplete Block Design

In some applications of blocking there are not enough experimental units in each block to accommodate all treatments. **Incomplete block designs** are designs in which only a subset of the treatments are applied in each block. The treatments that go into each block should be selected in order to provide the most information relative to the objectives of the experiment.

Three types of incomplete block designs are the so-called **balanced** incomplete block design (BIBD), **partially balanced** incomplete block design (PBIBD), and **unbalanced** incomplete block design. This does not mean “balanced” in the usual sense of the word, such that each treatment appears the same number of times in each block. In fact, any incomplete block design is “unbalanced” by this common definition.

The BIB and PBIB designs result in all treatments having the same variance (and hence the same standard error). Also, the variances of differences between two treatment means are the same for all pairs of treatments with BIBDs and for sets of treatments with PBIBDs. As you may suspect, it is not possible to construct BIB or PBIB designs for all possible numbers of treatments and blocks. Discovery of numbers of blocks and treatments for which BIBDs and PBIBDs can be constructed was once an active area of statistical research. With the advent of fast computers and good statistical software, the existence of BIBDs and PBIBDs for given numbers of blocks and treatments has become a less important problem. Incomplete block designs without any balance are commonly used in many fields of research. Mead (1988) has an excellent discussion of this issue.

This section presents analyses for a PBIBD using PROC GLM and PROC MIXED to further illustrate some of the similarities and differences between the two procedures. You can see some distinctions between PROC MIXED and PROC GLM that did not occur in the analyses of the RCB design. Although the example is a PBIBD, data analysis methods in this section apply to incomplete block designs in general.

Model

The equation for the model of an incomplete blocks design is the same as for an RCBD. That is, the response Y_{ij} that results from applying treatment i in block j is assumed to be equal to a treatment mean $\mu_i = \mu + \tau_i$ plus a block effect b_j , plus experimental error e_{ij} . Thus the equation

$$Y_{ij} = \mu + \tau_i + b_j + e_{ij}$$

where the block effects b_j are $iid N(0, \sigma_b^2)$, the experimental errors e_{ij} are $iid N(0, \sigma^2)$, the b_j are independent of the e_{ij} , and the indices (i, j) belong to a set of indices indicating which treatments occur in each block. An analysis of variance table for an incomplete blocks design is shown in Table 2.4.

Table 2.4 Type III Analysis of Variance Table for Incomplete Blocks Design

Source of Variation	df	F
Blocks (adjusted for treatments)	$r - 1$	
Treatments (adjusted for blocks)	$t - 1$	$MS(\text{Trts adj.}) / MS(\text{Error})$
Error	$N - r - t + 1$	

In the table, r is the number of blocks, t is the number of treatments, and N is the total number of observations. Notice that the treatments source of variation is adjusted for blocks. The treatments cannot be compared simply on the basis of the usual sum of squared differences between treatment means, because this would contain effects of blocks as well as treatment differences. Instead, a sum of squared differences must be computed between treatment means that have been adjusted to remove the block effects.

Analyses of BIBD and PBIBD data that are presented in most statistics textbooks are called intra-block analyses, because treatments are compared on the basis of differences computed within blocks. You can perform this type of analysis with PROC GLM. It is discussed first in Section 2.5.1. In Section 2.5.2, a mixed model analysis is presented using PROC MIXED that utilizes information about treatment means contained in differences between blocks. This type of analysis combines intra- and inter-block information.

2.5.1 The Usual Intra-block Analysis of PBIB Data Using PROC GLM

Data Set 2.5, “PBIB,” in Appendix 2, “Data Sets,” contains data from Cochran and Cox (1957, p. 456). The design is a PBIBD with fifteen blocks, fifteen treatments, and four treatments per block. Data are pounds of seed cotton per plot. The **block size** is the number of treatments per block. This PBIBD has a block size of four. Each treatment appears in four blocks. Some pairs of treatments appear together in one block (e.g., treatments 1 and 2), and others do not appear together in the same blocks (e.g., treatments 1 and 6).

Program

The data appear in multivariate form in Appendix 2, with one observation per plot. The four treatment identifiers and responses are in separate variables. To arrange the data in the univariate form in which all responses are in a single variable, use the following DATA step:

```
data pbib;
  set pbib_mv; /* data arranged as in Appendix 2 */
  array tx{4} trt1-trt4;
  array yy{4} y1-y4;
  do i=1 to 4;
    treat = tx{i};
    response = yy{i};
    output;
  end;
  keep blk treat response;
run;
```

An intra-block analysis of the PBIBD data is obtained with the following statements:

```
proc glm data=pbib;
  class blk treat;
  model response = blk treat;
  random blk;
  lsmeans treat / stderr pdiff;
  estimate 'treat 1 mean' intercept 1 treat 1;
  estimate 'trt 1 mean' intercept 15 treat 15
            blk 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 / divisor=15;
  estimate 'trt 1 blk 1' intercept 1 treat 1 blk 1;
  estimate 'trt 1 vs trt 2' treat 1 -1;
  contrast 'trt 1 vs trt 2' treat 1 -1;
run;
```

Results

Selected results from this PROC GLM run appear in Output 2.9.

Output 2.9 Incomplete Blocks Design: PROC GLM Analysis

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	28	6.32855556	0.22601984	2.62	0.0050
Error	31	2.67077778	0.08615412		
Corrected Total	59	8.99933333			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
blk	14	3.33422222	0.23815873	2.76	0.0090
treat	14	1.48922222	0.10637302	1.23	0.3012

Source	Type III Expected Mean Square
blk	Var(Error) + 3.2143 Var(blk)
treat	Var(Error) + Q(treat)

Least Squares Means

treat	response LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	2.84555556	0.16342514	<.0001	1
2	2.41277778	0.16342514	<.0001	2
3	2.45166667	0.16342514	<.0001	3
4	2.68333333	0.16342514	<.0001	4
5	2.80666667	0.16342514	<.0001	5
6	2.90388889	0.16342514	<.0001	6
7	2.77111111	0.16342514	<.0001	7
8	2.81000000	0.16342514	<.0001	8
9	2.93333333	0.16342514	<.0001	9
10	2.51500000	0.16342514	<.0001	10
11	2.85388889	0.16342514	<.0001	11
12	3.01277778	0.16342514	<.0001	12
13	2.66833333	0.16342514	<.0001	13
14	2.53333333	0.16342514	<.0001	14
15	2.84833333	0.16342514	<.0001	15

Least Squares Means for effect treat Pr > t for H0: LSMean(i)=LSMean(j)															
		Dependent Variable: response													
i/j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		0.0711	0.0989	0.4887	0.8677	0.8093	0.7500	0.8789	0.7072	0.1634	0.9725	0.4756	0.4498	0.1873	0.9905
2	0.0711		0.8677	0.2515	0.0989	0.0420	0.1450	0.0962	0.0318	0.6619	0.0661	0.0178	0.2782	0.6063	0.0694
3	0.0989	0.8677		0.3248	0.1354	0.0599	0.1776	0.1450	0.0458	0.7863	0.0923	0.0214	0.3729	0.7267	0.0967
4	0.4887	0.2515	0.3248		0.5981	0.3482	0.7072	0.5882	0.3049	0.4727	0.4669	0.1648	0.9488	0.5360	0.4814
5	0.8677	0.0989	0.1354	0.5981		0.6774	0.8789	0.9886	0.5882	0.2328	0.8397	0.3802	0.5545	0.2468	0.8631
6	0.8093	0.0420	0.0599	0.3482	0.6774		0.5705	0.6879	0.8996	0.1031	0.8361	0.6414	0.3168	0.1196	0.8120
7	0.7500	0.1450	0.1776	0.7072	0.8789	0.5705		0.8677	0.4887	0.2772	0.7231	0.3211	0.6602	0.3124	0.7410
8	0.8789	0.0962	0.1450	0.5882	0.9886	0.6879	0.8677		0.5981	0.2121	0.8509	0.3879	0.5587	0.2412	0.8696
9	0.7072	0.0318	0.0458	0.3049	0.5882	0.8996	0.4887	0.5981		0.0805	0.7338	0.7338	0.2612	0.1052	0.7160
10	0.1634	0.6619	0.7863	0.4727	0.2328	0.1031	0.2772	0.2121	0.0805		0.1533	0.0395	0.5127	0.9374	0.1742
11	0.9725	0.0661	0.0923	0.4669	0.8397	0.8361	0.7231	0.8509	0.7338	0.1533		0.4977	0.4290	0.1761	0.9810
12	0.4756	0.0178	0.0214	0.1648	0.3802	0.6414	0.3211	0.3879	0.7338	0.0395	0.4977		0.1469	0.0468	0.4829
13	0.4498	0.2782	0.3729	0.9488	0.5545	0.3168	0.6602	0.5587	0.2612	0.5127	0.4290	0.1469		0.5641	0.4428
14	0.1873	0.6063	0.7267	0.5360	0.2468	0.1196	0.3124	0.2412	0.1052	0.9374	0.1761	0.0468	0.5641		0.1835
15	0.9905	0.0694	0.0967	0.4814	0.8631	0.8120	0.7410	0.8696	0.7160	0.1742	0.9810	0.4829	0.4428		0.1835

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
trt 1 vs trt 2	1	0.30101240	0.30101240	3.49	0.0711

Parameter	Estimate	Standard Error	t Value	Pr > t
treat 1 mean	2.84555556	0.16342514	17.41	<.0001
trt 1 mean	2.84555556	0.16342514	17.41	<.0001
trt 1 blk 1	2.39666667	0.20406165	11.74	<.0001
trt 1 vs trt 2	0.43277778	0.23153188	1.87	0.0711

Interpretation

From the table of Type III sums of squares and the overall analysis of variance in Output 2.9, you can construct the analysis of variance table (Table 2.5).

Table 2.5 Type III Analysis of Variance Table for Incomplete Blocks Design from Cochran and Cox (1957)

Source of Variation	df	SS	MS	F	p
Blocks (adjusted for treatments)	14	3.334	0.238		
Treatments (adjusted for blocks)	14	1.489	0.106	1.23	0.3012
Error	31	2.671	0.086		

The F -test for differences between (adjusted) treatment differences has a significance probability of $p = 0.3012$, which presents no evidence of differences between treatments. In agreement with this conclusion, the table of significance probabilities for the least-squares means shows only six (out of 120) p -values less than 0.05. You would expect that many by chance.

The least-squares means, obtained from the LSMEANS statement, are usually called adjusted means in standard textbooks. In PROC GLM, these means and their standard errors stem from the OLS estimation of the treatment means. Thus, they do not take into account the fact that blocks are random. The adjustment of treatment means to remove block effects is a computation that treats blocks simply as another fixed effect. This analysis of variance, along with adjusted treatment means and differences between them and their standard errors, composes the so-called intra-block analysis of PBIBD data. The intra-block analysis does not use all available information about the treatment effects, and thus it is suboptimal compared to the combined intra- and inter-block estimators provided by PROC MIXED.

Four ESTIMATE statements and one CONTRAST statement are part of the PROC GLM program. The first ESTIMATE statement (labeled “treat1 mean”) specifies coefficients for the INTERCEPT and the TREAT 1 parameter. By default, PROC GLM averages across the BLK parameters to form the linear combination of model terms. The second ESTIMATE statement (labeled “trt 1 mean”) explicitly specifies the same linear combination of model terms; that is, it specifies the coefficients 1/15 for each of the BLK terms. The results from these two ESTIMATE statements are identical. Moreover, the ESTIMATE statements duplicate the least-squares mean for treatment 1 and its standard error. Unfortunately, however, these standard errors do not take into account the random block effects. Section 2.5.2 shows that corresponding ESTIMATE statements in PROC MIXED produce different results. The third ESTIMATE statement provides the prediction of the response for treatment 1 in block 1. The estimated standard error is a little larger than the error for the mean across all blocks. The fourth ESTIMATE statement (labeled “trt 1 vs trt 2”) computes the difference between the least-squares means for treatments 1 and 2. The CONTRAST statement labeled “trt 1 vs trt 2” computes an F -test that is equivalent to the t -test from this ESTIMATE statement.

The RANDOM statement in GLM, as already mentioned, causes only expected mean squares to be computed. The expected mean squares table in Output 2.8 shows that the correct denominator for the F -test for TREAT is MS(Error).

2.5.2 The Combined Intra- and Inter-block Analysis of PBIB Data Using PROC MIXED

When blocks are really treated as random, as in PROC MIXED, the result is the combined intra- and inter-block analysis. You can obtain this analysis with the MIXED procedure by using the following program.

Program

```
proc mixed data=pbib;
  class blk treat;
  model response=treat;
  random blk;
  estimate 'treat 1 mean' intercept 1 treat 1;
  estimate 'trt 1 mean'    intercept 15 treat 15 |
    blk 1 1 1 1 1 1 1 1 1 1 1 1 / divisor=15;
  estimate 'trt 1 blk 1'   intercept 1 treat 1 | blk 1;
  estimate 'trt 1 vs trt 2' treat 1 -1;
  contrast 'trt 1 vs trt 2' treat 1 -1;
```

```

lsmeans treat / diff adjust=tukey;
ods exclude diffs;
ods output diffs=difmix;
run;

```

The ESTIMATE and CONTRAST statements are the same as in the PROC GLM code, except for the second ESTIMATE statement. The BLK effect is a random effect in PROC MIXED, and coefficients for the random effect are separated from coefficients for fixed effects with a vertical bar (|). You will see below that the results for the first two ESTIMATE statements are not identical; they did produce the same results with PROC GLM.

The DIFF option of the LSMEANS statement generates all pairwise comparisons among the 15 treatments, a total of $15 \times 14 / 2 = 60$ comparisons. The listing of the least-squares means differences is suppressed by an ODS EXCLUDE statement. Instead, the corresponding table is written to a data set with the ODS OUTPUT statement for subsequent processing (see program below).

Results

Selected PROC MIXED results appear in Output 2.10 for the combined intra- and inter-block analysis. The differences of least-squares means are addressed later on.

Output 2.10 Incomplete Block Design: PROC MIXED Analysis

Model Information	
Data Set	WORK.PBIB
Dependent Variable	response
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions	
Covariance Parameters	2
Columns in X	16
Columns in Z	15
Subjects	1
Max Obs Per Subject	60

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.04652
Residual	0.08556

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
treat	14	31	1.53	0.1576

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
treat 1 mean	2.8175	0.1664	31	16.93	<.0001
trt 1 mean	2.8175	0.1568	31	17.97	<.0001
trt 1 blk 1	2.5288	0.1845	31	13.70	<.0001
trt 1 vs trt 2	0.4122	0.2221	31	1.86	0.0729

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
trt 1 vs trt 2	1	31	3.45	0.0729

Least Squares Means						
Effect	treat	Estimate	Standard Error	DF	t Value	Pr > t
treat	1	2.8175	0.1664	31	16.93	<.0001
treat	2	2.4053	0.1664	31	14.45	<.0001
treat	3	2.4549	0.1664	31	14.75	<.0001
treat	4	2.7838	0.1664	31	16.73	<.0001
treat	5	2.8049	0.1664	31	16.86	<.0001
treat	6	2.9107	0.1664	31	17.49	<.0001
treat	7	2.7890	0.1664	31	16.76	<.0001
treat	8	2.7816	0.1664	31	16.72	<.0001
treat	9	2.8913	0.1664	31	17.37	<.0001
treat	10	2.4911	0.1664	31	14.97	<.0001
treat	11	2.8987	0.1664	31	17.42	<.0001
treat	12	3.0528	0.1664	31	18.34	<.0001
treat	13	2.6178	0.1664	31	15.73	<.0001
treat	14	2.4913	0.1664	31	14.97	<.0001
treat	15	2.8592	0.1664	31	17.18	<.0001

Interpretation

The “Model Information” table specifies that the variance components are estimated by restricted maximum likelihood (REML). The “Dimensions” table indicates that the fixed effects part of the model (\mathbf{X}) has 16 columns (one for the intercept and 15 for the treatments) and the random effects part of the model (\mathbf{Z} matrix) has 15 columns, one for each block. The REML estimate of the residual variance component is 0.08556, compared to 0.086154 from PROC GLM (Output 2.9). The F -statistic in the “Type 3 Tests of Fixed Effects” table is 1.53 with p -value 0.1576. Compare this to the results from PROC GLM (Output 2.9, $F = 1.23, p = 0.3012$). This smaller p -value in the mixed model analysis is the result of the combined intra- and inter-block estimates of the treatment effects (see Chapter 7 for a discussion of the combining process).

The result from the “Contrasts” table provides a comparison of the means of treatments 1 and 2. The p -value is identical to the p -value comparing treatments 1 and 2 in Output 2.12.

You will note several differences between the PROC MIXED analysis and the intra-block analysis given by PROC GLM. Referring first to the results of the ESTIMATE statement labeled “treat 1 mean,” the estimates of the treatment means are different. Granted, the differences are not major, but they are certainly numerically different. In other applications the distinction can be dramatic. The PROC MIXED estimate of the treatment 1 mean is 2.817, compared with the PROC GLM estimate of 2.846. The PROC GLM estimate of the fixed effects is an ordinary least squares (OLS) estimate, whereas the PROC MIXED estimate is (estimated) generalized least squares (GLS). Theoretically, the GLS estimate is superior. PROC MIXED accounts for BLK being random and computes the best linear unbiased estimate (BLUE) accordingly, substituting estimates of the variance components for unknown parameters in the variance matrix \mathbf{V} . The standard errors in PROC MIXED likewise are different from those in PROC GLM. The standard error of the OLS estimate is 0.163 from GLM. This is not a valid estimate of the true standard error of the OLS estimate, for the same reason that PROC GLM did not compute a valid standard error estimate for a treatment mean for the RCBD data in Section 1.1.1: the random effects of blocks were ignored.

Next, we notice that the standard errors of the estimates labeled “treat 1 mean” and “trt 1 mean” are not the same. Recall that the difference in the ESTIMATE statements was the explicit specification of the BLK coefficients following the vertical bar operator (|) in the second ESTIMATE statement. The ESTIMATE statement with the label “treat 1 mean” did not specify coefficients for the block terms. This made no difference with PROC GLM, but it does make a difference with PROC MIXED. The standard error from the ESTIMATE statement labeled “treat 1 mean” correctly estimates the standard error of the GLS estimate considering blocks to be random. Thus it can be used to produce a confidence interval for the mean that would be valid for inference across the population of blocks from which those in the experiment were randomly drawn. The standard error from the ESTIMATE statement labeled “trt 1 mean,” however, does not involve the block variance component. Thus a confidence interval based on this standard error is valid only for the blocks in the experiment. Standard errors of the least-squares means are the same as for the “treat 1 mean” estimate. The “trt 1 mean” is an example of a **best linear unbiased predictor** (BLUP) and linear combination of fixed and random effects. Similarly, the estimate labeled “trt 1 blk 1” specifies a coefficient for the random block effect. It is a best linear unbiased predictor of the response to treatment 1 of experimental units in block 1. BLUPs are unique to mixed model theory and are discussed in Chapter 6.

The fourth ESTIMATE statement, comparing treatments 1 and 2, also produces different results between PROC MIXED and PROC GLM, because PROC MIXED recovers information about the treatments in the intra- and inter-block analysis.

In the previous PROC MIXED run the differences of the least-squares means were saved to a data set with the ODS OUTPUT statement. We now want to carry out additional processing on these differences. First, a data set (PAIRS) is created that contains the pairs of observations that occur together in a block in this partially balanced incomplete block design.

Program

```

data pairs;
  set pbib_mv;
  array tx{4} trt1-trt4;
  array yy{4} y1-y4;
  do i=1 to 3; do j=(i+1) to 4;
    treat = min(tx{i},tx{j});
    _treat = max(tx{i},tx{j});
    output;
  end; end;
  keep blk treat _treat;
run;
proc sort data=pairs nodupkey; by treat _treat; run;
proc print data=pairs(obs=23); run;

```

The data set of pairs is created from the original data in multivariate format. The variables TREAT and _TREAT are set up to match the variables by the same name in the DIFMIX data set created in the PROC MIXED call.

Results

Output 2.11 shows the first 23 observations of the PAIRS data set. These observations correspond to the pairings of treatments within a block that involve the first two treatments.

Output 2.11 Pairs within a Block Involving Treatments 1 and 2

Obs	blk	treat	_treat
1	3	1	2
2	6	1	3
3	6	1	4
4	2	1	5
5	2	1	7
6	2	1	8
7	1	1	9
8	3	1	10
9	6	1	12
10	1	1	13
11	3	1	14
12	1	1	15
13	4	2	3
14	9	2	4

Obs	blk	treat	_treat
15	9	2	5
16	12	2	6
17	12	2	8
18	12	2	9
19	3	2	10
20	4	2	11
21	9	2	13
22	3	2	14
23	4	2	15

Interpretation

Treatment 1 occurs with all other treatments somewhere in a block, except for treatments 6 and 11. Similarly, treatment 2 appears with all but treatments 7 and 12.

Next, the pairs data set is merged with the differences of the least-squares means to extract the results that correspond to treatment pairs that do not appear in the same plot (disconnected pairs, Output 2.12) and pairs that do appear in the same plot (connected pairs, Output 2.13).

Program

```

proc sort data=difmix; by treat _treat;
data diffs;
    merge difmix pairs; by treat _treat;
run;
proc print data=diffs(where=(blk=..)) noobs;
    var treat _treat estimate stderr probt;
run;
proc print data=diffs(where=((blk ne .) and
                           (treat <=2))) noobs;
    var treat _treat estimate stderr probt;
run;

```

Output 2.12 Least-Squares Means Differences for Disconnected Pairs

treat	_treat	Estimate	StdErr	Probt
1	6	-0.09317	0.2272	0.6846
1	11	-0.08118	0.2272	0.7233
2	7	-0.3837	0.2272	0.1013
2	12	-0.6475	0.2272	0.0077
3	8	-0.3267	0.2272	0.1605
3	13	-0.1628	0.2272	0.4789
4	9	-0.1075	0.2272	0.6395
4	14	0.2925	0.2272	0.2075
5	10	0.3138	0.2272	0.1771
5	15	-0.05434	0.2272	0.8126
6	11	0.01199	0.2272	0.9582

treat	_treat	Estimate	StdErr	Probt
7	12	-0.2638	0.2272	0.2544
8	13	0.1638	0.2272	0.4762
9	14	0.4000	0.2272	0.0882
10	15	-0.3682	0.2272	0.1153

Output 2.13 Least-Squares Means Differences for Connected Pairs Involving Treatments 1 and 2

treat	_treat	Estimate	StdErr	Probt
1	2	0.4122	0.2221	0.0729
1	3	0.3626	0.2221	0.1126
1	4	0.03369	0.2221	0.8804
1	5	0.01262	0.2221	0.9550
1	7	0.02854	0.2221	0.8986
1	8	0.03592	0.2221	0.8726
1	9	-0.07379	0.2221	0.7419
1	10	0.3265	0.2221	0.1516
1	12	-0.2353	0.2221	0.2975
1	13	0.1998	0.2221	0.3753
1	14	0.3262	0.2221	0.1519
1	15	-0.04171	0.2221	0.8522
2	3	-0.04963	0.2221	0.8246
2	4	-0.3785	0.2221	0.0983
2	5	-0.3996	0.2221	0.0817
2	6	-0.5054	0.2221	0.0299
2	8	-0.3763	0.2221	0.1002
2	9	-0.4860	0.2221	0.0363
2	10	-0.08575	0.2221	0.7020
2	11	-0.4934	0.2221	0.0337
2	13	-0.2125	0.2221	0.3461
2	14	-0.08600	0.2221	0.7012
2	15	-0.4539	0.2221	0.0495

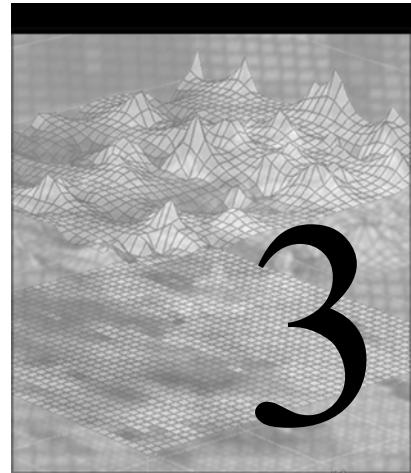
Interpretation

Although only the results for the connected pairs of treatments 1 and 2 are shown in Output 2.13, similar results are obtained for the other treatments. Notice that the estimated standard errors are identical within Output 2.12 and Output 2.13, and that the standard errors are larger in Output 2.12. Although in this example the difference is small, it is an important difference, since it reflects the decreased precision that is the result of disconnected treatment pairs.

Contrasts involving treatments that do not appear in the same block are not estimated with the same precision as contrasts involving treatments that do appear in the same block.

2.6 Summary

Chapter 2 commenced with an example of a randomized blocks design with fixed treatments and random blocks. The importance of accounting for random effects in such a basic situation as computing a variance for a treatment mean was demonstrated. The use of PROC MIXED was introduced with explanations of how to set up the MODEL and RANDOM statements. The chapter continued with illustrations of CONTRAST, ESTIMATE, and LSMEANS statements. Then, PROC GLM was applied to the same example to illustrate similarities with and differences from PROC MIXED. We emphasized the basic applications that are handled correctly by PROC MIXED but not by PROC GLM. A brief explanation of mixed model theory was presented in relation to the randomized blocks design, including explicit descriptions of the matrices in the general linear mixed model. Then, an incomplete block design was used in Section 2.5 to illustrate some of the problems presented by unbalanced mixed model data.



Random Effects Models

3.1	Introduction: Descriptions of Random Effects Models	58
3.1.1	Using PROC MIXED to Estimate the Variance Components.....	61
3.1.2	The Mixed Model Equations	62
3.1.3	Method of Moments Estimators Using PROC MIXED	63
3.2	Example: One-Way Random Effects Treatment Structure	64
3.2.1	Using the MIXED Procedure	64
3.2.2	Using the GLM Procedure.....	71
3.2.3	Confidence Intervals about the Variance Components.....	73
3.3	Example: A Simple Conditional Hierarchical Linear Model.....	75
3.3.1	Using PROC MIXED to Analyze the Data	76
3.3.2	PROC GLM Part of the Analysis	80
3.4	Example: Three-Level Nested Design Structure	81
3.4.1	Three-Level Nested Linear Model, an Unconditional Hierarchical Nested Linear Model	81
3.4.2	Data Analysis Using the PROC MIXED METHOD=REML to Estimate the Variance Components	82
3.4.3	Using the PROC MIXED METHOD=TYPE1 to Estimate the Variance Components	83
3.4.4	Conditional Hierarchical Linear Model or Mixed Model	84
3.4.5	Using the MIXED Procedure	84
3.4.6	Unequal Variance Model	85
3.5	Example: A Two-Way Random Effects Treatment Structure to Estimate Heritability	88
3.5.1	Using PROC MIXED METHOD=REML.....	88
3.5.2	Using PROC MIXED METHOD=TYPE3.....	90
3.6	Summary	91

3.1 Introduction: Descriptions of Random Effects Models

Random effects models are linear models where all of the factors in the models are random variables. Random effects models are used in studies and experiments where the levels of the factors have been selected at random from a population of possible levels and where you want to obtain information about the parameters of the distribution of those levels. The main goals of the analysis of random effects models are to do the following:

- estimate the parameters of the covariance structure of the random effects model
- test hypotheses about the parameters or functions of the parameters
- construct confidence intervals about the parameters or functions of the parameters

This chapter presents methods available in PROC MIXED for estimating parameters of random effects models. It also compares the information available from PROC GLM to that which can be obtained from PROC MIXED. Procedures for **testing hypotheses** and for constructing **confidence intervals** are presented for some simple models. Extensions of the procedures are used for some of the complex examples described in the final sections of the chapter.

The random effects model where the levels of the random effects and/or the experimental units form a **nested hierarchical** structure is also called the **unconditional hierarchical linear model**, as defined by Bryk and Raudenbush (1992). Models can be constructed to describe part of the variance between the levels of a random effect either by classifying the levels of a random effect into categories or by using continuous variables in the analysis of covariance context (see Chapter 7). **Conditional hierarchical linear models** are linear models that attempt to describe part of the variance between the levels of a random effect either by classifying the levels of a random effect into categories or by using continuous variables in the analysis of covariance context. The categorical variables and the continuous variables (called covariates) are considered to be fixed effects. Thus conditional hierarchical linear models are mixed models. We describe several conditional hierarchical linear models in this chapter, while additional models are discussed in other chapters.

Random effects models are applied to experiments where all of the factors in the treatment structure of the experimental design or study are random effects. The factors in the design structure are always considered as random effects (see Milliken and Johnson 1992, Chapter 4, for discussions of design and treatment structures). A factor is called a **random effect** if the levels of that factor selected to be included in the study or experiment are randomly selected from a population of possible levels of that factor.

Suppose you are studying the source of nitrogen contamination in the Mississippi River at New Orleans. You could go out and obtain water samples from every source of influent that eventually reaches the Mississippi River. There are hundreds of influents into the river, and it would be very expensive and time-consuming to sample every one of them (also, you most likely would want to sample them over time). Another strategy would be to identify all of the influents into the river (specify the population of influents) and then randomly select, say, 100 of the influents to be actually measured. In this case, we assume that the concentration of nitrogen in the water from influents in the population forms a distribution of nitrogen levels with mean μ_N and variance σ^2_{Infl} . By randomly selecting influents to be sampled, we use the

sample mean and variance to provide estimates of the population parameters. A model to describe the observed nitrogen concentration from the i^{th} sample is as follows:

$$Y_i = \mu_N + a_i, i = 1, 2, \dots, s, \text{ where } a_i \sim \text{iid } N(0, \sigma^2_{\text{Infl}}) \quad (3.1)$$

Model (3.1) is that of a single sample of size s obtained from a population with parameters μ_N and σ^2_{Infl} . This is the type of data set you may have encountered in your first course in statistical methods, but it most likely was not represented as a model. Model (3.1) is a mixed model where μ_N is the fixed effects part and a_i is the random effects part of the model. μ_N denotes the intercept or overall mean of the model. The estimates of the parameters can be obtained with PROC MIXED using the following statements:

```
proc mixed;
  model y = /solution;
run;
```

The MODEL statement above includes an intercept (unless you specify the NOINT option in the MODEL statement) that provides the estimate of μ_N . The estimate of σ^2_{Infl} is computed from the residuals of the model, which are the deviations of the observations from the estimate of μ_N . (Using PROC MIXED in this context is like using a sledgehammer to drive a carpet tack, but the chapter uses this example to begin the introduction to using PROC MIXED for random effects models.) The normality assumption is made to provide a distributional basis for testing hypotheses and for constructing confidence intervals about the parameters of the model. The **best linear unbiased estimate** (BLUE) of μ_N is \bar{y}_\square , which has sampling distribution $\bar{y}_\square \sim N(\mu_N, s^{-1} \sigma^2_{\text{Infl}})$. The **best quadratic unbiased estimate** of σ^2_{Infl} is

$$\hat{\sigma}_{\text{Infl}}^2 = \sum_{i=1}^s (y_i - \bar{y}_\square)^2 / (s-1)$$

The sampling distribution associated with $\hat{\sigma}_{\text{Infl}}^2$ is $(s-1)\hat{\sigma}_{\text{Infl}}^2 / \sigma^2_{\text{Infl}} \sim \chi^2_{(s-1)}$. Tests of hypotheses and confidence intervals are computed as described in any elementary or basic text on statistics. We include the **confidence intervals** for completeness. A $(1-\alpha)100\%$ confidence interval about μ_N is $\bar{y}_\square \pm (t_{\alpha/2, (s-1)})\hat{\sigma}_{\text{Infl}} / \sqrt{s}$. A $(1-\alpha)100\%$ confidence interval about σ^2_{Infl} is

$$\frac{(s-1)\hat{\sigma}_{\text{Infl}}^2}{\chi^2_{\alpha/2, (s-1)}} < \sigma^2_{\text{Infl}} < \frac{(s-1)\hat{\sigma}_{\text{Infl}}^2}{\chi^2_{1-\alpha/2, (s-1)}}$$

In our scenario, the environmentalist most likely wants to obtain information from multiple sites from each selected influent. In this case suppose that n_i sites were randomly selected from the population of possible sites within each randomly selected influent. The number of sites could vary from influent to influent, possibly depending on the size of the influent, as one may wish to select the same proportion of the possible sites from each selected influent. Let Y_{ij} denote the nitrogen concentration in the water sample taken from the j^{th} site at the i^{th} influent. A model to describe the collection of measurements is

$$Y_{ij} = \mu_N + a_i + e_{ij}, \quad i = 1, 2, \dots, s, \quad j = 1, 2, \dots, n_i \quad (3.2)$$

where

$$\begin{aligned} a_i &\sim iid N(0, \sigma^2_{Infl}) \\ e_{ij} &\sim iid N(0, \sigma^2_{site}) \end{aligned}$$

Model (3.2) has two variance component parameters. The variance component σ^2_{Infl} measures the influent-to-influent variability as in model (3.1). The variance component σ^2_{site} measures the site-to-site variability within an influent. Model (3.2) is a mixed effects model where μ_N is the **fixed effects part** of the model, a_i is the **random effects part** of the model, and e_{ij} is the **residual part** of the model. Model (3.2) is also called a **two-level hierarchical linear model** (Bryk and Raudenbush 1992) because the sampled sites are nested within each influent. The PROC MIXED statements to fit model (3.2) are as follows:

```
proc mixed;
  class influent;
  model y = ;
  random influent;
run;
```

Model (3.2) is the usual one-way random-effects treatment structure model where the levels of the treatment are the randomly selected influents. If you apply usual analysis of variance to compute sums of squares and to evaluate their expected means squares, you would obtain the results displayed in Table 3.1. The information in Table 3.1 can be obtained by using PROC MIXED to analyze the data by specifying METHOD=TYPE1 as in the following program.

```
proc mixed method=type1;
  class influent;
  model y = ;
  random influent;
run;
```

The TYPE1 analysis produces the analysis of variance table with sums of squares, degrees of freedom, mean squares, and expected mean squares. The expected mean squares are used to determine the appropriate divisor for F -statistics of the effects in the analysis.

Table 3.1 Analysis of Variance Table for a One-Way Random Effects Treatment Structure

Source	DF	Sum of Squares	Expected Mean Squares
Influents	$s - 1$	$\sum_{i=1}^s n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$\sigma^2_{site} + C\sigma^2_{Infl}$
Sites (Influents)	$\sum_{i=1}^s (n_i - 1) = n_{\cdot} - s$	$\sum_{i=1}^s \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$	σ^2_{site}

The coefficient C is computed as

$$C = \frac{n_{\square} - \frac{\sum_{i=1}^s n_i^2}{n_{\square}}}{s-1}$$

(See Chapter 18 of Milliken and Johnson 1992.) **Method of moments** estimators of the variance components are obtained by equating the observed mean squares to the expected mean squares replacing the variances with estimators. The **method of moments equations** are

$$\begin{aligned} MSI &= \frac{\sum_{i=1}^t n_i (\bar{y}_{i\square} - \bar{y}_{\square\square})^2}{s-1} = \hat{\sigma}_{site}^2 + C \hat{\sigma}_{Infl}^2 \\ MSsite(I) &= \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\square})^2}{n_{\square} - s} = \hat{\sigma}_{site}^2 \end{aligned} \quad (3.3)$$

The solution to the method of moment equations is

$$\begin{aligned} \tilde{\sigma}_{Infl}^2 &= \frac{MSI - MSsite(I)}{C} \\ \tilde{\sigma}_{site}^2 &= MSsite(I) \end{aligned} \quad (3.4)$$

The method of moments estimators are

$$\begin{aligned} \hat{\sigma}_{site}^2 &= \tilde{\sigma}_{site}^2 \\ \hat{\sigma}_{Infl}^2 &= \begin{cases} \tilde{\sigma}_{Infl}^2 & \text{if } \tilde{\sigma}_{Infl}^2 > 0 \\ 0 & \text{if } \tilde{\sigma}_{Infl}^2 \leq 0 \end{cases} \end{aligned} \quad (3.5)$$

The process of solving the method of moments equations does not guarantee that the solution is in the parameter space, and thus the solution can yield negative values for the variance components. But when that occurs, the estimate is set to zero as shown in equation (3.5).

3.1.1 Using PROC MIXED to Estimate the Variance Components

PROC MIXED obtains six types of estimators of the variance components. **Restricted maximum likelihood (REML)** estimators and **maximum likelihood (ML)** estimators are based on the normality assumptions stated in model (3.2). The third procedure is **MIVQUE(0)**, which provides estimates that are a form of method of moments estimator. The other three are methods of moments estimators based on Type 1, Type 2, or Type 3 sums of squares. The estimators obtained with the REML, MIVQUE0, TYPE1, TYPE2, and TYPE3 methods are identical for balanced data sets when the solution to the method of moments equations is in the parameter space. All of the estimators can be different when the data are unbalanced. Method of moments estimators are unbiased estimators under the assumption that random effects and errors are independently distributed. When you assume that the random effects are normally distributed, then REML and ML estimators possess the usual large sample properties of maximum

likelihood estimators. The MIVQUE0 estimators are minimum variance within the class of quadratic unbiased estimators without the assumption of normality. The likelihood estimators are obtained by maximizing the likelihood (either full or residual) function over the parameter space (see Appendix 1 for details).

For REML and ML estimates, PROC MIXED provides estimates of the standard errors of the estimates of the variance components that are computed from the inverse of the estimated information matrix. The z -score computed as the ratio of the estimator to the corresponding estimate of the standard error is available to test the hypothesis $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$. The z -score is valid only when the sampling distribution of $\hat{\sigma}^2_{Infl}$ can be approximated by a normal distribution. But this approximation is not appropriate when the number of levels of the random treatment is small, as is the case in this example. When the number of levels is large, the z -score can be used to test $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$ (the alternative hypothesis is one-sided).

The -2 REML Log Likelihood value is obtained by evaluating the likelihood at the selected estimators. If you want to compare two models for which the random effects of one model are a subset of the random effects of the other, you can formally test whether the additional parameters are zero with a likelihood ratio test defined on the respective -2 REML Log Likelihoods of the two models. This is discussed in Appendix 1.

3.1.2 The Mixed Model Equations

The fixed effects part of model (3.2) is μ_N and is estimated by solving the mixed models equations (see Appendix 1). To review briefly, the general mixed model can be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where

\mathbf{Y} is the data vector

$\boldsymbol{\beta}$ is the coefficient vector corresponding to the fixed effects

\mathbf{X} is the design matrix for the fixed effects

\mathbf{u} is the coefficient vector corresponding to the random effects

\mathbf{Z} is the design matrix for the random effects part of the model

\mathbf{e} is the error vector

In this chapter it is assumed that \mathbf{u} and \mathbf{e} are uncorrelated random variables with zero means and covariance matrices \mathbf{G} and \mathbf{R} , respectively; thus, the covariance matrix of the data vector is $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ (see Chapter 6 for more details). The solution of the mixed model equations for $\boldsymbol{\beta}$ and \mathbf{u} is as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} &= \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned}$$

In PROC MIXED, the SOLUTION option in the MODEL statement requests a listing of $\hat{\beta}$, and the SOLUTION option in the RANDOM statement requests a listing of $\hat{\mathbf{u}}$. Estimates of estimable linear combinations of β added to linear combinations of \mathbf{u} provide predicted values of predictable functions. For model (3.2), the matrices above are $\beta = \mu_N \mathbf{j}_n$, $\mathbf{X} = \mathbf{j}_s \otimes \mathbf{j}_{ni}$, $\mathbf{u} = \mathbf{a}$, $\mathbf{Z} = \mathbf{I}_s \otimes \mathbf{j}_{ni}$, $\mathbf{G} = \sigma_a^2 \mathbf{I}_s$, $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$, and \mathbf{e} is the random error. The notation $\mathbf{A} \otimes \mathbf{B}$ denotes the *right* direct product of matrices \mathbf{A} and \mathbf{B} , \mathbf{I}_t denotes a $t \times t$ identity matrix, and \mathbf{j}_n denotes a $n \times 1$ vector of ones.

3.1.3 Method of Moments Estimators Using PROC MIXED

The solution to the method of moments of equations can be obtained with PROC MIXED by specifying either Type1, Type2, or Type3 as the estimation method for the variance components. The following program uses METHOD=TYPE3:

```
proc mixed method=type3 covtest cl;
  class influent;
  model y = ;
  random influent;
run;
```

A $(1-\alpha)100\%$ confidence interval about σ_{site}^2 is

$$\frac{(n_s - s)\hat{\sigma}_{site}^2}{\chi_{\alpha/2, (n_s - t)}^2} < \sigma_{site}^2 < \frac{(n_s - s)\hat{\sigma}_{site}^2}{\chi_{1-\alpha/2, (n_s - t)}^2}$$

An approximate $(1-\alpha)100\%$ confidence interval about σ_{Infl}^2 is

$$\frac{v\hat{\sigma}_{Infl}^2}{\chi_{\alpha/2, v}^2} < \sigma_{Infl}^2 < \frac{v\hat{\sigma}_{Infl}^2}{\chi_{1-\alpha/2, v}^2}$$

where v are degrees of freedom determined by using the Satterthwaite approximation. In this case, the approximate degrees of freedom are calculated as

$$v = \frac{(\hat{\sigma}_{Infl}^2)^2}{\frac{\left[\frac{MSI}{C}\right]^2}{s-1} + \frac{\left[\frac{MSsite(I)}{C}\right]^2}{n_s - s}}$$

The Satterthwaite approximation provides an approximation to the sampling distribution of linear combinations of mean squares. In this case $\hat{\sigma}_{Infl}^2 = MSI/C - MSsite(I)/C$, and the approximating sampling distribution of $v\hat{\sigma}_{Infl}^2 / \sigma_{Infl}^2$ is χ_v^2 . As the number of influents increases, the sampling distribution of $\hat{\sigma}_{Infl}^2$ can be approximated by a normal distribution. When the number of influents is not large, the normal approximation will not be appropriate, as it is a symmetric distribution and the estimates of variances generally have a right-skewed sampling distribution.

Another statistic to test the hypothesis $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$ is

$$F_c = \frac{MSI}{MSsite(I)}$$

where the decision rule is to reject H_0 if and only if $F_c > F_{\alpha,(t-1),(n_t-t)}$. The sampling distribution of F_c is exact when the random effects are independently distributed as normal random variables. This test statistic is preferred over the z -score when the number of degrees of freedom associated with a random effect is small or when the test is exact. In reality we know that σ^2_{Infl} cannot be equal to 0, but σ^2_{Infl} may be small enough compared to σ^2_{site} to be considered negligible.

3.2 Example: One-Way Random Effects Treatment Structure

The data in Data Set 3.2, “Mississippi River,” in Appendix 2, “Data Sets,” are the nitrogen concentrations in parts per million from several sites at six of the randomly selected influents to the Mississippi River (you would want to select many more than six sites to monitor the Mississippi River, but we use only six for demonstration purposes). The model in equation (3.2) is used to describe the process, and we want to estimate μ_N , σ^2_{Infl} , and σ^2_{site} . The model is also called a two-level hierarchical linear model. The method of moments estimates of the variance components are obtained with PROC MIXED.

3.2.1 Using the MIXED Procedure

PROC MIXED is used to compute the estimates of the variance components and the population mean, predicted values for each level of influent, and predicted values for the deviations of the influent effects from the population mean using all of the methods available for estimating variance components. The PROC MIXED program to provide REML estimates of the variance components, etc., is as follows:

```
proc mixed data=influent covtest cl;
  class influent;
  model y = /solution;
  random influent / solution;

  estimate 'influent 1' intercept 1 | influent 1 0 0 0 0 0;
  estimate 'influent 2' intercept 1 | influent 0 1 0 0 0 0;
  estimate 'influent 3' intercept 1 | influent 0 0 1 0 0 0;
  estimate 'influent 4' intercept 1 | influent 0 0 0 1 0 0;
  estimate 'influent 5' intercept 1 | influent 0 0 0 0 1 0;
  estimate 'influent 6' intercept 1 | influent 0 0 0 0 0 1;

  estimate 'influent 1U' | influent 1 0 0 0 0 0;
  estimate 'influent 2U' | influent 0 1 0 0 0 0;
  estimate 'influent 3U' | influent 0 0 1 0 0 0;
  estimate 'influent 4U' | influent 0 0 0 1 0 0;
  estimate 'influent 5U' | influent 0 0 0 0 1 0;
  estimate 'influent 6U' | influent 0 0 0 0 0 1;

run;
```

The SOLUTION option in the MODEL statement requests the estimates of the fixed effects, which in this case is just the population mean and corresponds to the INTERCEPT. The SOLUTION option in the RANDOM statement provides predicted values of the random effects with expectation zero, which are listed as a solution for the random effect. The COVTEST option provides the estimates of the standard errors of the variance components, *z*-scores, and associated *p*-values. The CL option provides 95% confidence intervals about the variance components where the level of confidence can be controlled by the option ALPHA=xx, which yields (1-2xx)100% confidence intervals. The results are displayed in Output 3.1.

ESTIMATE statements are used to compute predictions of predictable functions. The first set of ESTIMATE statements provides predicted values for the concentrations of nitrogen levels at each influent. These predictions are the best linear unbiased predictors (BLUP) of the nitrogen level at each influent (see Chapter 6 for a detailed discussion of BLUP). The second set of ESTIMATE statements requests predictions of the deviations from the overall mean for each influent. This set of ESTIMATE statements is included to simply show that you can obtain the solutions for the random effects through ESTIMATE statements. To obtain predictions involving the random effects, the random effects coefficients are listed after the vertical bar (|) in the ESTIMATE statement.

Results

The results from this PROC MIXED run are shown in Output 3.1.

Output 3.1 PROC MIXED Results: REML Estimates of the Variance Components and Estimated BLUP for Each INFLUENT

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
influent	63.3211	45.2315	1.40	0.0808	0.05	22.5560	539.98
Residual	42.6583	10.8571	3.93	<.0001	0.05	27.3960	75.4984

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	21.2231	3.4290	5	6.19	0.0016

Solution for Random Effects						
Effect	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent	1	0.3093	3.8193	31	0.08	0.9360
influent	2	-6.7193	3.9170	31	-1.72	0.0963
influent	3	-3.8979	4.0805	31	-0.96	0.3468
influent	4	2.9461	3.9870	31	0.74	0.4655
influent	5	-6.0130	4.0805	31	-1.47	0.1507
influent	6	13.3748	4.0805	31	3.28	0.0026

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 1	21.5324	2.1135	5	10.19	0.0002
influent 2	14.5038	2.3769	5	6.10	0.0017
influent 3	17.3252	2.7721	5	6.25	0.0015
influent 4	24.1692	2.5518	5	9.47	0.0002
influent 5	15.2102	2.7721	5	5.49	0.0027
influent 6	34.5979	2.7721	5	12.48	<.0001
influent 1U	0.3093	3.8193	31	0.08	0.9360
influent 2U	-6.7193	3.9170	31	-1.72	0.0963
influent 3U	-3.8979	4.0805	31	-0.96	0.3468
influent 4U	2.9461	3.9870	31	0.74	0.4655
influent 5U	-6.0130	4.0805	31	-1.47	0.1507
influent 6U	13.3748	4.0805	31	3.28	0.0026

For comparison purposes, the results using ML, MIVQUE(0), and method of moments estimates of the variance components follow. The results are similar to those in Output 3.1 for REML estimates of the variance components. The ESTIMATE statements included in the first PROC MIXED run are not included in the program listing for brevity. The respective sets of programs are as follows.

Program

For ML:

```
proc mixed data=influent method=ml;
  class influent;
  model y = / solution;
  random influent / solution;
run;
```

For MIVQUE(0):

```
proc mixed data=influent method=mivque0;
  class influent;
  model y = / solution;
  random influent / solution;
run;
```

For TYPE1:

```
proc mixed data=influent method=type1;
  class influent;
  model y = / solution;
  random influent / solution;
run;
```

Results

The results of these sets of code are shown in Outputs 3.2–3.4.

Output 3.2 PROC MIXED Results: ML Estimates of the Variance Components and Estimated BLUP for Each INFLUENT

Covariance Parameter Estimates	
Cov Parm	Estimate
influent	51.2509
Residual	42.6979

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	21.2171	3.1219	5	6.80	0.0010

Solution for Random Effects						
Effect	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent	1	0.3098	3.5365	31	0.09	0.9308
influent	2	-6.5772	3.6379	31	-1.81	0.0803
influent	3	-3.7862	3.8054	31	-0.99	0.3275
influent	4	2.8826	3.7099	31	0.78	0.4430
influent	5	-5.8434	3.8054	31	-1.54	0.1348
influent	6	13.0144	3.8054	31	3.42	0.0018

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 1	21.5269	2.1005	5	10.25	0.0002
influent 2	14.6399	2.3582	5	6.21	0.0016
influent 3	17.4309	2.7420	5	6.36	0.0014
influent 4	24.0997	2.5285	5	9.53	0.0002
influent 5	15.3737	2.7420	5	5.61	0.0025
influent 6	34.2315	2.7420	5	12.48	<.0001
influent 1U	0.3098	3.5365	31	0.09	0.9308
influent 2U	-6.5772	3.6379	31	-1.81	0.0803
influent 3U	-3.7862	3.8054	31	-0.99	0.3275
influent 4U	2.8826	3.7099	31	0.78	0.4430
influent 5U	-5.8434	3.8054	31	-1.54	0.1348
influent 6U	13.0144	3.8054	31	3.42	0.0018

Output 3.3 PROC MIXED Results: MIVQUE(0) Estimates of the Variance Components and Estimated BLUP for Each INFLUENT

Covariance Parameter Estimates	
Cov Parm	Estimate
influent	45.7541
Residual	51.3914

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	21.2068	3.0122	5	7.04	0.0009

Solution for Random Effects						
Effect	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent	1	0.3100	3.4997	31	0.09	0.9300
influent	2	-6.3334	3.6144	31	-1.75	0.0896
influent	3	-3.5985	3.8004	31	-0.95	0.3510
influent	4	2.7739	3.6949	31	0.75	0.4585
influent	5	-5.5582	3.8004	31	-1.46	0.1537
influent	6	12.4062	3.8004	31	3.26	0.0027

Estimates						
Label	Estimate	Standard Error	DF	t Value	Pr > t	
influent 1	21.5169	2.2778	5	9.45	0.0002	
influent 2	14.8734	2.5495	5	5.83	0.0021	
influent 3	17.6084	2.9493	5	5.97	0.0019	
influent 4	23.9807	2.7277	5	8.79	0.0003	
influent 5	15.6486	2.9493	5	5.31	0.0032	
influent 6	33.6130	2.9493	5	11.40	<.0001	
influent 1U	0.3100	3.4997	31	0.09	0.9300	
influent 2U	-6.3334	3.6144	31	-1.75	0.0896	
influent 3U	-3.5985	3.8004	31	-0.95	0.3510	
influent 4U	2.7739	3.6949	31	0.75	0.4585	
influent 5U	-5.5582	3.8004	31	-1.46	0.1537	
influent 6U	12.4062	3.8004	31	3.26	0.0027	

Output 3.4 PROC MIXED Results: Methods of Moments Estimates of Variance Components and Estimated BLUP for Each INFLUENT

Type 1 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
influent	5	1925.193608	385.038722	Var(Residual) + 6.0973 Var(influent)		MS(Residual)	31	9.04 <.0001
Residual	31	1319.779365	42.573528	Var(Residual)		.	.	.

Covariance Parameter Estimates	
Cov Parm	Estimate
influent	56.1667
Residual	42.5735

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	21.2199	3.2500	5	6.53	0.0013

Solution for Random Effects						
Effect	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent	1	0.3096	3.6535	31	0.08	0.9330
influent	2	-6.6434	3.7533	31	-1.77	0.0866
influent	3	-3.8381	3.9190	31	-0.98	0.3350
influent	4	2.9122	3.8244	31	0.76	0.4521
influent	5	-5.9221	3.9190	31	-1.51	0.1409
influent	6	13.1818	3.9190	31	3.36	0.0021

Estimates						
Label	Estimate	Standard Error	DF	t Value	Pr > t	
influent 1	21.5295	2.1040	5	10.23	0.0002	
influent 2	14.5765	2.3640	5	6.17	0.0016	
influent 3	17.3818	2.7526	5	6.31	0.0015	
influent 4	24.1321	2.5363	5	9.51	0.0002	
influent 5	15.2978	2.7526	5	5.56	0.0026	
influent 6	34.4017	2.7526	5	12.50	<.0001	
influent 1U	0.3096	3.6535	31	0.08	0.9330	
influent 2U	-6.6434	3.7533	31	-1.77	0.0866	
influent 3U	-3.8381	3.9190	31	-0.98	0.3350	

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 4U	2.9122	3.8244	31	0.76	0.4521
influent 5U	-5.9221	3.9190	31	-1.51	0.1409
influent 6U	13.1818	3.9190	31	3.36	0.0021

Interpretation

You should notice how the values of the predictable functions change as the magnitudes of the variance components change. The estimates of the variance components are somewhat variable, but the predictions of each influent's effect do not change very much. Only the TYPE1 sum of squares method is used here, as all of the sums of squares methods provide the same estimators. The choice among these four analyses depends on your choice of technique to estimate the variance components. Because none of the variance component estimation techniques can be shown to be superior to the others, there is no clear choice. The REML estimates seem to be used more frequently than the estimates from the other techniques.

PROC MIXED can be used to evaluate a given data set using estimates of variance components obtained from another study or a given set of hypothesized variance components. The NOPROFILE option in the PROC MIXED statement together with the NOITER option in the PARMS statement forces PROC MIXED to use these values without modification by iteration.

Program

```
proc mixed data=influent noprofile;
  class influent;
  model y = / solution;
  random influent / solution;
  parms (70) (25) / noiter;
run;
```

Results

The results are given in Output 3.5.

Output 3.5 PROC MIXED Results for Analysis with Fixed Values of the Variance Components and Estimated BLUP for Each INFLUENT

Covariance Parameter Estimates	
Cov Parm	Estimate
influent	70.0000
Residual	25.0000

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	21.2360	3.5176	5	6.04	0.0018

Solution for Random Effects						
Effect	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent	1	0.3074	3.7575	31	0.08	0.9353
influent	2	-7.0206	3.8210	31	-1.84	0.0758
influent	3	-4.1402	3.9301	31	-1.05	0.3003
influent	4	3.0807	3.8672	31	0.80	0.4317
influent	5	-6.3802	3.9301	31	-1.62	0.1146
influent	6	14.1531	3.9301	31	3.60	0.0011

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 1	21.5434	1.6401	5	13.14	<.0001
influent 2	14.2153	1.8513	5	7.68	0.0006
influent 3	17.0957	2.1729	5	7.87	0.0005
influent 4	24.3166	1.9929	5	12.20	<.0001
influent 5	14.8557	2.1729	5	6.84	0.0010
influent 6	35.3891	2.1729	5	16.29	<.0001
influent 1U	0.3074	3.7575	31	0.08	0.9353
influent 2U	-7.0206	3.8210	31	-1.84	0.0758
influent 3U	-4.1402	3.9301	31	-1.05	0.3003
influent 4U	3.0807	3.8672	31	0.80	0.4317
influent 5U	-6.3802	3.9301	31	-1.62	0.1146
influent 6U	14.1531	3.9301	31	3.60	0.0011

Interpretation

The predicted values of the influent means are influenced by the estimates of the variance components. In addition, the estimate of μ_N is also influenced by the method used to estimate the variance components.

3.2.2 Using the GLM Procedure

PROC GLM can be used to compute the usual analysis of the random effects model that involves the computation of mean squares, expected mean squares, and statistics to test hypotheses about the importance of the individual variance components. PROC GLM does not provide estimates of the variance components. You can, however, use the mean squares and expected mean squares to set up and solve a set of equations to obtain the method of moments estimates. The PROC GLM statements to compute the mean squares and expected mean squares and to test the hypothesis $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$ are as follows.

Program

```

proc glm data=influent;
  class influent;
  model y = influent;
  random influent / test;

  estimate 'Mean'      intercept 6 influent 1 1 1 1 1 1 / divisor=6;
  estimate 'Influent 1' intercept 6 influent 6 0 0 0 0 0 / divisor=6;
  estimate 'Influent 2' intercept 6 influent 0 6 0 0 0 0 / divisor=6;
  estimate 'Influent 3' intercept 6 influent 0 0 6 0 0 0 / divisor=6;
  estimate 'Influent 4' intercept 6 influent 0 0 0 6 0 0 / divisor=6;
  estimate 'Influent 5' intercept 6 influent 0 0 0 0 6 0 / divisor=6;
  estimate 'Influent 6' intercept 6 influent 0 0 0 0 0 6 / divisor=6;

  estimate 'Influent 1U' influent 5 -1 -1 -1 -1 -1 / divisor=6;
  estimate 'Influent 1U' influent -1 5 -1 -1 -1 -1 / divisor=6;
  estimate 'Influent 1U' influent -1 -1 5 -1 -1 -1 / divisor=6;
  estimate 'Influent 1U' influent -1 -1 -1 5 -1 -1 / divisor=6;
  estimate 'Influent 1U' influent -1 -1 -1 -1 5 -1 / divisor=6;
  estimate 'Influent 1U' influent -1 -1 -1 -1 -1 5 / divisor=6;
run;

```

The results of the statements above are shown in Output 3.6, where the set of estimate statements provides an estimate of the overall mean, predictions for each influent, and predictions for the influents that sum to zero.

Results

Output 3.6 PROC GLM Results: Expected Mean Squares and Test of INFLUENT Variance Component against Zero

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1925.193608	385.038722	9.04	<.0001
Error	31	1319.779365	42.573528		
Corrected Total	36	3244.972973			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
influent	5	1925.193608	385.038722	9.04	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
influent	5	1925.193608	385.038722	9.04	<.0001

Source	Type III Expected Mean Square
influent	Var(Error) + 6.0973 Var(influent)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
influent	5	1925.193608	385.038722	9.04	<.0001
Error: MS(Error)	31	1319.779365	42.573528		

Parameter	Estimate	Standard Error	t Value	Pr > t
Mean	21.2521164	1.09863589	19.34	<.0001
Influent 1	21.5555556	2.17494643	9.91	<.0001
Influent 2	13.8571429	2.46615745	5.62	<.0001
Influent 3	16.8000000	2.91799684	5.76	<.0001
Influent 4	24.5000000	2.66375449	9.20	<.0001
Influent 5	14.4000000	2.91799684	4.93	<.0001
Influent 6	36.4000000	2.91799684	12.47	<.0001
Influent 1U	0.3641270	2.50584466	0.15	0.8854
Influent 1U	-8.8739683	2.75258722	-3.22	0.0030
Influent 1U	-5.3425397	3.14836442	-1.70	0.0997
Influent 1U	3.8974603	2.92401191	1.33	0.1923
Influent 1U	-8.2225397	3.14836442	-2.61	0.0138
Influent 1U	18.1774603	3.14836442	5.77	<.0001

Interpretation

The F -value of 9.04 has a p -value of less than 0.0001, which indicates strong evidence that the null hypothesis is not true. Estimates of the variance components are not computed by PROC GLM, though you could carry out the computations yourself as the mean squares and the expected mean squares are provided.

3.2.3 Confidence Intervals about the Variance Components

The final step in the analysis is to construct a 95% confidence interval about σ^2_{Infl} . There are two methods easily applied to the construction of confidence intervals about variance components. PROC MIXED provides estimates of the standard errors associated with each estimate of a variance component. Large-sample normal approximation confidence intervals can be constructed using that information. The second method is to use the Satterthwaite approximation, described in the discussion following equation (3.4). The Satterthwaite confidence intervals are asymmetric about the estimate of the variance component, whereas the normal approximation confidence intervals are symmetric about the estimate of the variance component.

An asymptotic 95% confidence interval about σ^2_{Infl} can be computed using the estimated standard error of $\hat{\sigma}^2_{Infl}$. An asymptotic 95% confidence interval about σ^2_{Infl} , computed using the estimated standard error of $\hat{\sigma}^2_{Infl}$ from Output 3.1, is

$$\hat{\sigma}^2_{Infl} \pm z_{0.025} s\hat{e}_{\hat{\sigma}^2_{Infl}} \text{ or } 63.321 \pm 1.96 \times 45.231 \text{ or} \\ 63.321 \pm 88.652 \text{ or } 0 < \sigma^2_{Infl} < 151.973$$

The lower limit is truncated to zero because σ^2_{Infl} is a nonnegative parameter.

Program

The DATA step to compute the confidence interval based on the Satterthwaite approximation follows:

```

data satt;
  /* c = coefficient of var(influent) in e(ms influent) */
  c      = 6.0973;
  mssite = 1319.77936508/31;   * ms error;
  msi    = 1925.19360789/5;    * ms influent;
  sa2    = 56.16672059;        *estimate of var(influent);
  /* approximate degrees of freedom*/
  v      = (sa2**2)/((((msi/c)**2)/5)+(((mssite/c)**2)/31));
  c025  = cinv(.025,v);      * lower 2.5 chi square percentage point;
  c975  = cinv(.975,v);      * upper 97.5 chi square percentage point;
  low   = v*sa2/C975;        * lower limit;
  high  = v*sa2/C025;        * upper limit;
run;
proc print data=satt;
run;

```

Results

Output 3.7 Intermediate Computations for the Satterthwaite Approximation to the Confidence Interval about σ^2_{Infl}

Obs	c	mssite	msi	sa2	v	c025	c975	low	high
1	6.0973	42.5735	385.039	56.1667	3.94765	0.46822	11.0523	20.0615	473.550

Interpretation

The confidence interval in Output 3.7 is based on the method of moments information in Output 3.5. The estimate of σ^2_{Infl} is 56.17 with approximate degrees of freedom equal to 3.95. The 95% approximate confidence interval about σ^2_{Infl} is $20.06 < \sigma^2_{Infl} < 473.55$ (a very non-symmetric interval). The chi-square distribution can be approximated by a normal distribution for large numbers of degrees of freedom (asymptotic distribution theory). Similarly, confidence intervals about variance components with large numbers of associated degrees of freedom can be constructed using the normal distribution and the estimated standard error of the variance components.

The sampling distribution of a sample variance is that of a chi-square distribution with the associated degrees of freedom, which is an asymmetric distribution. When the degrees of freedom associated with an estimate of a variance are small, the resulting equal-tailed confidence intervals are also asymmetric about the estimate of the variance. When the degrees of freedom are large, the confidence interval is nearly symmetric about the estimate of the variance component. Because the confidence intervals based on the normal distribution are symmetric about the estimates, the intervals can be very misleading when the degrees of freedom are small. For example, when there are only six levels of influent, the symmetric confidence interval based on the normal distribution would not be appropriate. An appropriate confidence interval based on the chi-square distribution with equal-tailed probabilities would be asymmetric.

3.3 Example: A Simple Conditional Hierarchical Linear Model

One purpose of conditional hierarchical linear models is to attempt to explain part of the variability in the levels of a random treatment by using characteristics of the levels. For example, a characteristic may be used to classify the levels into various groups. When the levels of the random factor are classified into groups, the groups are generally considered to be levels of a fixed effect. Thus, the resulting model is a mixed model where the fixed effects parameters correspond to the means of the newly formed groups and the random effects are the levels of the random effect nested within the levels of the fixed effect. This example applies the classification technique to Data Set 3.2, "Mississippi River," in Appendix 2, "Data Sets." The influents were classified according to the type of watershed above the influent as follows:

Type 1	Type 2	Type 3
No farm land in watershed	Less than 50% farm land in watershed	More than 50% farm land in watershed
Influents 3 and 5	Influents 1, 2, and 4	Influent 6

In this case the model is used to determine how much of the variation in the influents in Section 3.2 can be attributed to the different levels of TYPE. Thus, TYPE is considered as a fixed effect and the influents nested within a TYPE are considered as random effects. Although this section considers only the REML estimates of the variance components, other estimation methods could be applied as well.

Let Y_{ijk} denote the amount of nitrogen measured from the k^{th} site at the j^{th} influent of the i^{th} type, and a model to describe the collection of measurements is

$$Y_{ijk} = \mu_i + a_{j(i)} + e_{ikl}, \quad i=1,2,3, \quad j = 1, \dots, n_i, \quad k = 1, \dots, m_{ij} \quad (3.5)$$

where

$$\begin{aligned} a_{j(i)} &\sim \text{iid } N(0, \sigma_a^2) \\ e_{ikl} &\sim \text{iid } N(0, \sigma_e^2) \end{aligned}$$

Model (3.5) has two variance component parameters to be estimated, the same as model (3.2). Model (3.5) is a mixed effects model where μ_i ($i = 1, 2, 3$) is the fixed effects part of the model, $a_{j(i)}$ is the random effects part of the model, and e_{ikl} is the residual part of the model. The notation $j(i)$ denotes that the j^{th} level of influent is nested within the i^{th} type. By classifying the influents according to types, the $\mu_N + a_i$ of model (3.2) have been expressed as $\mu_N + a_i = \mu_i + a_{j(i)}$, where μ_i denotes that the mean level of nitrogen for influents of type i and $a_{j(i)}$ denotes the effect of the j^{th} randomly selected influent of type i . Model (3.5) is also called a two-level conditional hierarchical linear model (Bryk and Raudenbush 1992) since the sampling points are nested within each influent and the influents are classified according to levels of the type, a fixed effect.

3.3.1 Using PROC MIXED to Analyze the Data

Program

The following PROC MIXED program obtains REML estimates of the variance components, estimates of the fixed effects, and predictions of the respective influent means:

```
proc mixed data=influent covtest cl;
  class type influent;
  model y=type/solution;
  random influent(type)/solution;
  estimate 'influent 1' intercept 1 type 0 1 0 | 
             influent(type) 1 0 0 0 0 0;
  estimate 'influent 2' intercept 1 type 0 1 0 | 
             influent(type) 0 1 0 0 0 0;
  estimate 'influent 3' intercept 1 type 1 0 0 | 
             influent(type) 0 0 1 0 0 0;
  estimate 'influent 4' intercept 1 type 0 1 0 | 
             influent(type) 0 0 0 1 0 0;
  estimate 'influent 5' intercept 1 type 1 0 0 | 
             influent(type) 0 0 0 0 1 0;
  estimate 'influent 6' intercept 1 type 0 0 1 | 
             influent(type) 0 0 0 0 0 1;
  lsmeans type / diff;
run;
```

The results in Output 3.8 correspond to the random effects part of the analysis.

Results

Output 3.8 Results of the Random Effects Part of the Model Using PROC MIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
influent(type)	14.9702
Residual	42.5136

Solution for Random Effects							
Effect	type	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent(type)	1	3	0.7653	3.1932	31	0.24	0.8122
influent(type)	1	5	-0.7653	3.1932	31	-0.24	0.8122
influent(type)	2	1	1.2295	2.7593	31	0.45	0.6590
influent(type)	2	2	-4.3259	2.8007	31	-1.54	0.1326
influent(type)	2	4	3.0964	2.8314	31	1.09	0.2826
influent(type)	3	6	4.97E-14	3.8691	31	0.00	1.0000

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 1	20.7034	4.1423	3	5.00	0.0154
influent 2	19.1727	4.1423	3	4.63	0.0190
influent 3	16.8295	4.3989	3	3.83	0.0314
influent 4	15.6122	2.2137	3	7.05	0.0059
influent 5	18.6964	4.4444	3	4.21	0.0245
influent 6	36.4000	2.9159	3	12.48	0.0011

Interpretation

The REML estimates of the variance components are $\hat{\sigma}_{Infl}^2 = 14.970$ and $\hat{\sigma}_e^2 = 42.514$. The “Solution for Random Effects” table displays the predicted values of the effect of the individual influents with mean zero. The ESTIMATE statements provide predictions of the predictable functions for the influents with means equal to the mean of the respective types. Output 3.9 contains the analysis of the fixed effects part of the model using the REML estimates of the variance components.

Output 3.9 Results of the Fixed Effects Part of the Model Using PROC MIXED

Solution for Fixed Effects						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		36.4000	4.8449	3	7.51	0.0049
Type	1	-20.8000	5.9337	3	-3.51	0.0393
Type	2	-16.4619	5.5168	3	-2.98	0.0584
Type	3	0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
type	2	3	6.37	0.0832

Least Squares Means						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
type	1	15.6000	3.4259	3	4.55	0.0198
type	2	19.9381	2.6386	3	7.56	0.0048
type	3	36.4000	4.8449	3	7.51	0.0049

Differences of Least Squares Means							
Effect	type	_type	Estimate	Standard Error	DF	t Value	Pr > t
type	1	2	-4.3381	4.3242	3	-1.00	0.3897
type	1	3	-20.8000	5.9337	3	-3.51	0.0393
type	2	3	-16.4619	5.5168	3	-2.98	0.0584

The statistic to test the equal means hypothesis has a value of 6.37 with a p -value of 0.0832. The LSMEANS or adjusted means for the levels of type and differences between the LSMEANS are included in Output 3.9. There is some evidence that the TYPE 3 mean is different from both TYPE 1 ($p = 0.0393$) and 2 ($p = 0.0584$), and there is no difference between the means of TYPE 1 and 2 ($p = 0.3897$).

The method of moments estimators of the variance components from the TYPE3 option are obtained using the following program.

Program

```
proc mixed data=influent method=type3;
  class type influent;
  model y=type/solution;
  random influent(type)/solution;
  estimate 'influent 1' intercept 1 type 0 1 0 |
             influent(type) 1 0 0 0 0 0;
  estimate 'influent 2' intercept 1 type 0 1 0 |
             influent(type) 0 1 0 0 0 0;
  estimate 'influent 3' intercept 1 type 1 0 0 |
             influent(type) 0 0 1 0 0 0;
  estimate 'influent 4' intercept 1 type 0 1 0 |
             influent(type) 0 0 0 1 0 0;
  estimate 'influent 5' intercept 1 type 1 0 0 |
             influent(type) 0 0 0 0 1 0;
  estimate 'influent 6' intercept 1 type 0 0 1 |
             influent(type) 0 0 0 0 0 1;
  lsmeans type / diff;
run;
```

Output 3.10 PROC MIXED Results with METHOD=TYPE3

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
type	2	1500.033180	750.016590	Var(Residual) + 5.4393 Var(influent(type)) + Q(type)	0.8388 MS(influent(type)) + 0.1612 MS(Residual)	3.3584	6.01	0.0777
influent(type)	3	421.638817	140.546272	Var(Residual) + 6.4848 Var(influent(type))	MS(Residual)	31	3.30	0.0331
Residual	31	1319.779365	42.573528	Var(Residual)

Covariance Parameter Estimates	
Cov Parm	Estimate
influent(type)	15.1079
Residual	42.5735

Solution for Fixed Effects						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		36.4000	4.8603	3	7.49	0.0049
type	1	-20.8000	5.9526	3	-3.49	0.0396
type	2	-16.4618	5.5347	3	-2.97	0.0589
type	3	0

Solution for Random Effects							
Effect	type	influent	Estimate	Std Err Pred	DF	t Value	Pr > t
influent(type)	1	3	0.7675	3.2057	31	0.24	0.8124
influent(type)	1	5	-0.7675	3.2057	31	-0.24	0.8124
influent(type)	2	1	1.2317	2.7692	31	0.44	0.6596
influent(type)	2	2	-4.3357	2.8107	31	-1.54	0.1331
influent(type)	2	4	3.1040	2.8415	31	1.09	0.2831
influent(type)	3	6	-195E-16	3.8869	31	-0.00	1.0000

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
type	2	3	6.33	0.0838

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
influent 1	20.7057	4.1578	3	4.98	0.0156
influent 2	19.1707	4.1578	3	4.61	0.0192
influent 3	16.8317	4.4136	3	3.81	0.0317
influent 4	15.6025	2.2167	3	7.04	0.0059
influent 5	18.7040	4.4593	3	4.19	0.0247
influent 6	36.4000	2.9180	3	12.47	0.0011

Least Squares Means						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
type	1	15.6000	3.4368	3	4.54	0.0200
type	2	19.9382	2.6478	3	7.53	0.0049
type	3	36.4000	4.8603	3	7.49	0.0049

Differences of Least Squares Means							
Effect	type	_type	Estimate	Standard Error	DF	t Value	Pr > t
type	1	2	-4.3382	4.3384	3	-1.00	0.3910
type	1	3	-20.8000	5.9526	3	-3.49	0.0396
type	2	3	-16.4618	5.5347	3	-2.97	0.0589

Interpretation

The main addition to the analysis is that you get an exact F -test for the hypothesis $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$, which has an F -value of 3.30 and a p -value of 0.0331, while the z -score test has a p -value of 0.1967 in Output 3.8. The sum of squares method is a better method for testing hypotheses about the variance components than the z -score method when the number of levels of the random effect is not large, or when the test based on the former is exact.

3.3.2 PROC GLM Part of the Analysis

To start the analysis of the data in Data Set 3.2 using model (3.5), use the following PROC GLM statements to obtain a Type III analysis for testing $H_0: \mu_1 = \mu_2 = \mu_3$ versus H_a : (not H_0) and $H_0: \sigma^2_{Infl} = 0$ versus $H_a: \sigma^2_{Infl} > 0$:

```
proc glm data=influent;
  class type influent;
  model y = type influent(type);
  random influent(type) / test;
run;
```

The TEST option in the RANDOM statement requests that the expected mean squares are used in constructing tests of hypotheses. The mixed model test of the hypotheses above is shown in Output 3.11.

Output 3.11 PROC GLM Results: Type III Sums of Squares and Expected Mean Squares and Tests of Equal Type Means and Zero Variance Component

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1925.193608	385.038722	9.04	<.0001
Error	31	1319.779365	42.573528		
Corrected Total	36	3244.972973			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
type	2	1500.033180	750.016590	17.62	<.0001
influent(type)	3	421.638817	140.546272	3.30	0.0331

Source	Type III Expected Mean Square
type	Var(Error) + 5.4393 Var(influent(type)) + Q(type)
influent(type)	Var(Error) + 6.4848 Var(influent(type))

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type III SS	Mean Square	F Value	Pr > F
type	2	1500.033180	750.016590	6.01	0.0777
Error	3.3584	418.964526	124.750889		
Error: 0.8388*MS(influent(type)) + 0.1612*MS(Error)					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
influent(type)	3	421.638817	140.546272	3.30	0.0331
Error: MS(Error)	31	1319.779365	42.573528		

The F -statistic to test the equal means hypothesis has a value of 6.01 with a p -value of 0.0777. The statistic to test $H_0: \sigma^2_{Infl} = 0$ has a value of 3.30 with a p -value of 0.0331. Thus there is some evidence that both of the hypotheses are false.

3.4 Example: Three-Level Nested Design Structure

The data in Data Set 3.4, “Semiconductor,” in Appendix 2, “Data Sets,” are from a passive data collection study in the semiconductor industry where the objective is to estimate the variance components to determine assignable causes for the observed variability. The measurements are thicknesses of the oxide layer on silicon wafers determined at three randomly selected sites on each wafer. The wafers stem from eight different lots (each lot consists of 25 wafers, but only 3 wafers per lot were used in the passive data collection study). The process consisted of randomly selecting eight lots of 25 wafers from the population of lots of 25 wafers. Then 3 wafers were selected from each lot of 25 for use in the oxide deposition process. After the layer of oxide was deposited, the thickness of the layer was determined at three randomly selected sites on each wafer. The structure of the study involves three sizes of experimental units in the design structure with a uniform application of a single treatment in the treatment structure.

3.4.1 Three-Level Nested Linear Model, an Unconditional Hierarchical Nested Linear Model

A model to describe the data in Data Set 3.4 is

$$Y_{ijk} = \mu + a_i + w_{j(i)} + s_{k(j)}, \quad i = 1, 2, \dots, 8, \quad j = 1, 2, 3, \quad k = 1, 2, 3 \quad (3.6)$$

where

$$\begin{aligned} a_i &\sim \text{iid } N(0, \sigma_L^2) \\ w_{j(i)} &\sim \text{iid } N(0, \sigma_w^2) \\ s_{k(ij)} &\sim \text{iid } N(0, \sigma_s^2) \end{aligned}$$

and

a_i is the effect of the i^{th} randomly selected lot
 $w_{j(i)}$ is the effect of the j^{th} randomly selected wafer from the i^{th} lot
 $s_{k(ij)}$ is the effect of the k^{th} randomly selected site from the j^{th} wafer of the i^{th} lot

In the linear models literature, model (3.6) has been called a **three-level nested linear model** or an **unconditional hierarchical nested linear model**. The objective of the passive data collection study is to estimate the variance components, σ_L^2 , σ_w^2 , and σ_s^2 .

3.4.2 Data Analysis Using the PROC MIXED METHOD=REML to Estimate the Variance Components

The PROC MIXED program to fit model (3.6) is as follows, where lot and wafer(lot) effects are specified in the RANDOM statement and site(Wafer Lot) is the residual.

```
proc mixed data=e_3_4 Method=REML;
  class lot wafer site;
  model Thick=;
  random lot wafer(lot);
run;
```

The mean is the only fixed effect and the residual variance corresponds to the site-to-site variance. The variance components corresponding to LOT and Wafer(LOT) measure the variability in the mean thickness of the population of lots and the variation in the mean thickness of the wafers within the population of lots, respectively.

Results

The REML estimates of the variance components are shown in Output 3.12.

Output 3.12 Results of PROC MIXED for the Three-Level Nested Random Effects Model

Covariance Parameter Estimates	
Cov Parm	Estimate
lot	129.91
wafer(lot)	35.8657
Residual	12.5694

Interpretation

For this study, the estimate of the lot-to-lot variance of 129.9 is four times larger than the wafer-to-wafer within-a-lot variance of 35.9, which is 2.85 times larger than the site-to-site within a wafer variance of 12.6. If possible, one may want to evaluate the cause of the lot-to-lot variance in order to improve the consistency of the oxide layer across the population of wafers.

3.4.3 Using the PROC MIXED METHOD=TYPE1 to Estimate the Variance Components

Method of moments estimation provides tests of hypotheses that each of the variance components is equal to zero. The following program requests moment estimators with the METHOD=TYPE1 option.

Program

```
proc mixed data=e_3_4 Method=Type1 covtest cl;
  class lot wafer site;
  model Thick=;
  random lot wafer(lot);
run;
```

Results

The results are given in Output 3.13.

Output 3.13 METHOD=TYPE1 Results for the Three-Level Nested Random Effects Model

Type 1 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
lot	7	9025.319444	1289.331349	Var(Residual) + 3 Var(wafer(lot)) + 9 Var(lot)	MS(wafer(lot))	16	10.73	<.0001
wafer(lot)	16	1922.666667	120.166667	Var(Residual) + 3 Var(wafer(lot))	MS(Residual)	48	9.56	<.0001
Residual	48	603.333333	12.569444	Var(Residual)

Covariance Parameter Estimates	
Cov Parm	Estimate
lot	129.91
wafer(lot)	35.8657
Residual	12.5694

Interpretation

The method of moments estimates are

$$\hat{\sigma}_s^2 = MSERROR$$

$$\hat{\sigma}_w^2 = \frac{MSWAFER(LOT) - MSERROR}{3}$$

$$\hat{\sigma}_L^2 = \frac{MSLOT - MSWAFER(LOT)}{9}$$

For the test of the hypothesis $H_0: \sigma_L^2 = 0$ versus $H_a: \sigma_L^2 > 0$, the value of the F -statistic is 10.73 with $p < 0.0001$. For the test of the hypothesis $H_0: \sigma_w^2 = 0$ versus $H_a: \sigma_w^2 > 0$, the value of the F -

statistic is 9.56 with $p < 0.0001$. These tests indicate that the variability among lots and the variability among wafers within a lot are significantly different from zero; they are important sources of variability in the system.

The REML and method of moments estimates of the variance components are identical in this case, as this data set is balanced.

3.4.4 Conditional Hierarchical Linear Model or Mixed Model

The next part of the analysis is to take into account the information that the lots are from two different sources, denoted by SOURCE in the code for model (3.7). Because the levels of source are fixed effects, we change from the unconditional hierarchical nested linear model in (3.6) to the following conditional hierarchical nested linear model:

$$Y_{ijkm} = \mu_i + a_{j(i)} + w_{k(j)} + s_{m(jk)}, \quad i = 1, 2, \quad j = 1, 2, 3, 4, \quad k = 1, 2, 3, \quad m = 1, 2, 3 \quad (3.7)$$

where

$$\begin{aligned} a_{j(i)} &\sim \text{iid } N(0, \sigma_a^2) \\ w_{k(j)} &\sim \text{iid } N(0, \sigma_w^2) \\ s_{m(jk)} &\sim \text{iid } N(0, \sigma_s^2) \\ \mu_i &\text{ is the mean of the } i^{\text{th}} \text{ source level} \end{aligned}$$

and

- $a_{j(i)}$ is the effect of the j^{th} randomly selected lot from source i
- $w_{k(j)}$ is the effect of the k^{th} randomly selected wafer from the j^{th} lot from source i
- $s_{m(jk)}$ is the effect of the m^{th} randomly selected site from the k^{th} wafer of the j^{th} lot from source i

The μ_i represent the fixed effects part of the model, the $a_{j(i)} + w_{k(j)}$ represent the random effects part of the model, and $s_{m(jk)}$ is the residual part of the model. Since the model involves both random and fixed factors, model (3.7) is a mixed model. The discussion here is to demonstrate the process of moving from the purely random effects linear model or unconditional hierarchical linear model to the mixed model or conditional hierarchical linear model.

3.4.5 Using the MIXED Procedure

The PROC MIXED program to fit model (3.7) is as follows:

```
proc mixed data=e_3_4;
  class source lot wafer site;
  model Thick = source / ddfm=kr;
  random lot(source) wafer(source lot);
  lsmeans source / diff;
run;
```

The levels of the sources, SOURCE, denote the fixed effects part of the model. Hence SOURCE is listed in the MODEL statement. The DDFM=KR option in the MODEL statement uses a Satterthwaite-type approximation for the degrees of freedom and adjusts the estimated standard errors for fixed effects and predictable functions.

Results

The results of fitting the mixed model to the thickness data are shown in Output 3.14. The output includes estimates of the means of the fixed effects and comparisons of the means that are produced by the LSMEANS statement.

Output 3.14 PROC MIXED Results for the Nested Random Effects Model Using Source as a Fixed Effect for the Conditional Hierarchical Linear Model

Covariance Parameter Estimates	
Cov Parm	Estimate
lot(source)	119.89
wafer(source*lot)	35.8657
Residual	12.5694

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
source	1	6	1.53	0.2629

Least Squares Means						
Effect	source	Estimate	Standard Error	DF	t Value	Pr > t
source	1	1995.11	5.7716	6	345.68	<.0001
source	2	2005.19	5.7716	6	347.43	<.0001

Differences of Least Squares Means							
Effect	source	_source	Estimate	Standard Error	DF	t Value	Pr > t
source	1	2	-10.0833	8.1622	6	-1.24	0.2629

Interpretation

The estimate of the lot-to-lot variance component is a little smaller when the lot source is included in the model (compare to Output 3.12), because the lot-to-lot variance component now measures the variability among lots within a source. Some of the variability of the lots is attributable to the levels of SOURCE and thus is reflected in the smaller estimate in Output 3.14. Because the data set is balanced, the wafer-to-wafer variance component and the site-to-site variance component are identical for both the unconditional and conditional hierarchical linear models.

3.4.6 Unequal Variance Model

Oftentimes factors in an experiment have an effect on the variance of the responses as well as on the mean of the responses. Because there are two sources of lots of wafers, it is possible that the lot-to-lot variability is different for each source. The last phase of the analysis is to fit a model where the lot-to-lot variance components varies with the source of lots. The heterogeneous variance component model is

$$Y_{ijkm} = \mu_i + a_{j(i)} + w_{k(ij)} + s_{m(ijk)}, \quad i = 1, 2, \quad j = 1, 2, 3, 4, \quad k = 1, 2, 3, \quad m = 1, 2, 3 \quad (3.8)$$

where

$$a_{j(i)} \sim \text{iid } N(0, \sigma_{L_1}^2)$$

$$w_{k(ij)} \sim \text{iid } N(0, \sigma_w^2)$$

$$s_{m(ijk)} \sim \text{iid } N(0, \sigma_s^2)$$

Model (3.8) has two variance components for LOTS; $\sigma_{L_1}^2$ and $\sigma_{L_2}^2$ are the variances of the lots within source 1 and source 2, respectively.

Program

The PROC MIXED program to fit the unequal variance model is as follows:

```
proc mixed data=e_3_4 covtest cl scoring=4;
  class source lot wafer site;
  model Thick = source/ ddfm=kr;
  random lot / group=source;
  random wafer(source lot);
  lsmeans source / diff;
run;
```

The SCORING=4 option for the MODEL statement requests that Fisher scoring be used in association with the estimation method up to iteration number 4. When the model is complex (the unequal variances make this model a little complex), the SCORING= *number* sometimes helps with convergence. The GROUP= option in the first RANDOM statement requests that the parameters of the associated covariance structure be varied by the group effect. Here, the lot variance component is varied by levels of the SOURCE effect. This provides a different estimate of $\sigma_{L_1}^2$ for each level of SOURCE.

Results

Output 3.15 PROC MIXED Results Using Source as a Fixed Effect for the Conditional Hierarchical Linear Model with Unequal Variances for the Two Sources of Wafers for the Three-Level Nested Random Effects Model

Covariance Parameter Estimates		
Cov Parm	Group	Estimate
lot	source 1	17.0761
lot	source 2	222.71
wafer(source*lot)		35.8657
Residual		12.5694

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
source	1	3.76	1.53	0.2883

Least Squares Means						
Effect	source	Estimate	Standard Error	DF	t Value	Pr > t
source	1	1995.11	2.7581	3	723.37	<.0001
source	2	2005.19	7.6821	3	261.02	<.0001

Differences of Least Squares Means							
Effect	source	_source	Estimate	Standard Error	DF	t Value	Pr > t
source	1	2	-10.0833	8.1622	3.76	-1.24	0.2883

Interpretation

The results are in Output 3.15, where the estimates of the lot-to-lot variance components are 17.08 and 222.71 for sources 1 and 2, respectively. The standard errors of the estimates of the SOURCE means reflect the unequal variances. That is, the estimated standard errors are higher for groups with greater variance: 2.76 for source 1 and 7.68 for sources 2, respectively, each based on 3 degrees of freedom. The common lot-to-lot variance component model provides the same estimated standard error for the two source means, 5.77 based on 6 degrees of freedom. (The estimated standard errors are equal for the common lot-to-lot variance model in this case because there are equal numbers of observations from each level of SOURCE.)

Next, compare the analysis in Output 3.14 to the analysis in Output 3.15 to see the effect of unequal variances on the analysis of the fixed effects in the model. The *p*-value for comparing the SOURCE means changes from 0.2629 for the equal variance model to 0.2883 for the unequal variance model. This change in the *p*-values is due to the differences in the denominator degrees of freedom. The estimated standard errors of the differences of the two SOURCE means are identical for both analyses. This similarity disappears if the number of observations per level of SOURCE is not equal. The estimates of the wafer-to-wafer and site-to-site variance components are unchanged (35.87 and 12.57, respectively). The mean of the two lot-to-lot variance components is equal to the single lot-to-lot variance component for the equal variances model ($(17.08 + 222.71)/2 = 119.89$). This equality occurs because there are equal numbers of observations per level of SOURCE. Because the mean of the two lot-to-lot variance components is equal to the lot-to-lot variance component for the equal variances model, the estimate of the standard error of the difference between the two sources is the same for both models (8.162). The least-squares means and the difference between the two means are identical for both models. A test of the hypothesis of equal variances can be obtained by comparing the difference between the -2 Res Log Likelihood values to a chi-square distribution based on $4-3 = 1$ degrees of freedom, because there are four parameters in the covariance structure for the unequal variances model and there are three parameters in the covariance structure for the equal variances model. In this case, the difference is 2.7148 based on 1 degree of freedom, and the *p*-value is 0.3173 (from the chi-square table based on 1 degree of freedom). The results of the equal variances test indicate that the equal variance model is adequate to describe the data.

3.5 Example: A Two-Way Random Effects Treatment Structure to Estimate Heritability

The data in Data Set 3.5, “Genetics,” in Appendix 2, “Data Sets,” represent yields of five wheat families grown at four randomly selected locations. The wheat families were randomly selected from a population of families in a breeding program. The locations were selected from locations where the type of wheat would be grown commercially. At each location, the design of the experiment was a one-way treatment structure in a randomized complete block design. The objective of the study was to estimate the heritability of yield, a measure of possible genetic advancement under selection (Allard 1966). Heritability is estimated by the ratio of the estimate of the genetic variance to the phenotypic variance of a family mean. A model to describe the data is

$$Y_{ijk} = \mu + l_i + f_j + lf_{ij} + b_{k(i)} + e_{ijk} \quad (3.9)$$

where

$$i = 1, 2, 3, 4 \quad j = 1, 2, 3, 4, 5 \quad k = 1, 2, 3$$

$$l_i \sim \text{iid } N(0, \sigma_L^2)$$

$$f_j \sim \text{iid } N(0, \sigma_F^2)$$

$$lf_{ij} \sim \text{iid } N(0, \sigma_{LF}^2)$$

$$b_{k(i)} \sim \text{iid } N(0, \sigma_B^2)$$

$$e_{ijk} \sim \text{iid } N(0, \sigma_e^2)$$

The additive genetic variance component is σ_F^2 , and the variance of a family mean is $\sigma_{\bar{y}_{j.}}^2 = \sigma_L^2 / 4 + \sigma_F^2 + \sigma_{LF}^2 / 4 + \sigma_B^2 / 12 + \sigma_e^2 / 12$. For this design, heritability is defined as

$$h^2 = \sigma_F^2 / \sigma_{\bar{y}_{j.}}^2.$$

3.5.1 Using PROC MIXED METHOD=REML

The following PROC MIXED program is used to fit model (3.9) to obtain REML estimates of the variance components:

```
proc mixed data=e_3_5 covtest cl;
  class loc fam block;
  model Yield =;
  random loc fam loc*fam block(loc);
run;
```

Because the fixed effects part of the model consists of only the intercept, the MODEL statement contains no terms (an intercept is automatically included by PROC MIXED). All other terms in the model are random effects and thus occur in the RANDOM statement.

Results

Output 3.16 PROC MIXED Results Providing REML Estimates of the Variance Components for the Genetics Example

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
loc	613.65	540.71	1.13	0.1282	0.05	185.08	12013
fam	188.00	149.52	1.26	0.1043	0.05	61.6103	2349.79
loc*fam	74.8616	37.8647	1.98	0.0240	0.05	33.9079	280.29
block(loc)	89.3208	49.9124	1.79	0.0368	0.05	37.9170	404.14
Residual	51.8458	12.9615	4.00	<.0001	0.05	33.5297	90.7052

Interpretation

The results of PROC MIXED are in Output 3.16. The estimates of the variance components are

$$\hat{\sigma}_L^2 = 613.6$$

$$\hat{\sigma}_F^2 = 188.0$$

$$\hat{\sigma}_{LF}^2 = 74.9$$

$$\hat{\sigma}_B^2 = 89.3$$

$$\hat{\sigma}_e^2 = 51.8$$

$$\hat{\sigma}_{\bar{y}_{ij}}^2 = \frac{613.6}{4} + 188.0 + \frac{74.9}{4} + \frac{89.3}{12} + \frac{51.8}{12} = 371.8$$

$$\hat{h}^2 = \frac{188.0}{371.8} = 0.506$$

The estimate of the heritability is 0.506, indicating that 50.6% of the variability in the family means is due to additive genetic variance.

The *z*-scores in Output 3.16 provide results for testing the following hypotheses:

$$H_0 : \sigma_L^2 = 0 \text{ versus } H_a : \sigma_L^2 > 0$$

$$H_0 : \sigma_F^2 = 0 \text{ versus } H_a : \sigma_F^2 > 0$$

$$H_0 : \sigma_{LF}^2 = 0 \text{ versus } H_a : \sigma_{LF}^2 > 0$$

The *p*-values for the hypotheses are 0.1282, 0.1043, and 0.0240, respectively.

3.5.2 Using PROC MIXED METHOD=TYPE3

The analysis of variance table for model (3.9) is constructed using the following PROC MIXED statements with the METHOD=TYPE3 option.

Program

```
proc mixed data=e_3_5 method=type3;
  class loc fam block;
  model Yield=;
  random loc fam loc*fam block(loc);
run;
```

The PROC MIXED results are displayed in Output 3.17.

Results

Output 3.17 PROC MIXED Analysis with METHOD=TYPE3 for Genetics Example

Type 3 Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value
loc	3	29783	9927.777778	Var(Residual) + 5 Var(block(loc)) + 3 Var(loc*fam) + 15 Var(loc)	MS(loc*fam) + MS(block(loc)) - MS(Residual)	13.938	13.73
fam	4	10130	2532.441667	Var(Residual) + 3 Var(loc*fam) + 12 Var(fam)	MS(loc*fam)	12	9.16
loc*fam	12	3317.166667	276.430556	Var(Residual) + 3 Var(loc*fam)	MS(Residual)	32	5.33
block(loc)	8	3987.600000	498.450000	Var(Residual) + 5 Var(block(loc))	MS(Residual)	32	9.61
Residual	32	1659.066667	51.845833	Var(Residual)	.	.	.

Type 3 Analysis of Variance	
Source	Pr > F
loc	0.0002
fam	0.0012
loc*fam	<.0001
block(loc)	<.0001
Residual	.

Covariance Parameter Estimates	
Cov Parm	Estimate
loc	613.65
fam	188.00
loc*fam	74.8616

Covariance Parameter Estimates	
Cov Parm	Estimate
block(loc)	89.3208
Residual	51.8458

Interpretation

The analysis of variance table and expected mean squares in Output 3.17 includes statistics to test the following hypotheses:

$$H_0 : \sigma_L^2 = 0 \text{ versus } H_a : \sigma_L^2 > 0$$

$$H_0 : \sigma_F^2 = 0 \text{ versus } H_a : \sigma_F^2 > 0$$

$$H_0 : \sigma_{LF}^2 = 0 \text{ versus } H_a : \sigma_{LF}^2 > 0$$

The *p*-values corresponding to the hypotheses above are 0.0002, 0.0012, and < 0.0001, respectively. These tests indicate that the variance components in the model are required to adequately describe the variation in the data. The only *F*-statistic with an approximate denominator in Output 3.17 corresponds to LOC.

There is a large discrepancy between the *p*-values from METHOD=REML and METHOD=TYPE3. One reason for the discrepancy is that with REML estimation, PROC MIXED uses the asymptotic normal sampling distribution while TYPE3 estimation relies on the *F*-distribution. In this example, there are very few levels of the random effects. Thus, the asymptotic normal distribution of the REML estimates is most likely not appropriate. Several more levels of the factor are necessary before the asymptotic distribution becomes appropriate.

3.6 Summary

The concept of a random effect was described, where the random effect can occur in the design structure and/or the treatment structure. Procedures for estimating the parameters of a random effects model were demonstrated using PROC MIXED. Small sample size approximations to confidence intervals using the Satterthwaite approximation were described and were constructed for several examples. Unconditional and conditional hierarchical linear models were described, which are special cases of random effects and mixed linear models, respectively. Random coefficient models (Chapter 8) are a special type of mixed analysis of covariance models (see Chapter 7) and are conditional hierarchical linear models when all of the random effects are hierarchical.



Multi-factor Treatment Designs with Multiple Error Terms

4.1	Introduction	94
4.2	Treatment and Experiment Structure and Associated Models.....	94
4.2.1	Possible Layouts of Factorial and Split-Plot Experiments.....	95
4.2.2	Determining the Appropriate Mixed Model for a Given Layout	99
4.3	Inference with Mixed Models for Factorial Treatment Designs	102
4.3.1	Effects of Interest in Factorial Treatment Designs	102
4.3.2	More on Effects of Interest—Estimability and Use of ESTIMATE, CONTRAST, and LSMEANS Statements	104
4.3.3	Standard Errors	106
4.4	Example: A Split-Plot Semiconductor Experiment	113
4.4.1	Tests of Interest in the Semiconductor Experiment.....	113
4.4.2	Matrix Generalization of Mixed Model F-tests	115
4.4.3	PROC MIXED Analysis of Semiconductor Data	117
4.4.4	Alternative Mean Comparisons Using PROC GLIMMIX.....	128
4.5	Comparison with PROC GLM	130
4.6	Example: Type × Dose Response	135
4.6.1	PROC MIXED Analysis of DOSE × TYPE Main Effects and Interactions.....	136
4.6.2	Regression Analysis over DOSE by TYPE	140
4.7	Example: Variance Component Estimates Equal to Zero.....	148
4.7.1	Default Analysis Using PROC MIXED	148
4.7.2	Recommended Alternative: Analysis with NOBOUND or METHOD=TYPE3 Options	150
4.7.3	Conceptual Alternative: Negative Variance or Correlation?.....	152
4.8	More on PROC GLM Compared to PROC MIXED: Incomplete Blocks, Missing Data, and Estimability	154
4.9	Summary	156

4.1 Introduction

Researchers often conduct multi-factor experiments in order to study the effects of two or more factors simultaneously. Factorial experiments are so named because of their treatment structure. Many textbooks about the design of experiments distinguish between the treatment structure, or **treatment design**, and the manner in which treatment combinations are assigned to experimental units, or **experiment design**. The factorial design is a type of treatment design and is not, by itself, completely descriptive, because it may be used in conjunction with several possible experiment designs. These include completely randomized and randomized blocks designs, described in Chapter 2, but also split-plot, strip-plot, nested, clustered, and other designs that give rise to two or more sources of random error. Consequently, these designs give rise to mixed models and hence, mixed models methods are essential to their proper analysis. Repeated measures and longitudinal data are closely related to split-plot structures but involve the additional complication of serially correlated errors. Repeated measures analysis is discussed separately in Chapter 5.

Many, especially older, textbooks describe split-plot analysis using analysis of variance and fixed effects linear model software such as PROC GLM. While these procedures are adequate, if somewhat awkward, for testing overall sources of variation, they are not designed to provide appropriate standard errors of many of the treatment differences that are of interest in most factorial experiments. Specifically, if scientifically important interactions exist among treatment factors, methods based on fixed effects linear models *cannot* correctly compute standard errors for simple effects of interest, and *will* compute *incorrect* results. PROC GLM *cannot* correctly compute a complete analysis of a split-plot experiment. Within the SAS System, PROC MIXED (or PROC GLIMMIX) is required. This chapter shows how.

Section 4.4.4 contains additional material using PROC GLIMMIX.¹ This procedure has certain features for comparing means that are simpler to implement than PROC MIXED and provides certain plots not available in PROC MIXED to help visualize relationships among means. The GLIMMIX procedure was released very late in the production of this book, so examples in this chapter are necessarily limited. Chapter 14 discusses PROC GLIMMIX in greater detail over a broader range of applications.

4.2 Treatment and Experiment Structure and Associated Models

Although factorial experiments are often viewed as a specific type of design, in reality they can be set up and conducted in a wide variety of ways. In this section the word *factorial* refers specifically to the 2×2 cross-classified treatment structure. Do not confuse it with the names used for the various experiment designs—that is, ways in which treatment combinations are assigned to experimental units, described below.

This section shows seven different ways to conduct a 2×2 factorial treatment design, each leading to a different model and analysis. The 2×2 is the *simplest* factorial treatment structure, and even the seven layouts shown here are not an exhaustive list of *possible* experiment designs! The purpose of this section is to show you how to visualize the layout (Section 4.2.1)

¹ The GLIMMIX procedure is an add-on in SAS 9.1 to SAS/STAT for the (32-bit) Windows platform. It does not ship with SAS 9.1. You can obtain the GLIMMIX procedure for SAS 9.1 as a download from www.sas.com/statistics. This site also contains the documentation for the GLIMMIX procedure.

and how to associate it with an appropriate mixed model (Section 4.2.2). The approach shown here generalizes to arbitrarily complex factorial experiments. Once you master the strategy, you can adapt it to your own data. The key concept is that different factors or factorial combinations are applied to different experimental units: identify the experimental units, and the model follows.

4.2.1 Possible Layouts of Factorial and Split-Plot Experiments

Suppose that you want to investigate two treatment factors, generically referred to as factor A and factor B, or A and B, respectively, each with two levels, denoted A_1 and A_2 for factor A and B_1 and B_2 for factor B. Factor A could be two types of drug, two varieties of a crop, two materials in a manufacturing process, or two different teaching methods. Factor B could be two different levels—e.g., each drug applied at a low dose (B_1) or high dose (B_2)—or it could be a distinct factor. In a factorial experiment, the treatments are cross-classified: all possible combinations of factor levels may be observed. Here, there are four treatments: A_1 applied in combination (or crossed) with B_1 (referred to as $A_1 \times B_1$, and abbreviated as AB_{11} later in this chapter); A_1 crossed with B_2 (denoted $A_1 \times B_2$, abbreviated as AB_{12}); A_2 crossed with B_1 (denoted $A_2 \times B_1$, abbreviated as AB_{21}); and A_2 crossed with B_2 (denoted $A_2 \times B_2$, abbreviated as AB_{22}). Note that factor levels are never applied alone—treatments always consist of a level of one factor in combination with a level of the other factor.

The four treatments can be observed in a variety of ways. First, each treatment must be assigned to an **experimental unit**. An experimental unit is defined as the smallest entity to which a treatment is independently applied. The word *independent* is important. For example, if you want to compare two teaching methods, you could design the experiment by assigning method A_1 to one teacher to use in her class with n students, and method A_2 to another teacher to use in his class. In this case, the *class* is the experimental unit: while there are individual students in each class, once the class is assigned to a given treatment, all students in the class are assigned as well. The assignment of individual students is not independent. On the other hand, you could design the experiment by assigning individual students within a class to different teaching methods (assuming it is logically possible to do so). For example, some students within the class could be assigned to learn data analysis using a calculator, whereas others could be assigned to data analysis using a computer. Then students would be the experimental unit and the class would be a **blocking factor**. These two approaches imply different models and require different analyses. It is important to be very clear about how a study was conducted in order to use the right model and do the right analysis.

In the 2×2 factorial, you can conduct the experiment in a variety of ways. Figure 4.1 shows seven common designs associated with a 2×2 factorial.

Figure 4.1 Possible Design Layouts for 2×2 Factorial Experiment

Treatment codes



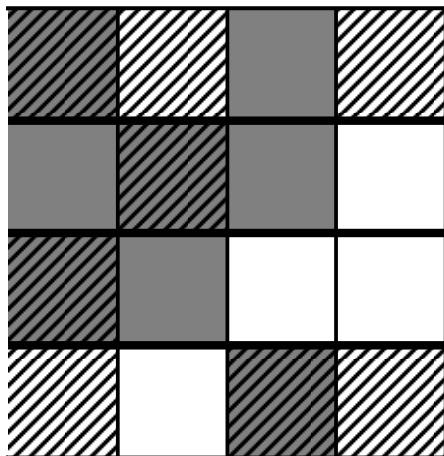
$A_1 \times B_1$

$A_1 \times B_2$

$A_2 \times B_1$

$A_2 \times B_2$

4.1.a Completely Randomized



4.1.b Randomized Complete Block

Blk 1				
Blk 2				
Blk 3				
Blk 4				

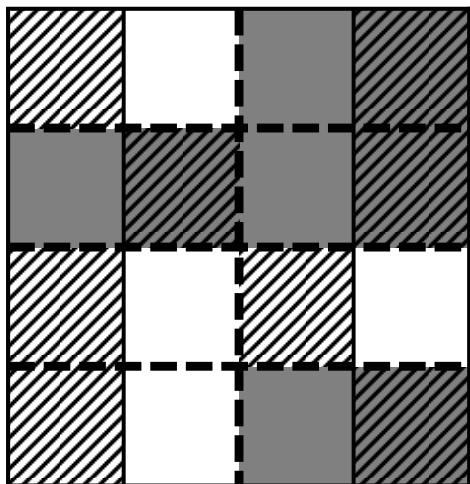
A 4x5 grid representing a Randomized Complete Block design. The columns are labeled 'Blk 1' through 'Blk 4'. The first column contains four hatched squares. The second column contains one white square, one solid gray square, one hatched square, and one solid gray square. The third column contains one solid gray square, one white square, one hatched square, and one solid gray square. The fourth column contains one white square, one hatched square, one solid gray square, and one hatched square. The fifth column contains one solid gray square.

4.1.c Row-Column (Latin Square)

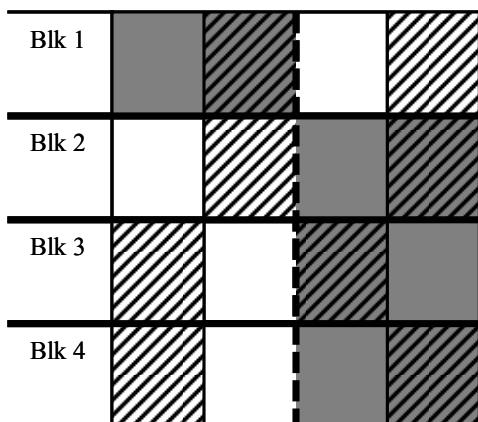
	col1	col2	col3	col4
row 1				
row2				
row3				
row4				

A 4x4 grid representing a Row-Column (Latin Square) design. The columns are labeled 'col1' through 'col4'. The rows are labeled 'row 1' through 'row4'. The first column contains four hatched squares. The second column contains one white square, one solid gray square, one hatched square, and one solid gray square. The third column contains one solid gray square, one white square, one solid gray square, and one hatched square. The fourth column contains one solid gray square, one hatched square, one solid gray square, and one white square.

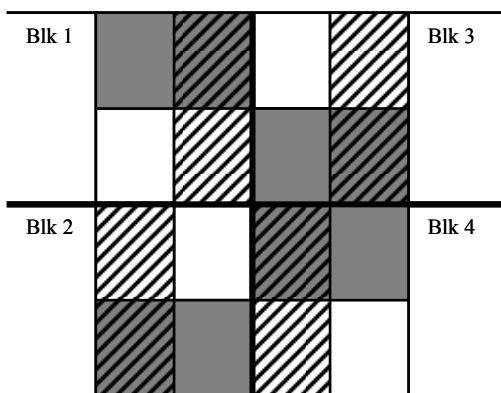
4.1.d Split Plot 1, Whole Plot Completely Randomized



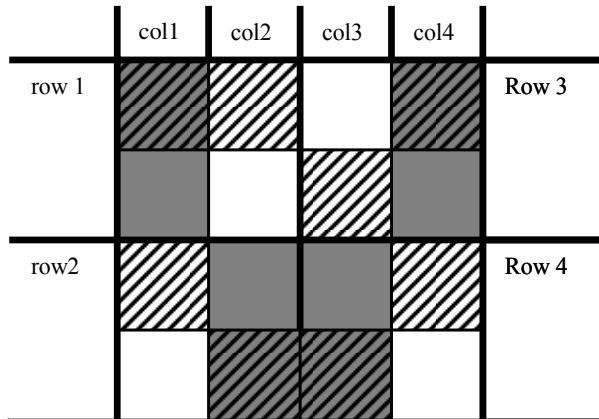
4.1.e Split Plot 2, Whole Plot in Randomized Complete Blocks



4.1.f Split Block, a.k.a. Strip-Split Plot



4.1.g Split Plot 3, Whole Plot in Row-Column (2 Latin Squares)



You could start with $4n$ experimental units and assign each treatment combination at random to n of them. Figure 4.1.a shows such a design for $n = 4$. This is a completely randomized design (CRD) with a 2×2 factorial treatment structure. If there is a basis for doing so, you could divide the experimental units into subsets, or **blocks**, so that the experimental units within each block are more similar to one another than to experimental units in different blocks. This gives rise to a randomized complete block (RCB) design (Figure 4.1.b), whose analysis was described in Chapter 2. Alternatively, you could block on rows and columns, yielding a Latin square design (Figure 4.1.c), whose analysis was described in Chapters 3 and 4 of Littell et al. (2002). Each of these three experiment designs uses the same treatment design and has a single source of experimental error.

The fourth experiment design for the 2×2 factorial is shown in Figure 4.1.d. There are 16 units depicted in the figure. These are grouped into 8 pairs of units, which appear as 1×2 rectangles, shown by the heavy dashed lines. The two levels of factor A are randomly assigned to 4 of these larger units each. Within each pair, the two levels of B are randomly assigned, one to each unit within the pair. This is a **split-plot experiment**. The units to which levels of A are assigned are called **whole-plot experimental units**, and the units to which levels of B are assigned are called **split-plot experimental units**. In general, a split-plot experiment is a factorial treatment design conducted such that the experimental unit with respect to one or more factors is a subunit of the experimental unit with respect to other factors. Why conduct split-plot experiments? Broadly speaking, there are three reasons:

1. Out of necessity when a factor, or factorial combination, must be applied to relatively large experimental units, whereas other factors are more appropriately applied to subunits.
2. For convenience: it is often simply easier to apply different factors to different sized units.
3. To increase the precision of the estimated effect of the factor applied to the subunits.

The split plot shown in Figure 4.1.d uses a completely randomized design to assign levels of factor A to whole-plot experimental units. Figures 4.1.e, 4.1.f, and 4.1.g show variations of the split plot using blocking. In Figure 4.1.e, the whole-plot experimental units are blocked as they were in Figure 4.1.b. The levels of A are assigned to units corresponding to half of the block (marked by the heavy dashed line), and the levels of B are assigned to the two units within each level of A. In Figure 4.1.f, the blocks consist of 2×2 squares. The levels of A are randomly assigned to the upper or lower half of each block, whereas the levels of B are randomly assigned

to the right or left half. Depending on the author, textbooks may refer to this as a **split-block** or **strip-split plot**. Finally, Figure 4.1.g uses a Latin square to arrange the whole-plot experimental units. There are two 2×2 Latin squares, one on the left (rows and columns 1 and 2) and one on the right (rows and columns 3 and 4). Levels of A are assigned within each square consistent with the constraints of the Latin square. Each row-column combination has 2 split-plot units: levels of B are randomly assigned, one level per unit.

There are other variations, but these seven serve to show the variety of experiment designs that can be used even with a simple treatment design.

4.2.2 Determining the Appropriate Mixed Model for a Given Layout

Once you have a picture of the way in which observations were obtained, like the layouts in Figure 4.1, you can determine the mixed model needed to describe the sources of variation. The basic idea is that the model structure is

$$\text{observation} = \text{treatment design components} + \text{experiment design components}$$

The treatment design components describe the sources of variation associated with the treatment factors, and the experiment design components describe additional sources of variation introduced by the way experimental units are assigned to treatments plus random variation. In the 2×2 factorial, all seven designs share the same generic model,

$$Y_{ijk} = \mu_{ij} + E_{ijk}$$

where Y_{ijk} denotes the observation on the k^{th} experimental unit assigned to the ij^{th} $A_i \times B_j$ treatment combination, μ_{ij} denotes the mean of the ij^{th} treatment combination, and E_{ijk} denotes all other variability on the observation.

The treatment mean is typically decomposed into main effects and interaction terms—i.e., $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, where μ is the overall mean or intercept, α_i and β_j are the main effects of A and B, respectively, and $(\alpha\beta)_{ij}$ is the $A \times B$ interaction term. Section 4.3 discusses the specific meaning of main effect and interaction in greater detail.

The E_{ijk} term is written to match the layout or design of the study, specifically the size and assignment of experimental units. For the seven designs shown in Figure 4.1, consider the following table.

Table 4.1 Model–Design Association

Effect	Figures						
	4.1.a CRD	4.1.b RCB	4.1.c LS	4.1.d split-plot CR	4.1.e split-plot RCB	4.1.f split-block	4.1.g split-plot LS
Block?	no	yes	row col	no	yes	yes	row col
A	eu(A*B)	blk*A*B	row*col	eu(A)	blk*A	blk*A	row*col
B	eu(A*B)	blk*A*B	row*col	B*eu(A)	blk*A*B	blk*B	row*col*B
A*B	eu(A*B)	blk*A*B	row*col	B*eu(A)	blk*A*B	blk*A*B	row*col*B

Table 4.1 give two types of information essential to associating a model with a design: (1) are there any blocking factors (if so, what?), and (2) what is the experimental unit with respect to each of the treatment main effects and interactions? For the CRD, there is no blocking, and the experimental unit (denoted eu in column 4.1.a) is assigned to $A \times B$ combinations and hence is shared by all treatment factors. Thus the experiment design component, E_{ijk} , is simply experimental error and the model can be written

$$Y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

where e_{ijk} is typically assumed iid $N(0, \sigma^2)$.

For the RCB, there is blocking and there is only one size of experimental unit assigned to $A \times B$ treatment combinations. This experimental unit corresponds to a block $\times A \times B$ combination as depicted in Figure 4.1.b. Hence, the experiment design component is decomposed as $E_{ijk} = r_k + e_{ijk}$, where r_k denotes the k^{th} block effect and e_{ijk} denotes experimental error. The model is thus

$$Y_{ijk} = \mu_{ij} + r_k + e_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + r_k + e_{ijk}$$

where the e_{ijk} are typically assumed iid $N(0, \sigma^2)$ and the r_k are typically assumed random but may be assumed fixed depending on the circumstances. The Latin square design (Figure 4.1.c) is similar, except that separate row and column terms replace the single block term in the model.

For the split-unit designs (Figures 4.1.d through 4.1.g), there are different-size experimental units with respect to the various treatment effects. For the CRD split-plot design (Figure 4.1.d), $eu(A)$ denotes the whole-plot experimental unit. There must be a column in the data set (called eu in Table 4.1) to identify these units and you must include this variable name in the CLASS statement. Note also that main effects and their interactions may have different-size experimental units. The extreme case is the split-block design (Figure 4.1.f), for which levels of A are applied to units defined by block $\times A$ combinations, level of B are assigned to units defined by block $\times B$, and thus $A \times B$ combinations are effectively assigned to units defined by block $\times A \times B$. The E_{ijk} term for each design reflects the information in Table 4.1.

For the design in Figure 4.1.d, there is no blocking, but levels of A are assigned to larger units (defined by “e.u.” within A), and levels of B (and hence $A \times B$ combinations) are assigned to smaller units defined by the intersection of B and “e.u.” within A. Thus, $E_{ijk} = w_{ik} + e_{ijk}$, where w_{ik} denotes random variation among “e.u.” within A units (whole-plot units) and e_{ijk} denotes random variation among split-plot experimental units. The model for the design in Figure 4.1.e is similar, except that the whole plot is arranged in randomized blocks, so you add a block effect, r_k , just as you did in going from the CRD to the RCB. In the split plot you must also modify the whole-plot error term accordingly (i.e., change $eu(A)$ to $block*A$).

For the split block, shown in Figure 4.1.f, each factor has a different-size experimental unit. Thus, $E_{ijk} = r_k + w_{ik} + v_{jk} + e_{ijk}$, to reflect block effect and random variation among block $\times A$, block $\times B$, and block $\times A \times B$ units, respectively. The full model is thus

$$Y_{ijk} = \mu + \alpha_i + w_{ik} + \beta_j + v_{jk} + (\alpha\beta)_{ij} + e_{ijk}$$

The final model, shown in Figure 4.1.g, is a split-plot design with the whole plots arranged as Latin squares. Thus the model requires row and column blocking effects and the experimental units with respect to levels of A, row \times column combinations, and with respect to B (and hence A \times B), intersections of B with row \times column combinations. Thus, E_{ijk} can be written as $r_k + c_l + w_{ikl} + e_{ijkl}$, where r_k and c_l denote row and column effects, respectively. Note the additional subscript, l , to denote column position. The final model is thus

$$Y_{ijkl} = \mu + r_k + c_l + \alpha_i + w_{ikl} + \beta_j + (\alpha\beta)_{ij} + e_{ijkl}$$

Important Note: Beginning users often add inappropriate random effects to models such as those shown in this example. *Only random effects that correspond to actual physical units in the design should appear in the model.* For example, in the split-plot design with whole-plot blocking (Figure 4.1.e), a block \times A term (w_{ij}) appears in the model because it reflects the units assigned to levels of A. On the other hand, there is no block \times B term because it is not an experimental unit with respect to any treatment factor or effect. Note the difference between this model and the split-block model, where block \times B does appear in the model because it is the experimental unit for levels of B. *It is essential to maintain the relationship between the design and the way in which the experiment design (E_{ijk}) component is written in order for the model to be appropriate.*

This section has focused on a particular set of layouts for the 2 \times 2 factorial. The main lesson, however, is that you can apply the methods of this section to any, arbitrarily complex factorial study. The basic steps are:

1. Visualize the layout, as shown by example in Figure 4.1.
2. List the blocking criteria, if any.
3. List the treatment factor and all possible interactions.
4. Identify the experimental unit with respect to each treatment main effect and interaction.
5. Determine the fixed effects component of the model from step 3, and from step 2 for those blocking criteria that imply fixed effects.
6. Determine the random effects component of the model from step 4 (and from step 2 for random blocks). In general, there will be one random effect per experimental unit size.
7. If there are sampling units within the smallest experimental unit (e.g., if class is the experimental unit but measurements are taken on individual students within the class), then residual error corresponds to sampling unit error. Otherwise, residual error corresponds to the smallest experimental unit error and is not mentioned explicitly in the RANDOM statement.

Table 4.2 shows the CLASS, MODEL, and RANDOM statements to be used with PROC MIXED for each of the seven layouts described in Figure 4.1. Note that all the blocking criteria appear in the RANDOM statements. This follows the convention discussed in Chapter 2: in most cases, blocks are more naturally viewed as random effects. Note, however, that this is not a one-size-fits-all rule. There are cases in which the blocking criteria may be more appropriately regarded as fixed. Also, as discussed in Chapter 2, in the case of balanced experiments—i.e., complete block designs with no missing data—you get identical inference for treatment effects (any test, estimate, or contrast defined on A or B) regardless of whether blocks are defined as fixed or random. This is NOT the case with incomplete blocks, as in the example discussed in Section 4.7. In such cases, defining blocks as random allows you to recover intra-block information. Provided the blocks satisfy assumptions discussed in Chapter 2, recovery of intra-block information yields more efficient inference on treatment effects.

Table 4.2 CLASS, MODEL, and RANDOM Statements for Each Design Shown in Figure 4.1

Design	SAS PROC MIXED – CLASS, MODEL, and RANDOM Statements
CRD (Figure 4.1.a)	class a b; model y = a b a*b;
RCB (Figure 4.1.b)	class block a b; model y = a b a*b; random block;
Latin square (Figure 4.1.c)	class row col a b; model y = a b a*b; random row col;
Split-plot CR (Figure 4.1.d)	class eu a b; model y = a b a*b; random eu(a);
Split-plot RCB (Figure 4.1.e)	class block a b; model y = a b a*b; random block block*a;
Split block (Figure 4.1.f)	class block a b; model y = a b a*b; random block block*a block*b;
Split-plot LS (Figure 4.1.g)	class row col a b; model y=a b a*b; random row col row*col; <i>(or equivalently random row col row*col*a;)</i>

4.3 Inference with Mixed Models for Factorial Treatment Designs

The basic quantity of interest in any factorial treatment design is the $A \times B$ factorial mean, μ_{ij} . There are several terms defined on μ_{ij} : simple effects, main effects, marginal means, interactions, and slices. Regardless of the layout and resulting random effects, at least some of the fixed effect terms defined on μ_{ij} will be of interest. This section defines the various terms of interest, in terms of both the **mean model** and the **effects model** (μ_{ij} decomposed as $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$), relates them to the ESTIMATE, CONTRAST, and LSMEANS statements in PROC MIXED, and shows how their standard errors, test statistics, and degrees of freedom are obtained.

4.3.1 Effects of Interest in Factorial Treatment Designs

The following parameters are typically of interest in a factorial experiment, either as quantities to be estimated or as hypotheses to be tested. The first part of this section presents them in terms of the factorial treatment mean, μ_{ij} . The second part of this section presents the resulting expressions that result for these parameters in term of the mixed model and shows how standard errors, test statistics, and other considerations for statistical inference are derived.

4.3.1.1 Parameters of Interest

A **simple effect** is the difference between two levels of one factor with levels of another factor held constant. In the 2×2 factorial, these are

simple effect of A given $B_1 = \mu_{11} - \mu_{21}$, which may be denoted $A|B_1$
 simple effect of B given $A_1 = \mu_{11} - \mu_{12}$, which may be denoted $B|A_1$

A **marginal mean** is the arithmetic average of a given level of a factor averaged over all levels of the other factor. **Main effect mean** is another name for the marginal mean. In the 2×2 factorial, these are the marginal mean of A_1 ,

$$\mu_{1\bar{1}} = (1/b) \sum_{j=1}^b \mu_{ij}$$

where b is the number of levels of factor B ($b = 2$ in a 2×2 factorial); and the marginal mean of B_1 ,

$$\mu_{\bar{1}\bar{1}} = (1/a) \sum_{i=1}^a \mu_{ij}$$

where a is the number of levels of factor A ($a = 2$ in a 2×2 factorial).

A **main effect** is the difference between marginal means of two levels of a factor. In the 2×2 factorial, these are

$$\text{main effect of } A = \mu_{1\bar{1}} - \mu_{2\bar{1}}$$

$$\text{main effect of } B = \mu_{11} - \mu_{12}$$

If a factor has more than two levels, the overall test of main effect is H_0 : “all marginal means equal,” e.g., main effect of A: H_0 : “all $\mu_{i\bar{1}}$ equal, $i=1, 2, \dots, a$.”

The **interaction** concerns equality of simple effects. In a 2×2 factorial, interaction is the difference between simple effects given the two levels of the other factor; i.e.,

$$\mu_{11} - \mu_{12} - (\mu_{21} - \mu_{22}) = \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$$

The hypothesis of no interaction states that simple effects of one factor are equal given all levels of the other factor, e.g., for a 2×2 factorial, H_0 : $A|B_1 = A|B_2$ or, equivalently, H_0 : $B|A_1 = B|A_2$. Note that this is equivalent to H_0 : $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$.

The **slice** tests the equality of simple effects of one factor for a given level of the other factor. For a 2×2 factorial, the slice is a single treatment difference, i.e., the simple effect defined above. For a factor with more than two levels, the slice tests the equality of all simple effects; e.g., for an $a \times 2$ factorial (a levels of factor A, 2 levels of factor B), the slice given B tests the following two hypotheses:

$$\text{Slice given } B_1: \quad H_0: \mu_{11} = \mu_{21} = \dots = \mu_{a1}$$

$$\text{Slice given } B_2: \quad H_0: \mu_{12} = \mu_{22} = \dots = \mu_{a2}$$

Contrasts allow you to assess linear combinations of treatment means. For the treatment means, μ_{ij} , a **linear combination** is defined as $\sum_{i,j} c_{ij} \mu_{ij}$. The terms c_{ij} are called **coefficients**. A contrast is a linear combination, $\sum_{i,j} c_{ij} \mu_{ij}$, where the coefficients obey the restrictions

$$\sum_i c_{ij} = \sum_j c_{ij} = \sum_{i,j} c_{ij} = 0$$

Contrasts are thus mean comparisons. If $c_{11} = 1$, $c_{12} = -1$, and $c_{21} = c_{22} = 0$, then the contrast defines the simple effect $B|A_1$. The coefficients $c_{11} = c_{22} = 1$ and $c_{12} = c_{21} = -1$ define the interaction. The main effect of A is defined by the coefficients $c_{11} = c_{12} = 1$ and $c_{21} = c_{22} = -1$, or, more precisely, $c_{11} = c_{12} = \frac{1}{2}$ and $c_{21} = c_{22} = -\frac{1}{2}$.

You can also define the main effect contrast on the marginal means. In fact, it is more typical to do so. For example, for A, a contrast is $\sum_i c_i \mu_{i\cdot}$ provided $\sum_i c_i = 0$.

In general, simple effects, slices, and contrasts defined on treatment combination means are of interest if the interaction effect is nonzero, or more accurately, non-negligible from a practical viewpoint. Marginal means and main effects (including main effect contrasts) are of interest *only if* interaction has been shown to be negligible from any practical viewpoint. Therefore, with rare exceptions, you should start the analysis of factorial experiments by testing interaction.

4.3.2 More on Effects of Interest—Estimability and Use of ESTIMATE, CONTRAST, and LSMEANS Statements

You can express simple effects, main effects, interactions, and contrasts in terms of the treatment mean model, as in the previous section, or you can express them in terms of the *effects model*, as in Section 4.2.2. Obtaining the statistics you need to estimate means and perform mean comparisons requires the ESTIMATE, CONTRAST, and LSMEANS statements. The effects model helps you to use these statements more effectively. This section shows the decomposition of simple and main effects, interactions, and contrasts with the effects model and how you translate these decompositions for use with PROC MIXED statements. You can also use these decompositions with PROC GLM.

4.3.2.1 Simple Effects

As an example, the simple effect $B|A_1$ was defined in the means model as $\mu_{11} - \mu_{12}$. In terms of the effects model, you re-express this as

$$[\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}] - [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}] = \beta_1 - \beta_2 + (\alpha\beta)_{11} - (\alpha\beta)_{12}$$

Similarly, the simple effect $A|B_2$ is $a_1 - a_2 + (\alpha\beta)_{11} - (\alpha\beta)_{21}$. These coefficients allow you translate directly into the ESTIMATE statement in either PROC MIXED or PROC GLM. For $B|A_1$ the needed SAS statements are as follows:

```
proc mixed;
  class a b;
  model y=a b a*b;
  estimate 'simple effect b given a_1'  b 1 -1 a*b 1 -1 0 0;
run;
```

Two important notes to address common problems for beginning users:

- The order of the coefficients follows from the order of the effects in the CLASS statement: CLASS A B means that SAS assigns coefficients for the levels of B nested within the first level of A first, the levels of B within the second level of A, and so forth if there are more than two levels of A. You can change the program to CLASS B A, but you must change the order of the coefficients for the A \times B effect as well.
- Also, one of the most common problems encountered by new PROC GLM and PROC MIXED users is including the coefficients for A \times B but omitting the coefficients for B. You have to include *all* of the terms in the effects model, or the relationship between the means and the effects model will be incompletely specified and the expression is therefore nonestimable.

PROC GLIMMIX simplifies comparing simple effects with the LSMESTIMATE statement. This statement is not available in PROC MIXED or PROC GLM. GLIMMIX is a new procedure as of this printing. Section 4.4.4 provides a brief introduction, including the LSMESTIMATE statement, of the GLIMMIX procedure for split-plot experiments.

4.3.2.2 Main Effects

As an example, the simple effect of A requires the marginal means $\mu_{1\cdot}$ and $\mu_{2\cdot}$. The marginal mean

$$\mu_{1\cdot} = \mu + \alpha_1 + \bar{\beta}_{\cdot} + (\bar{\alpha}\bar{\beta})_{1\cdot}$$

where $\bar{\beta}_{\cdot} = (\frac{1}{b}) \sum_j \beta_j$, the mean of the B effects, and $(\bar{\alpha}\bar{\beta})_{1\cdot} = (\frac{1}{b}) \sum_j (\alpha\beta)_{1j}$, the average of the A \times B effects within level A₁. Note that $\sum_j \beta_j$ and $\sum_j (\alpha\beta)_{1j}$ do not, in general, sum to zero.

Using the effects model, the main effect of A is thus

$$\mu_{1\cdot} - \mu_{2\cdot} = \left[\mu + \alpha_1 + \bar{\beta}_{\cdot} + (\bar{\alpha}\bar{\beta})_{1\cdot} \right] - \left[\mu + \alpha_2 + \bar{\beta}_{\cdot} + (\bar{\alpha}\bar{\beta})_{2\cdot} \right] = \alpha_1 - \alpha_2 + (\bar{\alpha}\bar{\beta})_{1\cdot} - (\bar{\alpha}\bar{\beta})_{2\cdot}$$

A similar derivation yields the main effect of B, $\beta_1 - \beta_2 + (\bar{\alpha}\bar{\beta})_{1\cdot} - (\bar{\alpha}\bar{\beta})_{2\cdot}$. You can obtain these main effects using the ESTIMATE statement in PROC GLM or PROC MIXED as follows:

```
proc mixed;
  class a b;
  model y=a b a*b;
  estimate 'main effect of a'  a 1 -1 a*b 0.5 0.5 -0.5 -0.5;
  estimate 'main effect of b'  b 2 -2 a*b 1 -1 1 -1/divisor=2;
run;
```

The DIVISOR option divides all of the coefficients by the number to the right of the equal sign. For the 2 \times 2 factorial, both specifications are equally easy, but for more levels, the DIVISOR option is very useful. The ESTIMATE statement by default computes the coefficients for higher-order effects that contain effects already mentioned in the statement (unless you provide coefficients of these higher-order effects). Thus, the statements

```
estimate 'main effect of a'  a 1 -1;
estimate 'main effect of b'  b 1 -1;
```

give you the same result as the ESTIMATE statements above that explicitly give the A×B coefficients. If you add the E option to the ESTIMATE statement, the coefficients for all effects are listed, so you can check how the procedure “fills in” coefficients for effects not specified in the ESTIMATE statement. Note that this fill-in does not work for the simple effects, because the default works in only one direction: you could not give coefficients for A×B and have anything assumed about A or B. You can also obtain main effects by using the LSMEANS statement

```
lsmeans a b / diff;
```

4.3.2.3 Interactions and Contrasts

The interaction was defined by the difference between simple effects, $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$. You can re-express this in the effects model as

$$\begin{aligned} & [\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}] - [\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}] \\ & - [\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}] + [\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}] = (\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{22} \end{aligned}$$

You can obtain the needed statistics by adding either of the following statements to the SAS program above:

```
estimate 'a x b interaction' a*b 1 -1 -1 1;
contrast 'a x b interaction' a*b 1 -1 -1 1;
```

The ESTIMATE statement gives you a *t*-statistic to test the interaction and the estimated difference between the two simple effects. The CONTRAST statement gives you only the *F*-statistic, i.e., the square of the *t*-statistic given in the ESTIMATE statement, but no estimate of the simple effect difference.

The contrast statement has the general form $\sum_{i,j} c_{ij} \mu_{ij}$, which can be re-expressed in the effects model as $\sum_{i,j} c_{ij} [\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}] = \sum_{i,j} c_{ij} [\alpha_i + \beta_j + (\alpha\beta)_{ij}]$. Note that $\sum_{i,j} c_{ij} \mu$ always equals zero. The terms $\sum_{i,j} c_{ij} \alpha_i$ and $\sum_{i,j} c_{ij} \beta_j$ may equal zero, as they do for the interaction, but they may not, as was the case for simple effects. Once you determine the coefficients, you can put them in the CONTRAST or ESTIMATE statements as with the other examples in this section. If your contrast is defined on a main effect, the A×B terms contained in the effects model expression of the contrast will be assumed by default.

4.3.3 Standard Errors

As shown in Sections 4.3.1 and 4.3.2, the fixed effect component of the model concerned with the treatment effects, in either its mean form or effects form, is the same for all factorial treatment structures. However, the random effect component of the model depends on the layout of the design and the underlying variability among the various experimental units, as shown in Section 4.2. This, in turn, affects the standard errors associated with the various effects. You use the standard errors of simple and main effects, interactions and contrasts, to construct their confidence intervals and test statistics. This section shows how standard errors are derived from the experiment design and the assumed sources of variation.

You can apply the methods shown in this section to any mixed model that follows from layouts such as those shown in Figure 4.1. The specific results vary depending on the random effects

implied by the layout. Space does not permit showing the results for all possible layouts. This section focuses on the split plot with randomized block whole plot shown in Figure 4.1.e. It is the simplest layout with the full array of issues that arise in mixed model analysis. While the results shown in this section specifically apply to the model that follows from Figure 4.1.e, the methods shown here can be adapted to any layout.

As shown in Section 4.2, you can write a mixed model for Figure 4.1.e in two forms:

$$\begin{aligned} \text{mean model: } Y_{ijk} &= \mu_{ij} + r_k + w_{ik} + e_{ijk} \\ \text{effects model: } Y_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + r_k + w_{ik} + e_{ijk} \end{aligned}$$

The effects model may be reordered as $Y_{ijk} = \mu + r_k + \alpha_i + w_{ik} + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$ —that is, listing the “whole-plot” elements first, then the split-plot elements. In this section, assume that the block effects, r_k , the whole-plot error effects, w_{ik} , and the split-plot error effects, e_{ijk} , are all random effects, with the following assumptions:

- All random effects, r_k , w_{ik} , and e_{ijk} are mutually independent.
- The block effects are assumed $iid N(0, \sigma_R^2)$.
- The whole-plot error effects are assumed $iid N(0, \sigma_W^2)$.
- The split-plot error effects are assumed $iid N(0, \sigma^2)$.

Under these assumptions, the standard errors of the estimates of the various terms of interest discussed in Sections 4.3.1 and 4.3.2 can be obtained. Recall that the **standard error** is the **estimate** of the square root of the variance of the estimate of the term of interest. Hence, there are four steps in determining the standard error:

- Write the estimate of the term of interest (simple or main effect, interaction or contrast).
- Derive the variance of the estimate from step 1. This will be a function of the variances of the random model effects (σ_R^2 , σ_W^2 , and σ^2 in this example), called the variance components.
- Determine the estimates of the variance components σ_R^2 , σ_W^2 , and σ^2 .
- Substitute the estimated variance components into the expression in step 2 and take the square root.

First, consider steps 1 and 2.

4.3.3.1 Variance of Treatment Mean and Difference Estimates

For treatment combination means, μ_{ij} , the best estimate is $\hat{\mu}_{ij} = \frac{1}{r} \sum_k y_{ijk} = \bar{y}_{ij\bar{k}}$. In model terms, the variance of the estimate is

$$\begin{aligned} \text{Var}[\bar{y}_{ij\bar{k}}] &= \text{Var}\left[\left(\frac{1}{r}\right) \sum_k \mu_{ij} + r_k + w_{ik} + e_{ijk}\right] \\ &= \left(\frac{1}{r}\right)^2 \sum_k \text{Var}[r_k + w_{ik} + e_{ijk}] \\ &= \left(\frac{1}{r}\right)^2 \sum_k [\sigma_R^2 + \sigma_W^2 + \sigma^2] = \frac{\sigma_R^2 + \sigma_W^2 + \sigma^2}{r} \end{aligned}$$

Similar derivations yield

$$\text{Var}[\bar{y}_{i\perp}] = \frac{b(\sigma_k^2 + \sigma_w^2) + \sigma^2}{br} \text{ for the estimated marginal mean of A}_i$$

$$\text{Var}[\bar{y}_{j\perp}] = \frac{b\sigma_k^2 + \sigma_w^2 + \sigma^2}{br} \text{ for the estimated marginal mean of B}_j.$$

For the main effect of A, e.g., $\bar{\mu}_1 - \bar{\mu}_2$, the variance of the estimate is

$$\begin{aligned} \text{Var}[\bar{y}_{1\perp} - \bar{y}_{2\perp}] &= \text{Var}\left[\frac{1}{br} \sum_{j,k} (y_{1jk} - y_{2jk})\right] \\ &= \left(\frac{1}{br}\right)^2 \text{Var}\left[\sum_{j,k} (\mu_{1j} + r_k + w_{1k} + e_{1jk}) - (\mu_{2j} + r_k + w_{2k} + e_{2jk})\right] \\ &= \left(\frac{1}{br}\right)^2 \text{Var}\left[\sum_{j,k} (\mu_{1k} - \mu_{2k} + w_{1k} - w_{2k} + e_{1jk} - e_{2jk})\right] \\ &= \left(\frac{1}{br}\right)^2 \text{Var}\left[b \sum_j (w_{1k} - w_{2k}) - \sum_{j,k} (e_{1jk} - e_{2jk})\right] \\ &= \left(\frac{1}{br}\right)^2 (2b^2 r \sigma_w^2 + 2br \sigma^2) \\ &= \frac{2(b\sigma_w^2 + \sigma^2)}{br} \end{aligned}$$

Similar derivation yields the variance of the main effect of B, e.g., $\bar{\mu}_1 - \bar{\mu}_2$:

$$\begin{aligned} \text{Var}[\bar{y}_{1\perp} - \bar{y}_{2\perp}] &= \text{Var}\left[\left(\frac{1}{ar}\right) \sum_{i,k} (y_{i1k} - y_{i2k})\right] \\ &= \left(\frac{1}{ar}\right)^2 \text{Var}\left[\sum_{i,k} (\mu_{i1} + r_k + w_{ik} + e_{i1k} - \mu_{i2} - r_k - w_{ik} - e_{i2k})\right] \\ &= \left(\frac{1}{ar}\right)^2 \text{Var}\left[\sum_{i,k} (e_{i1k} - e_{i2k})\right] \\ &= \frac{2\sigma^2}{ar} \end{aligned}$$

Note that the whole-plot variance, σ_w^2 , appears in the main effect of A but not in the main effect of B. This is because you estimate A differences across whole plots, whereas you estimate B differences within whole plots, so the whole-plot variation is not part of the estimate.

Similar derivations yield the variances of simple effect estimators. For the simple effect of A given B, e.g., $\mu_{11} - \mu_{21}$, the variance of the estimator is

$$\begin{aligned}
\text{Var}[\bar{y}_{110} - \bar{y}_{120}] &= \text{Var}\left[\left(\frac{1}{r}\right)\sum_k(y_{11k} - y_{12k})\right] \\
&= \left(\frac{1}{r}\right)^2 \text{Var}\left[\sum_k(\mu_{11} + r_k + w_{1k} + e_{11k} - \mu_{12} - r_k - w_{1k} - e_{12k})\right] \\
&= \left(\frac{1}{r}\right)^2 \text{Var}\left[\sum_k(w_{1k} - w_{2k}) + \sum_k(e_{11k} - e_{12k})\right] \\
&= \frac{2(\sigma_w^2 + \sigma_e^2)}{r}
\end{aligned}$$

On the other hand, the simple effect of B given A, e.g., $\mu_{11} - \mu_{12}$, yields a different variance of the estimate:

$$\begin{aligned}
\text{Var}[\bar{y}_{110} - \bar{y}_{210}] &= \text{Var}\left[\left(\frac{1}{r}\right)\sum_k(y_{11k} - y_{21k})\right] \\
&= \left(\frac{1}{r}\right)^2 \text{Var}\left[\sum_k(\mu_{11} + r_k + w_{1k} + e_{11k} - \mu_{21} - r_k - w_{2k} - e_{21k})\right] \\
&= \left(\frac{1}{r}\right)^2 \text{Var}\left[\sum_k(e_{11k} - e_{12k})\right] \\
&= \frac{2\sigma_e^2}{r}
\end{aligned}$$

Simple effects of A given B must be estimated from differences between observations in different whole plots, whereas B given A effects can be estimated from differences between observations entirely within the same whole plot.

4.3.3.2 Generalization: Variance of Estimates in Matrix Form

You can apply similar derivations to obtain standard errors for estimates of interactions, contrasts, and other terms of interest. These are not shown here. In general, estimates of functions of the μ_{ij} and their variances can be determined from the mixed model equations and their properties (see Appendix 1 for details). In this example, the matrix form of the model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}_1\mathbf{r} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}$$

where

\mathbf{Y} is the vector of observations

$\boldsymbol{\mu} = [\mu_{11} \ \mu_{12} \ \mu_{21} \ \mu_{22}]'$, the vector of A×B means

$\mathbf{r} = [r_1 \ r_2 \ \dots \ r_r]'$, the vector of block effects

$\mathbf{w} = [w_{11} \ w_{12} \ \dots \ w_{1r} \ w_{21} \ \dots \ w_{2r} \ \dots \ w_{41} \ \dots \ w_{4r}]'$, the vector of whole-plot errors

\mathbf{e} is the vector of split-plot errors

\mathbf{X} is the design matrix with respect to A×B treatment combinations

\mathbf{Z}_1 is the design matrix with respect to blocks

\mathbf{Z}_2 is the design matrix with respect to whole-plot errors

In matrix form, the model assumptions are

\mathbf{r} is distributed MVN(0, $\sigma_r^2 \mathbf{I}$)

\mathbf{w} is distributed MVN(0, $\sigma_w^2 \mathbf{I}$)

\mathbf{e} is distributed MVN(0, $\sigma^2 \mathbf{I}$)

The resulting distribution of \mathbf{Y} is MVN($\mathbf{X}\boldsymbol{\mu}, \mathbf{V}$), where $\mathbf{V} = \mathbf{Z}_1(\sigma_r^2 \mathbf{I})\mathbf{Z}_1' + \mathbf{Z}_2(\sigma_w^2 \mathbf{I})\mathbf{Z}_2' + \sigma^2 \mathbf{I}$.

In this model, $\mathbf{X}\boldsymbol{\mu}$ could be expressed in terms of the fixed effects. That is,

$$\mathbf{X}\boldsymbol{\mu} = \mathbf{1}\boldsymbol{\mu} + \mathbf{X}_A\boldsymbol{\alpha} + \mathbf{X}_B\boldsymbol{\beta} + \mathbf{X}_{AB}(\boldsymbol{\alpha}\boldsymbol{\beta})$$

where

$\mathbf{1}$ is a vector of ones

$\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2]'$, the vector of A effects

$\boldsymbol{\beta} = [\beta_1 \ \beta_2]'$, the vector of B effects

$(\boldsymbol{\alpha}\boldsymbol{\beta}) = [(\boldsymbol{\alpha}\boldsymbol{\beta})_{11} \ (\boldsymbol{\alpha}\boldsymbol{\beta})_{12} \ (\boldsymbol{\alpha}\boldsymbol{\beta})_{21} \ (\boldsymbol{\alpha}\boldsymbol{\beta})_{22}]'$, the vector of A×B effects

\mathbf{X}_A , \mathbf{X}_B , and \mathbf{X}_{AB} are the design matrices with respect to A, B, and A×B, respectively

The solution to the mixed model equations provides best linear unbiased estimates of the fixed effects, $\boldsymbol{\mu}$, or alternatively, **estimable functions** of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $(\boldsymbol{\alpha}\boldsymbol{\beta})$. For any linear combination $\mathbf{K}'\boldsymbol{\mu}$ (or estimable linear combination of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $(\boldsymbol{\alpha}\boldsymbol{\beta})$), the variance of the estimate is $\mathbf{K}'[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{K}$.

For example, for a main effect mean for the first level of A,

$$\mu_{11} = (1/2)(\mu_{11} + \mu_{12}) = (1/2)[1 \ 1 \ 0 \ 0]'\boldsymbol{\mu}$$

Thus $\mathbf{K}' = (1/2)[1 \ 1 \ 0 \ 0]'$. Applying the formula $\mathbf{K}'[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{K}$ yields

$$\text{Var}[\hat{\mu}_{11}] = \text{Var}\left[1/2[1 \ 1 \ 0 \ 0]'\boldsymbol{\mu}\right] = \frac{1}{br}(2\sigma_w^2 + \sigma^2)$$

This approach can also be used for a treatment difference. For example, for the main effect difference $\mu_{11} - \mu_{21}$, $\mathbf{K}'\boldsymbol{\mu}$ is

$$(1/2)[1 \ 1 \ -1 \ -1]'\boldsymbol{\mu}$$

Applying the variance formula, the variance of $\hat{\mu}_{11} - \hat{\mu}_{21}$ yields $(2/b)r[2\sigma_w^2 + \sigma^2]$.

4.3.3.3 Completing the Standard Error: Variance Component Estimates and Degrees of Freedom

From the variance of the treatment mean, difference, or contrast, you can do steps 3 and 4 described above—that is, obtain the standard error by substituting the estimates of the variance components into the variance and taking the square root. Applying these results allows **inference** in the form of either **confidence interval estimation** or **hypothesis testing**.

For *confidence intervals*, the general formula for linear combinations of the treatment combination means is **estimate $\pm t_{\nu} \times standard\ error(estimate)$** , or in matrix terms:

$$\mathbf{K}'\hat{\boldsymbol{\mu}} \pm t_{\nu} \sqrt{\mathbf{K}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{K}}$$

where $\hat{\mathbf{V}}$ denotes the estimate of the variance matrix \mathbf{V} . The degrees of freedom, ν , are determined by the degrees of freedom required to estimate the linear combination of variance components in $\mathbf{K}'[\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}]^{-1}\mathbf{K}$. Variance component estimation and degrees of freedom are discussed immediately after the next paragraph.

For *hypothesis testing*, the general approach for linear combinations of treatment means is via the ratio

$$\frac{estimate}{std\ error(estimate)}$$

or, in matrix terms,

$$\frac{\mathbf{K}'\hat{\boldsymbol{\mu}}}{\sqrt{\mathbf{K}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{K}}}$$

These ratios form t -statistics with ν degrees of freedom. Alternatively, you can square these ratios to form the F -statistic with 1 degree of freedom in the numerator and ν denominator degrees of freedom. The ESTIMATE statement in PROC MIXED computes the t -statistic as well as the standard error; the CONTRAST statement computes only the F -statistic. You can also use the CONTRAST statement to compute the general form of the F -statistic with k numerator degrees of freedom. In the 2×2 factorial considered in this section, such hypotheses do not typically arise. However, Section 4.4 presents an example with more than two levels of a factor. The F -statistic with k numerator and ν denominator degrees of freedom will be discussed in greater detail.

To obtain confidence intervals or test hypotheses, you need estimates of the components of variance. PROC MIXED allows you to obtain variance component estimates in several ways. The default is **restricted maximum likelihood** (hereafter referred to by its commonly used acronym, **REML**), described in Appendix 1. For the purposes of this discussion, estimates based on the **analysis of variance** (ANOVA) are easier to describe; for designs described in this chapter that are **balanced** (equal number of observations per $A \times B \times$ block combination), REML and ANOVA variance component estimates are identical.

For the example design based on the layout in Figure 4.1.e, the analysis of variance is shown in Table 4.3.

Table 4.3 Analysis of Variance for Layout in Figure 4.1.e

Source of Variation	d.f.	Expected Mean Square
BLOCK	3	
A	1	$\sigma^2 + 2\sigma_w^2 + \phi_A$
BLOCK \times A (whole-plot error)	3	$\sigma^2 + 2\sigma_w^2$
B	1	$\sigma^2 + \phi_B$
A \times B	1	$\sigma^2 + \phi_{A \times B}$
split-plot error	6	σ^2

In Table 4.3, the ϕ are quadratic expressions for the fixed effects and have no bearing on variance component estimation. Littell et al. (2002) discuss them in detail. Also, in this discussion, the expected mean square for BLOCK is not shown, as the variance for BLOCK is not of interest. The whole-plot and split-plot error variance components can be estimated as

$$\hat{\sigma}^2 = \text{MS(Split-Plot Error), or MS(SPE) for convenience}$$

$$\hat{\sigma}_w^2 = 1/2(\text{MS(BLOCK} \times \text{A}) - \text{MS(SPE)})$$

For a whole-plot main effect mean, μ_{i1} , the variance is $(1/8)[2\sigma_w^2 + \sigma^2]$, which is estimated as $(1/8)\text{MS[BLOCK} \times \text{A]}$. Thus, the degrees of freedom for a confidence interval correspond to the degrees of freedom for BLOCK \times A, $v=3$.

For a split-plot main effect difference, $\mu_{11} - \mu_{21}$, the variance is $(1/4)\sigma^2$, which is estimated by $(1/4)\text{MS(SPE)}$. Thus, the degrees of freedom for the confidence interval correspond to the degrees of freedom for split-plot error, $v=6$.

As a final example, consider the simple effect difference $\mu_{11} - \mu_{21}$, whose variance is $(1/2)[\sigma_w^2 + \sigma^2]$. The linear combination $\sigma_w^2 + \sigma^2$ cannot be estimated by a single mean square. Instead, it is estimated by

$$\hat{\sigma}_w^2 + \hat{\sigma}^2 = \text{MS(SPE)} + \frac{1}{2}(\text{MS(BLOCK} \times \text{A}) - \text{MS(SPE)})$$

$$= \frac{1}{2}\text{MS(SPE)} + \frac{1}{2}\text{MS(BLOCK} \times \text{A})$$

The degrees of freedom can be approximated using Satterthwaite's formula. Let

$\text{MS} = a_1\text{MS}_1 + \dots + a_k\text{MS}_k$. The approximate degrees of freedom for MS are

$$v \cong \frac{(\text{MS})^2}{\frac{(a_1\text{MS}_1)^2}{df_1} + \dots + \frac{(a_k\text{MS}_k)^2}{df_k}}$$

Thus, the approximate degrees of freedom for the simple effect of A for a given B are

$$v \cong \frac{\left(\frac{1}{2} MS(SPE) + \frac{1}{2} MS(BLOCK \times A)\right)^2}{\frac{\left(\frac{1}{2} MS(SPE)\right)^2}{6} + \frac{\left(\frac{1}{2} MS(BLOCK \times A)\right)^2}{3}}$$

4.4 Example: A Split-Plot Semiconductor Experiment

This example, which appeared in Littell et al. (1991, 1996, and 2002), is an experiment conducted in a semiconductor plant to study the effect of several modes of a process condition (ET) on resistance in computer chips. Twelve silicon wafers (WAFER) were drawn from a lot, and three wafers were randomly assigned to four modes of ET. Resistance in the chips was measured on chips at four different positions (POS) on each wafer after processing. The measurement was recorded as the variable RESISTANCE. The data are given as Data Set 4.4, “Semiconductor Split-Plot Experiment,” in Appendix 2, “Data Sets.”

The semiconductor experiment consists of two factors, ET and POS. The experimental unit with respect to ET is the wafer. The experimental unit with respect to POS is the individual chip, a subdivision of the wafer. Thus, the wafer is the whole-plot unit, and the chip is the split-plot unit.

The layout of this experiment is similar to that of Figure 4.1.d, the only difference being the number of levels per factor. Thus, following Table 4.1, a model for this experiment is

$$Y_{ijk} = \mu + \alpha_i + w_{ik} + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

where

$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ is the mean of the ij^{th} ET \times POS combination, and α_i , β_j , and $(\alpha\beta)_{ij}$ are the ET, POS, and ET \times POS effects, respectively

w_{ik} is the whole-plot error effect, assumed $iid N(0, \sigma_w^2)$

e_{ijk} is the split-plot error effect, assumed $iid N(0, \sigma^2)$

w_{ik} and e_{ijk} are assumed to be independent of one another

Littell et al. (1991) mention that correlation among the chips may actually depend on their proximity on the wafer, thus violating the assumption of independence among split-plot errors. If so, the data may be analyzed by revising the model to assume spatially dependent correlated errors. This is discussed in Chapter 5, “Analysis of Repeated Measures Data,” and Chapter 11, “Spatial Variability.”

4.4.1 Tests of Interest in the Semiconductor Experiment

Several hypotheses are potentially of interest in this experiment. These include:

- No ET \times POS interaction: $H_0: \mu_{ij} = \mu_{ij'}$ for all i , given $j \neq j'$; i.e., all $(\alpha\beta)_{ij} = 0$
- All ET main effect means equal: $H_0: \mu_{1\cdot} = \mu_{2\cdot} = \dots = \mu_{4\cdot}$
- All POS main effect means equal: $H_0: \mu_{\cdot 1} = \mu_{\cdot 2} = \dots = \mu_{\cdot 4}$

- Simple effect comparisons among specific ET levels for a given POS, e.g.,
 $H_0: \mu_{11} - \mu_{21} = 0$
- Simple effect comparisons among specific positions for a given ET level, e.g.,
 $H_0: \mu_{11} - \mu_{12} = 0$
- Contrasts, e.g., average of ET1 and ET2 versus average of ET3 and ET4 given POS2,
 $H_0: \frac{1}{2}(\mu_{12} + \mu_{22}) - \frac{1}{2}(\mu_{32} + \mu_{42}) = 0$

The first three hypotheses are tested using F -ratios from the analysis of variance, shown in Table 4.4.

Table 4.4 Analysis of Variance for Semiconductor Example

Source of Variation	d.f.	Expected Mean Square
ET	3	$\sigma^2 + 4\sigma_w^2 + \phi_{ET}$
WAFER(ET) (whole-plot error)	8	$\sigma^2 + 4\sigma_w^2$
POS	3	$\sigma^2 + \phi_{POS}$
ET \times POS	9	$\sigma^2 + \phi_{ET*POS}$
Split-plot error (SPE)	24	σ^2

The variance components can be estimated as

$$\hat{\sigma}^2 = \text{MS}(\text{SPE})$$

$$\hat{\sigma}_w^2 = 1/4(\text{MS}[\text{WAFER}(ET)] - \text{MS}(\text{SPE}))$$

From the ANOVA table for the semiconductor experiment, the required test statistics are

MS(ET) / MS[WAFER(ET)]	for the ET main effect
MS(POS) / MS(SPE)	for the POS main effect
MS(ET \times POS) / MS(SPE)	for the ET \times POS interaction

The remaining hypotheses are all single-degree-of-freedom contrasts. They can be tested using the methods shown in Section 4.3. The same degree-of-freedom issues discussed in Section 4.3.3.3 arise here. For example, the main effect comparison of ET1 versus ET2 can be tested using the ratio

$$t = \frac{\hat{\mu}_{1\cdot} - \hat{\mu}_{2\cdot}}{\sqrt{\frac{1}{12}(\hat{\sigma}^2 + 4\hat{\sigma}_w^2)}} = \frac{\hat{\mu}_{1\cdot} - \hat{\mu}_{2\cdot}}{\sqrt{\frac{1}{12}\text{MS}[\text{WAFER}(ET)]}}$$

Thus, the degrees of freedom for the t-test correspond to the degrees of freedom for WAFER(ET), $v = 8$.

In many cases, the denominator term of the t -statistic is a linear combination of mean squares. If so, the degrees of freedom for the t -test must be approximated, e.g., using Satterthwaite's method. For example, the t -ratio for the test of ET1 and ET2 versus ET3 and ET4 in POS2 is

$$t = \frac{\hat{\mu}_{12} + \hat{\mu}_{22} - \hat{\mu}_{32} - \hat{\mu}_{42}}{\sqrt{\frac{4}{3}(\hat{\sigma}^2 + \hat{\sigma}_w^2)}}$$

The denominator $\hat{\sigma}^2 + \hat{\sigma}_w^2$ is estimated by $MS(SPE) + (1/4)\{MS[WAFER(ET)] - MS(SPE)\} = (1/4)MS[WAFER(ET)] + (3/4)MS(SPE)$. The Satterthwaite approximation yields

$$v \approx \frac{\left(\frac{3}{4}MS(SPE) + \frac{1}{4}MS[WAFER(ET)]\right)^2}{\frac{\left(\frac{3}{4}MS(SPE)\right)^2}{24} + \frac{\left(\frac{1}{4}MS[WAFER(ET)]\right)^2}{8}}$$

Alternatively, the single-degree-of-freedom hypotheses can be tested using F -ratios of the form

$$F = t^2 = \frac{estimate^2}{std\ err^2} = \frac{MS(\text{contrast})}{MS(\text{relevant error})}$$

where $MS(\text{relevant error})$ is the appropriate mean square or linear combination of mean squares. Thus, the F -ratio for the main effect of ET1 versus ET2 is

$$F = \frac{MS(\text{ET1 versus ET2})}{MS(WAFER(ET))}$$

The numerator has one degree of freedom and the denominator has $v = 8$, corresponding to the degrees of freedom for WAFER(ET). For the test of ET1 and ET2 versus ET3 and ET4 in POS2, the F -ratio is

$$F = \frac{MS(\text{ET1, ET2 versus ET3, ET4 given POS2})}{\frac{1}{4}MS[WAFER(ET)] + \frac{3}{4}MS(SPE)}$$

There is one degree of freedom for the numerator. The denominator degrees of freedom are determined from the Satterthwaite approximation for the linear combination of $MS[WAFER(ET)]$ and $MS(SPE)$.

4.4.2 Matrix Generalization of Mixed Model F -tests

All of the preceding F -ratios, for the overall main effects, interactions, and single-degree-of-freedom tests, are special cases of a general result for mixed models. In general, you can test hypotheses with one or more degrees of freedom by defining a \mathbf{K} matrix consisting of any number of *independent* \mathbf{k} vectors.

Common examples of multiple-degree-of-freedom hypotheses are overall main effect or interaction terms in the analysis of variance. For example, consider the main effect of ET.

The overall hypothesis of no ET main effect is

$$H_0: \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$$

This can be expressed in terms of a \mathbf{K} matrix several ways. Because there are 4 levels of ET, there are 3 degrees of freedom for the ET main effect. Therefore, \mathbf{K} consists of 3 independent \mathbf{k} vectors, one per degree of freedom. *Any set of 3 comparisons such that H_0 is true* can be used to form \mathbf{K} . For example,

$$\mathbf{K} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}$$

defines the comparisons $\mu_{1\Box} = \mu_{2\Box}$, $\mu_{1\Box} = \mu_{3\Box}$, and $\mu_{1\Box} = \mu_{4\Box}$. If all three equalities hold, then $\mu_{1\Box} = \mu_{2\Box} = \mu_{3\Box} = \mu_{4\Box}$, i.e., H_0 is true. On the other hand,

$$\mathbf{K}' = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -2 & -2 & -2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 & -3 & -3 & -3 \end{bmatrix}$$

defines the comparisons $\mu_{1\Box} = \mu_{2\Box}$, $1/2(\mu_{1\Box} + \mu_{2\Box}) = \mu_{3\Box}$, and $1/3(\mu_{1\Box} + \mu_{2\Box} + \mu_{3\Box}) = \mu_{4\Box}$. If these three equalities hold, they *also* guarantee that $H_0: \mu_{1\Box} = \mu_{2\Box} = \mu_{3\Box} = \mu_{4\Box}$ is true. For the main effect of ET, \mathbf{K} can be composed of any three comparisons that imply $\mu_{1\Box} = \mu_{2\Box} = \mu_{3\Box} = \mu_{4\Box}$. For any \mathbf{K} so constructed, the resulting F -statistic, discussed below, will be the same.

For any linear combination $\mathbf{K}'\hat{\boldsymbol{\mu}}$ (or estimable linear combination of the fixed effects μ , $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $(\boldsymbol{\alpha}\boldsymbol{\beta})$),

$$(\mathbf{K}'\hat{\boldsymbol{\mu}})' \left[\mathbf{K}' (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{K} \right]^{-1} (\mathbf{K}'\hat{\boldsymbol{\mu}}) / \text{rank}(\mathbf{K})$$

has an approximate chi-square distribution with $\text{rank}(\mathbf{K})$ degrees of freedom *when \mathbf{V} is known*. When \mathbf{V} is unknown and the components of \mathbf{V} are estimated (by far the more common case),

$$(\mathbf{K}'\hat{\boldsymbol{\mu}})' \left[\mathbf{K}' (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{K} \right]^{-1} (\mathbf{K}'\hat{\boldsymbol{\mu}}) / \text{rank}(\mathbf{K})$$

has an approximate F -distribution, whose numerator degrees of freedom are $\text{rank}(\mathbf{K})$ and whose denominator degrees of freedom correspond to the degrees of freedom required to estimate $\mathbf{K}'[\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}]^{-1} \mathbf{K}$. For balanced split-plot experiments, $\mathbf{K}'[\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}]^{-1} \mathbf{K}$ is estimated from either a single mean square or a linear combination of mean squares from the analysis of variance.

For example, for the main effect comparison of ET1 versus ET2,

$$\mathbf{K}' = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{K}' (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{K} \propto \text{MS}[\text{WAFER(ET)}]$$

For the comparison of ET1 and ET2 versus ET3 and ET4 in POS2,

$$\mathbf{K}' = \frac{1}{2} [0 \ 0 \ 0 \ 0 \ 1 \ 1 \ -1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

$$\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K} \propto \frac{1}{4} \text{MS}[\text{WAFER(ET)}] + \frac{3}{4} \text{MS(SPE)}$$

The F -ratio for other main effects and interactions, e.g., POS and ET \times POS, can be constructed from appropriately defined \mathbf{K} using the same logic as for ET. The numerator degrees of freedom for the resulting F -test is the rank of \mathbf{K} . The denominator degrees of freedom is determined by the ANOVA mean square implied by $\mathbf{K}'[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{K}$. The general form of the Satterthwaite approximation, from Giesbrecht and Burns (1985), is

$$\nu = \frac{2E\left[\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K}\right]^2}{\text{Var}\left[\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K}\right]}$$

Appendix 1 contains additional details; evaluation of the denominator requires further approximation beyond the scope of this discussion. Table 4.5 shows degree of freedom results for the semiconductor example.

Table 4.5 Degrees of Freedom for Semiconductor Example

Effect Tested	Numerator d.f. (Rank)	Denominator d.f $\mathbf{K}'[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{K}$ depends on
ET	3	MS[WAFER(ET)]
POS	3	MS(SPE)
ET \times POS	9	MS(SPE)

4.4.3 PROC MIXED Analysis of Semiconductor Data

This section shows you how to use PROC MIXED to compute the analysis of the semiconductor experiment. Three basic procedures are covered:

1. Fitting the basic model
2. Estimating the treatment means and differences
3. Testing the hypotheses relevant to these data

Subsequent sections present examples with additional contrast and mean comparison options relevant to those applications, but not to the semiconductor data.

4.4.3.1 Fitting the Mixed Model

Following Table 4.2, and noting that the layout for the semiconductor experiment is a split-plot structure with whole plots conducted in a completely randomized design (Figure 4.1.d), you can fit the model for these data with the following PROC MIXED statements:

```
proc mixed data=ex_4_4;
  class et wafer position;
  model resistance=et position et*position;
  random wafer(et);
run;
```

Output 4.1 shows the results.

Output 4.1 Basic PROC MIXED Model Fitting Output for Semiconductor Data

Covariance Parameter Estimates	
Cov Parm	Estimate
wafer(et)	0.1058
Residual	0.1111

Effect	Type 3 Tests of Fixed Effects			
	Num DF	Den DF	F Value	Pr > F
et	3	8	1.94	0.2015
pos	3	24	3.39	0.0345
et*pos	9	24	0.81	0.6125

Interpretation

The main items of information are as follows:

- The estimates of σ^2 and σ_w^2 , displayed in the “Covariance Parameter Estimates” table as “Residual” and “wafer(et),” respectively. The default variance component estimation procedure used by PROC MIXED is the REML algorithm. For balanced experiments, REML estimates are identical to the estimates obtained from the expected mean squares discussed in Section 4.4.1, provided all estimates are positive.
- The F -tests for the ET and POS main effects and the ET \times POS interaction, given under “Tests of Fixed Effects.” The F -statistics, numerator and denominator degrees of freedom, and p -values (column “Pr > F”) are identical to those you would obtain from computing the analysis of variance (compare to Output 4.2).

You can compute the analysis of variance using PROC MIXED by using the METHOD=TYPE3 option. Modify the first line of the program above to

```
proc mixed method=type3;
```

This produces Output 4.2.

Output 4.2 Analysis of Variance Using PROC MIXED for Semiconductor Data

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
et	3	3.112158	1.037386	Var(Residual) + 4 Var(wafer(et)) + Q(et,et*pos)	MS(wafer(et))	8	1.94	0.2015
pos	3	1.128892	0.376297	Var(Residual) + Q(pos,et*pos)	MS(Residual)	24	3.39	0.0345
et*pos	9	0.809475	0.089942	Var(Residual) + Q(et*pos)	MS(Residual)	24	0.81	0.6125
wafer(et)	8	4.274483	0.534310	Var(Residual) + 4 Var(wafer(et))	MS(Residual)	24	4.81	0.0013
Residual	24	2.667583	0.111149	Var(Residual)

In addition to the F -values for the tests of ET and POSITION main effects and ET \times POSITION interaction, the output also includes the sum of squares and mean squares, the expected mean squares (as previously described in Table 4.4), and the error terms for each test, as described in Table 4.5. This output also gives you an F -statistic for the test of $H_0: \sigma_w^2 = 0$ ($F = 4.81$ for the wafer(et) effect with p -value of 0.0013). Usually in a split-plot experiment you assume that σ_w^2 is nonzero by the nature of the design, and this test is therefore of no interest. However, it is a valid test in case you do want to test σ_w^2 .

4.4.3.2 Treatment Means

To obtain the ET \times POSITION means μ_{ij} and the marginal means μ_i and μ_{ij} , you use the LSMEANS statement

```
lsmeans et position et*position;
```

Place this line after the RANDOM statement in the program above. This gives you the results shown in Output 4.3.

Output 4.3 Semiconductor Data Least-Squares Means Using PROC MIXED Default

Least Squares Means							
Effect	et	position	Estimate	Standard Error	DF	t Value	Pr > t
et	1		5.6258	0.2110	8	26.66	<.0001
et	2		5.9658	0.2110	8	28.27	<.0001
et	3		6.0875	0.2110	8	28.85	<.0001
et	4		6.3325	0.2110	8	30.01	<.0001
position		1	6.0208	0.1345	24	44.78	<.0001
position		2	6.1342	0.1345	24	45.62	<.0001
position		3	5.7475	0.1345	24	42.75	<.0001
position		4	6.1092	0.1345	24	45.44	<.0001

Least Squares Means							
Effect	et	position	Estimate	Standard Error	DF	t Value	Pr > t
et*position	1	1	5.6133	0.2689	24	20.87	<.0001
et*position	1	2	5.4500	0.2689	24	20.27	<.0001
et*position	1	3	5.5533	0.2689	24	20.65	<.0001
et*position	1	4	5.8867	0.2689	24	21.89	<.0001
et*position	2	1	5.9933	0.2689	24	22.29	<.0001
et*position	2	2	6.1867	0.2689	24	23.01	<.0001
et*position	2	3	5.7667	0.2689	24	21.44	<.0001
et*position	2	4	5.9167	0.2689	24	22.00	<.0001
et*position	3	1	6.1367	0.2689	24	22.82	<.0001
et*position	3	2	6.3467	0.2689	24	23.60	<.0001
et*position	3	3	5.7733	0.2689	24	21.47	<.0001
et*position	3	4	6.0933	0.2689	24	22.66	<.0001
et*position	4	1	6.3400	0.2689	24	23.58	<.0001
et*position	4	2	6.5533	0.2689	24	24.37	<.0001
et*position	4	3	5.8967	0.2689	24	21.93	<.0001
et*position	4	4	6.5400	0.2689	24	24.32	<.0001

The standard errors are obtained from the method presented in Section 4.4.1. For example, for the main effect mean of ET1, $\mu_{1\Box}$, the standard error is

$$\sqrt{\frac{1}{12}(\hat{\sigma}^2 + 4\hat{\sigma}_w^2)}$$

Using the variance component estimates given in Output 4.1, the standard error for $\mu_{1\Box}$ is

$$\sqrt{\frac{1}{12}[0.1111 + 4 \times 0.1058]} = 0.2110$$

The t -ratios and their probabilities are obtained by the ratio of the estimate to the standard error, as given in Section 4.4.1. For example, the t -ratio to test $H_0: \mu_{1\Box} = 0$ is

$$t = \frac{\bar{y}_{1\Box}}{\text{std err}(y_{1\Box})} = \frac{5.6258}{0.2110} = 26.66$$

4.4.3.3 Degrees of Freedom

Default Degrees of Freedom in PROC MIXED

PROC MIXED uses containment as the default method for determining denominator degrees of freedom. For the t -values for all least-squares means, the containment method identifies the degrees of freedom for the random effect containing the effect of interest. If no random model

effect contains the LSMEAN effect, then the error degrees of freedom are used. Here, ET is contained within MS[WAFER(ET)], and therefore the degrees of freedom equal 8. POS and ET \times POS are not contained in any random effect, so they have default denominator degrees of freedom of 24.

Overriding the Default

Depending on the form of the standard error, the default may or may not be correct. For example, the standard error for the estimated ET mean is a function of the WAFER(ET) expected mean square *only*, so the default degrees of freedom are correct. However, from Section 4.4.1, the standard errors of the estimated POS and ET \times POS means involve linear combinations of the WAFER(ET) and error expected mean squares.

For example, the variance of an estimated POS mean is $(1/12)(\sigma^2 + \sigma_w^2)$. The estimate of $\sigma^2 + \sigma_w^2$ is $(3/4)\text{MS}[\text{SPE}] + (1/4)\text{MS}[\text{WAFER}(ET)]$. Using Satterthwaite's procedure, the approximate degrees of freedom are

$$v = \frac{\left(\frac{3}{4}\text{MS}(\text{SPE}) + \frac{1}{4}\text{MS}[\text{WAFER}(ET)]\right)^2}{\frac{\left(\frac{3}{4}\text{MS}(\text{SPE})\right)^2}{24} + \frac{\left(\frac{1}{4}\text{MS}[\text{WAFER}(ET)]\right)^2}{8}}$$

$\text{MS}[\text{WAFER}(ET)]$ can be obtained from its expected mean square formula $\sigma^2 + 4\sigma_w^2$. Thus,

$$\text{MS}[\text{WAFER}(ET)] = 0.1111 + 4(0.1058) = 0.5343$$

Also, $\text{MS}(\text{SPE}) = \hat{\sigma}^2 = 0.1111$. The approximate degrees of freedom are

$$v = \frac{\left[\left(\frac{3}{4}\right)0.1111 + \left(\frac{1}{4}\right)0.5343\right]^2}{\frac{\left[\left(\frac{3}{4}\right)(0.1111)\right]^2}{24} + \frac{\left[\left(\frac{1}{4}\right)(0.5343)\right]^2}{8}} = 18.7$$

To get the correct degrees of freedom, you must override the default. PROC MIXED provides you with several ways to override the default and obtain the correct degrees of freedom. By far the most straightforward way is the use of the DDFM option in the MODEL statement. For example, the MODEL statement

```
model y=resistance=et position et*position/ddfm=satterth;
```

causes the Satterthwaite approximation to be computed for every statistic throughout the procedure for which it is appropriate. In more advanced applications, the DDFM=KR option invokes an adjustment to standard errors and test statistics and the degree of freedom approximation due to Kenward and Roger (1997). In a balanced split plot such as this example, the DDFM=KR and DDFM=SATTERTH options yield identical results. However, with split plots that are unbalanced or have missing data and, more importantly, with repeated measures and spatial data, there are important differences and the DDFM=KR option is recommended. The DDFM=KR option is introduced in Section 4.7.3. However, Chapter 5 gives the most detailed explanation and rationale for the Kenward-Roger adjustment.

Alternatively, you can directly specify denominator degrees of freedom. For example, you can use the statement

```
lsmeans position / df=18.7;
```

The results are given in Output 4.4.

Output 4.4 Least-Squares Means for POS Using Specified Degrees of Freedom

Least Squares Means						
Effect	position	Estimate	Standard Error	DF	t Value	Pr > t
position	1	6.0208	0.1345	18.7	44.78	<.0001
position	2	6.1342	0.1345	18.7	45.62	<.0001
position	3	5.7475	0.1345	18.7	42.75	<.0001
position	4	6.1092	0.1345	18.7	45.44	<.0001

Interpretation

The estimates, standard errors, and *t*-ratios are the same as in Output 4.3, but the denominator degrees of freedom are different and *p*-values may be affected. In this case, the reported probabilities are the same because the *t*-ratios are large.

4.4.3.4 Differences among Means

Main Effect Means

From the ANOVA results, only the POSITION main effects are statistically significant. Therefore, the next logical step is specific comparisons among POSITION means. You can do this in two ways. The DIFF option in the LSMEANS statement performs pairwise comparisons. The CONTRAST and ESTIMATE statements allow you to test or estimate various linear combinations of treatment means. For example, you run the following SAS statements to pursue inference on the POSITION means. The ESTIMATE and CONTRAST statements shown here are illustrative only, and are not meant to be an exhaustive set of appropriate comparisons.

```
proc mixed data=ex_4_4;
  class wafer et position;
  model resistance=et|position / ddfm=satterth;
  lsmeans position / diff;
  estimate 'pos1 vs pos3'      position 1 0 -1 0;
  contrast 'pos1 vs pos3'      position 1 0 -1 0;
  estimate 'pos 3 vs others'   position 1 1 -3 1 / divisor=3;
  estimate 'pos3 v oth - wrong' position 1 1 -3 1;
  contrast 'pos 3 vs others'   position 1 1 -3 1;
run;
```

Note the DIVISOR option for the ‘POS 3 VS OTHERS’ estimate. If you omit it, you estimate $\mu_1 + \mu_2 + \mu_4 - 3\mu_3$, the difference between the sum of position 1, 2, and 4 and three times the mean of position 3. The correct statement, with the DIVISOR=3 option, estimates $(1/3)(\mu_1 + \mu_2 + \mu_4) - \mu_3$, the difference between the average mean of positions 1, 2, and 4 and the mean of

position 3. For the CONTRAST statement, the DIVISOR option is unnecessary because the change of scale occurs in both the numerator and the denominator. Output 4.5 shows the results.

Output 4.5 Comparison of POSITION Least-Squares Means Using LSMEANS (/ DIFF), ESTIMATE, and CONTRAST Statements

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
pos1 vs pos3	0.2733	0.1361	24	2.01	0.0560
pos 3 vs others	0.3406	0.1111	24	3.06	0.0053
pos3 vs oth - wrong	1.0217	0.3334	24	3.06	0.0053

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
pos1 vs pos3	1	24	4.03	0.0560
pos 3 vs others	1	24	9.39	0.0053

Differences of Least Squares Means							
Effect	position	_position	Estimate	Standard Error	DF	t Value	Pr > t
position	1	2	-0.1133	0.1361	24	-0.83	0.4132
position	1	3	0.2733	0.1361	24	2.01	0.0560
position	1	4	-0.08833	0.1361	24	-0.65	0.5225
position	2	3	0.3867	0.1361	24	2.84	0.0090
position	2	4	0.02500	0.1361	24	0.18	0.8558
position	3	4	-0.3617	0.1361	24	-2.66	0.0138

Interpretation

First, notice that the ‘POS 1 vs POS 3’ estimate and the difference between the position 1 and position 3 least-squares means from the DIFF option produce identical results: they represent different ways of programming the same mean difference. In general, it is easier to obtain a difference between a pair of means with the DIFF option. Use the ESTIMATE statement for linear combinations, such as ‘POS 3 vs OTHERS’, that cannot be obtained using the DIFF option. You can see the impact of the DIVISOR statement in the two ESTIMATE statements for position 3 versus the others: with the DIVISOR option you get a sensible estimate of the mean difference; without it you get a value multiplied by the number of means in the “others” group (in this case 3) and the result makes no sense. Note, however, that the *t*-statistic is unaffected: changes of scale affect the estimates but not the tests (the standard error of the estimate changes by the same factor as the estimate itself). You can also see that for testing purposes, the CONTRAST or ESTIMATE statement output can be used interchangeably for 1 degree-of-freedom comparison. The CONTRAST statement gives you an *F*-statistic; the ESTIMATE statement yields a *t*-statistic. The *F* is equal to t^2 and the *p*-values are identical. The main advantage of the ESTIMATE statement is that you also get the difference estimate and its

standard error. This allows you to report *how different* the means are, not merely that they are, or are not, significantly different.

Defining a Specific Treatment as a Control

You can define a specific position as the “control” treatment. This restricts the treatment differences printed to those involving the “control.” You can also adjust the *p*-value, e.g., using Dunnett’s test, to control the type I error. For example, suppose POSITION 3 is the control or reference treatment. You can modify the LSMEANS statement as follows to obtain Output 4.6:

```
lsmeans position/diff=control('3') adjust=dunnett;
```

The level (“3”) in parentheses identifies the control level of position. Note that it must be enclosed in single quotes. Consult the *SAS/STAT User’s Guide* for a complete list of ADJUST= options.

Output 4.6 Dunnett Test of Position 3 versus Other Positions in Semiconductor Data

Differences of Least Squares Means									
Effect	position	_position	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
position	1	3	0.2733	0.1361	24	2.01	0.0560	Dunnett-Hsu	0.1363
position	2	3	0.3867	0.1361	24	2.84	0.0090	Dunnett-Hsu	0.0240
position	4	3	0.3617	0.1361	24	2.66	0.0138	Dunnett-Hsu	0.0361

The _POSITION variable is the control. The column labeled “Pr > |t|” gives the unadjusted *p*-values, e.g., for the least significant difference test. The “Adj P” column gives the Dunnett-adjusted *p*-values. In many cases, your conclusions would not be affected by the adjustment, but you can see that they could. For example, if you use a significance level of $\alpha = 0.10$, your conclusion about the difference between position 1 and position 3 would change.

Simple Effects and ET \times POSITION Means

Although the ET \times POSITION interaction is not significant, you may still want to look at various differences among specific ET \times POSITION means. You can do this in several ways: you can use the DIFF and SLICE options of the LSMEAN statement, or you can use ESTIMATE or CONTRAST statements for specific simple effects of interest. The following statements provide examples of comparisons that might be of interest. The methods described in Section 4.3.2.1 were used to obtain the coefficients. Note that the ESTIMATE statements for the simple effects tend to be lengthy. See Section 4.4.4 for a somewhat less tedious alternative using PROC GLIMMIX. The interpretation of the various estimates and contrasts is given following the output below.

```
lsmeans et*position / slice=(et position);
estimate 'pos 1 vs 3 in et4'
    position 1 0 -1 0
    et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1 0;
estimate 'pos 1 vs 3 in et1'
    position 1 0 -1 0
    et*position 1 0 -1 0;
estimate 'pos3 vs others in et=4'
    position 1 1 -3 1
    et*position 0 0 0 0 0 0 0 0 0 0 0 0 1 1 -3 1 / divisor=3;
estimate 'pos3 vs others in et<4'
    position 3 3 -9 3
    et*position 1 1 -3 1 1 1 -3 1 1 1 -3 1 0 / divisor=9;
```

```

estimate 'et1 vs et2 in pos 2'
    et 1 -1 0 0
    et*position 0 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0;
estimate 'et1 vs others in pos 1'
    et -3 1 1 1
    et*position -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 / divisor=3;
estimate 'et1 vs others in pos 2'
    et -3 1 1 1
    et*position 0 -3 0 0 0 1 0 0 0 1 0 0 0 1 0 / divisor=3;
contrast 'pos 1 vs 3 in et4'
    position 1 0 -1 0
    et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1 0;
contrast 'pos 1 vs 3 in et1'
    position 1 0 -1 0
    et*position 1 0 -1 0;
contrast 'pos3 vs others in et=4'
    position 1 1 -3 1
    et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 -3 1;
contrast 'pos3 vs others in et<4'
    position 3 3 -9 3
    et*position 1 1 -3 1 1 1 -3 1 1 1 -3 1 0;
contrast 'et1 vs et2 in pos 2'
    et 1 -1 0 0
    et*position 0 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0;
contrast 'et1 vs others in pos 1'
    et -3 1 1 1
    et*position -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0;
contrast 'et1 vs others in pos 2'
    et -3 1 1 1
    et*position 0 -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0;

```

The SLICE option obtains simple effect tests among POSITION means given each ET level and among ET means given each level of POSITION. Note that you can combine both sets of tests in a single statement. The ESTIMATE and CONTRAST statements are not meant to be exhaustive: they show various comparisons that are possible.

For the ESTIMATE and CONTRAST statements, there are several conventions to note. First, the order of the coefficients for the ET \times POSITION effects follows from the CLASS statement. Because ET is listed first, the first four coefficients refer to position 1 through position 4 with ET 1, the next four to position 1 through position 4 with ET 2, etc. Second, when the last several coefficients are zero, you do not need to list all of them: all coefficients after the last zero mentioned are assumed to be zero. For example, in the ESTIMATE ‘pos 1 vs pos 3 in et 1’ you do not need to list the last 12 zeros. They are assumed to be zero.

Output 4.7 Simple Effect Inference with Semiconductor Data: SLICE Option and Selected ESTIMATE and CONTRAST Statement Results

Tests of Effect Slices						
Effect	et	position	Num DF	Den DF	F Value	Pr > F
et*position	1		3	24	0.94	0.4373
et*position	2		3	24	0.82	0.4934
et*position	3		3	24	1.51	0.2362
et*position	4		3	24	2.54	0.0807
et*position		1	3	18.7	1.30	0.3038

Tests of Effect Slices						
Effect	et	position	Num DF	Den DF	F Value	Pr > F
et*position		2	3	18.7	3.19	0.0477
et*position		3	3	18.7	0.28	0.8383
et*position		4	3	18.7	1.26	0.3181

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
pos 1 vs 3 in et4	0.4433	0.2722	24	1.63	0.1164
pos 1 vs 3 in et1	0.06000	0.2722	24	0.22	0.8274
pos3 vs others in et=4	0.5811	0.2223	24	2.61	0.0152
pos3 vs others in et<4	0.2604	0.1283	24	2.03	0.0537
et1 vs et2 in pos 2	-0.7367	0.3803	18.7	-1.94	0.0680
et1 vs others in pos 1	0.5433	0.3105	18.7	1.75	0.0966
et1 vs others in pos 2	0.9122	0.3105	18.7	2.94	0.0086

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
pos 1 vs 3 in et4	1	24	2.65	0.1164
pos 1 vs 3 in et1	1	24	0.05	0.8274
pos3 vs others in et=4	1	24	6.84	0.0152
pos3 vs others in et<4	1	24	4.12	0.0537
et1 vs et2 in pos 2	1	18.7	3.75	0.0680
et1 vs others in pos 1	1	18.7	3.06	0.0966
et1 vs others in pos 2	1	18.7	8.63	0.0086

Interpretation

The SLICE output gives you two sets of tests. The first four (where the numbers 1 through 4 appear in the ET column) test $H_0: \mu_{i1} = \mu_{i2} = \mu_{i3} = \mu_{i4}$, that is, the equality of the POSITION simple effects given ET level $i = 1, 2, 3$, and 4, respectively. The second set of four (where numbers 1 through 4 appear in the “position” column) test the equality of the ET simple effect means given position $j = 1, 2, 3$, and 4, respectively, that is $H_0: \mu_{1j} = \mu_{2j} = \mu_{3j} = \mu_{4j}$. From the results, there is some evidence ($p = 0.0807$) of differences among POSITION simple effect means given ET 4, but not for the other ET levels. There is stronger evidence ($p = 0.0477$) of a difference among ET means given position 2, but no evidence of ET simple effects given the other positions. Note that slices comparing ET simple effects at specific positions use Satterthwaite’s approximation for their degrees of freedom. This is because the denominator term for the F -statistics is a linear combination of MS[Wafer(ET)] and MS[split-plot error]. The F -statistic for the POSITION simple effects within given ET levels uses MS[split-plot error] only, since these comparisons occur within the same whole plot, and hence they use the split-plot error degrees of freedom, in this case 24.

The ESTIMATE and CONTRAST statements pursue the SLICE results in more depth. You can see from the ET \times POSITION means in Output 4.3 that POSITION 3 has a noticeably different mean from the others in ET 4 but not in the other ET levels. ‘POS 3 VS POS 1 IN ET 4’ and ‘POS 3 VS POS 1 IN ET 1’ are examples of estimates or tests of differences among specific POSITION means given a specific ET level. You could obtain such differences by adding a DIFF option to the LSMEANS statement, but that would give you all possible differences (in this case there are 120 pairs of ET \times POSITION means to compare) and you may want to target your comparisons more selectively. ‘POS 3 VS OTHERS IN ET 4’ compares POSITION 3 to the average of the other positions given ET 4 only. Recall that this contrast was significant in the tests of POSITION main effects. ‘POS 3 VS OTHERS IN ET<4’ compares $(\mu_{13} + \mu_{23} + \mu_{33})/3$, the average of the ET \times POSITION means averaged over ET levels 1, 2, and 3 at position 3, with the ET \times POSITION mean average $(1/9)(\mu_{11} + \mu_{12} + \mu_{14} + \mu_{21} + \mu_{22} + \mu_{24} + \mu_{31} + \mu_{32} + \mu_{34})$, the average of positions 1, 2, and 4 at ET levels 1, 2, and 3. The results show stronger differences for position 3 given ET4 compared to position 3 given the other ET levels. For example, given ET 4, the estimated position 1 vs. position 3 difference is 0.4433 with standard error 0.2722 and $p = 0.1164$, whereas given ET 1, the estimated position 1 vs. position 3 difference is 0.06 with standard error 0.2722 and $p = 0.8274$.

Analogous ESTIMATE and CONTRAST statements appear for ET simple effects at given positions. Note that the degrees of freedom for ET simple effects at specific positions require Satterthwaite’s approximation, since their standard errors involve estimates of whole-plot and split-plot error. From comparing the ‘ET 1 VS OTHERS IN POS 1’ results (estimated difference 0.5433, $p = 0.0966$) to ‘ET 1 VS OTHERS IN POS 2’ (estimated difference 0.9122, $p = 0.0086$), you can gain some insight into the SLICE results showing ET effects at POSITION 2 but not at the other positions.

Using a Control for Simple Effects

You can use the DIFF option to obtain all 120 pairwise comparisons among ET \times POSITION means. In some cases, it may make sense to define a specific treatment combination as a control or reference treatment. You may even want to define a specific level of one factor as the reference and compare the other levels of that factor within each level of the other factor, essentially a Dunnett-style elaboration of the SLICE tests. For example, the following PROC MIXED statements define ET level one in conjunction with POSITION 1, 2, 3, and 4, respectively, in the four statements shown:

```
lsmeans et*position / diff=control('1' '1');
lsmeans et*position / diff=control('1' '2');
lsmeans et*position / diff=control('1' '3');
lsmeans et*position / diff=control('1' '4');
```

Note the syntax that defines the control: it identifies the level of ET and POSITION. They must be in the same order as they are listed in the CLASS statement and their levels must be in single quotes. You can run all four statements in the same PROC MIXED program. The output will have four sets of 15 comparisons: μ_{11} versus all 15 other ET \times POSITION treatment means, μ_{12} versus all 15 other ET \times POSITION treatment means, etc. Because this output will be lengthy and contain many comparisons of no interest, you can add the following statements at the end of the MIXED procedure:

```
ods exclude lsmeans diffs;
ods output diffs=simpeff;
run;
```

```

data smpeff_dunnett;
  set simpeff;
  if position=_position;
proc print data=smpeff_dunnett;
run;

```

The first ODS statement suppresses printing of the LSMEANS and the sets of comparisons called for by DIFF. Otherwise your SAS listing will contain four sets of LSMEANS and the four sets of differences described above. The second ODS statement outputs the differences to a new SAS data set, in this case called *simpeff*. The DATA step defines a new SAS data set *simpeff_dunnett*, which contains only those differences with a common position. This is accomplished by the “if” statement. The results, obtained from PROC PRINT, appear in Output 4.8.

Output 4.8 ET 1 versus Other ET Simple Effects by Position

Obs	Effect	et	position	_et	_position	Estimate	StdErr	DF	tValue	Probt
1	et*position	2	1	1	1	0.3800	0.3803	18.7	1.00	0.3305
2	et*position	3	1	1	1	0.5233	0.3803	18.7	1.38	0.1851
3	et*position	4	1	1	1	0.7267	0.3803	18.7	1.91	0.0715
4	et*position	2	2	1	2	0.7367	0.3803	18.7	1.94	0.0680
5	et*position	3	2	1	2	0.8967	0.3803	18.7	2.36	0.0295
6	et*position	4	2	1	2	1.1033	0.3803	18.7	2.90	0.0093
7	et*position	2	3	1	3	0.2133	0.3803	18.7	0.56	0.5815
8	et*position	3	3	1	3	0.2200	0.3803	18.7	0.58	0.5698
9	et*position	4	3	1	3	0.3433	0.3803	18.7	0.90	0.3781
10	et*position	2	4	1	4	0.03000	0.3803	18.7	0.08	0.9380
11	et*position	3	4	1	4	0.2067	0.3803	18.7	0.54	0.5933
12	et*position	4	4	1	4	0.6533	0.3803	18.7	1.72	0.1023

Interpretation

The twelve comparisons appear in sets of three, the first set for POSITION 1, the next set for POSITION 2, etc. You can see that for positions 3 and 4, there is no evidence of any statistically significant differences between ET 1 and the other ET levels. At POSITION 1 there is weak evidence ($p = 0.0715$). Only at POSITION 2 are there substantial differences between ET 1 and the other ET levels. This is consistent with the SLICE results discussed earlier. You could add the ADJUST=DUNNETT option to the LSMEANS statements shown above to obtain Dunnett-Hsu-adjusted p -values, similar to what was shown for the POSITION main effects. For example, for position 1, the statement is

```
lsmeans et*position / diff=control('1' '1') adjust=dunnett;
```

4.4.4 Alternative Mean Comparisons Using PROC GLIMMIX

The ESTIMATE statements for simple effects shown in the previous section require a great deal of care and can be quite tedious. PROC GLIMMIX is a new (as of 2005) procedure for analyzing generalized linear mixed models (GLMM). GLMMs were defined in Chapter 1 and are discussed in greater detail in Chapter 14. The linear mixed model (LMM) is a special case. While PROC MIXED is the primary tool for analyzing linear mixed models, PROC GLIMMIX

has features that make it easier for you to program certain analyses of interest. One example involves simple effects in split-plot experiments.

The following PROC GLIMMIX statements reproduce the simple effect analysis shown in Section 4.4.3.4, Output 4.7.

```
proc glimmix data=ex_4_4;
  class et wafer position;
  model resistance=et position et*position/ddfm=kr;
  random wafer(et);
  lsmeans et*position / slice=(et position) diff=control ('1' '2');
  lsmeans et*position
    'pos 1 vs 3 in et1' 1 0 -1 0,
    'pos 1 vs 3 in et4' 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1 0;
  lsmeans et*position
    'pos3 vs others in et=4' 0 0 0 0 0 0 0 0 0 0 0 0 1 1 -3 1,
    'pos3 vs others in et<4' 1 1 -3 1 1 1 -3 1 1 1 -3 1 0 /
      divisor=3,9;
  lsmeans et*position
    'et1 vs et2 in pos 2' 0 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 0;
  lsmeans et*position
    'et1 vs others in pos 1' -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0,
    'et1 vs others in pos 2' 0 -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 /
      divisor=3;
run;
```

The syntax for the CLASS, MODEL, RANDOM, and LSMEANS statements is identical to the PROC MIXED syntax. The LSMESTIMATE statement performs a similar role to that of the ESTIMATE statement in PROC MIXED, but there are three important differences. First, the coefficients define linear contrasts on the *least-squares means*, not on the model effects as in PROC MIXED. For example, the first part of the second LSMESTIMATE statement, ‘pos3 vs others in et=4’, defines the comparison $(1/3)(\mu_{41} + \mu_{42} - 3\mu_{43} + \mu_{44})$. The coefficients involve ET \times POSITION only. In MIXED, you would have to re-express the coefficients in terms of the effects model, i.e., $\frac{1}{3}[\beta_1 + \beta_2 - 3\beta_3 + \beta_4] + \frac{1}{3}[(\alpha\beta)_{41} + (\alpha\beta)_{42} - 3(\alpha\beta)_{43} + (\alpha\beta)_{44}]$, and write the ESTIMATE statement including coefficients for POSITION as well as ET \times POSITION.

The second difference is that you can put more than one comparison in the LSMESTIMATE statement. In theory you could put all seven comparisons defined above in one statement. *P*-values for multiple-row LSMESTIMATE statements in PROC GLIMMIX can be adjusted for multiplicity. Note that the ESTIMATE statement of PROC GLIMMIX also has this feature; you can specify multiple-row estimates and adjust *p*-values for multiplicity. This program shows them divided into subsets with common comparisons to avoid confusion. Finally, the order of the effect and the label is reversed: in the ESTIMATE statement, you put the label after ESTIMATE, then the effect, then the coefficients. Here, the effect goes first, then the label, then the coefficient. The LSMESTIMATE statement results are shown in Output 4.9.

Output 4.9 LSMESTIMATE Results for Selected Simple Effects in Semiconductor Experiment (PROC GLIMMIX)

Least Squares Means Estimates						
Effect	Label	Estimate	Standard Error	DF	t Value	Pr > t
et*position	pos 1 vs 3 in et1	0.060	0.272	24	0.22	0.8274
et*position	pos 1 vs 3 in et4	0.443	0.272	24	1.63	0.1164
et*position	pos3 vs others in et=4	0.581	0.222	24	2.61	0.0152
et*position	pos3 vs others in et<4	0.260	0.128	24	2.03	0.0537
et*position	et1 vs et2 in pos 2	-0.737	0.380	18.68	-1.94	0.0680
et*position	et1 vs others in pos 1	1.630	0.932	18.68	1.75	0.0966
et*position	et1 vs others in pos 2	0.912	0.311	18.68	2.94	0.0086

You can see that the results match those of the ESTIMATE statement in PROC MIXED (Output 4.7) aside from some minor differences in rounding and format.

In the interest of space, the other PROC GLIMMIX output is not shown. Like the output above, it is similar to PROC MIXED output except for minor differences due to format and rounding. PROC GLIMMIX has several other features for presenting means, including plots of the results of multiple comparison procedures similar to the LINES option of the MEANS statement in PROC GLM.

4.5 Comparison with PROC GLM

Although PROC MIXED has been available since the early 1990s and designs with split-plot features are clearly mixed models, many data analysts were trained to compute analysis of split-plot designs using PROC GLM. Many contemporary textbooks persist in presenting SAS analysis of split-plot experiments using PROC GLM. It is therefore instructive to look at what you can and cannot do with split plots using PROC GLM. Several essential aspects of analysis described in the last section are either done incorrectly or cannot be done at all using PROC GLM.

The following SAS statements duplicate as much of the analyses from Section 4.4 as PROC GLM can do.

```
proc glm data=ex_4_4;
  class et wafer position;
  model resistance=et wafer(et) position et*position;
  random wafer(et)/test;
  lsmeans et/ e=wafer(et) stderr tdiff;
  lsmeans position/ stderr tdiff;
  lsmeans et*position / slice=(et position);
  estimate 'pos 1 vs 3 in et4'
    position 1 0 -1 0
    et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1 0;
  estimate 'pos 1 vs 3 in et1'
    position 1 0 -1 0
    et*position 1 0 -1 0;
  estimate 'pos3 vs others in et=4'
```

```

position 1 1 -3 1
et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 -3 1 /
divisor=3;
estimate 'pos3 vs others in et<4'
position 3 3 -9 3
et*position 1 1 -3 1 1 1 -3 1 1 1 -3 1 0/ divisor=9;
estimate 'et1 vs et2 in pos 2'
et 1 -1 0 0
et*position 0 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0;
estimate 'et1 vs others in pos 1'
et -3 1 1 1
et*position -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 /
divisor=3;
estimate 'et1 vs others in pos 2'
et -3 1 1 1
et*position 0 -3 0 0 0 1 0 0 0 1 0 0 0 1 0  / 
divisor=3;
contrast 'pos 1 vs 3 in et4'
position 1 0 -1 0
et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1 0;
contrast 'pos 1 vs 3 in et1'
position 1 0 -1 0
et*position 1 0 -1 0;
contrast 'pos3 vs others in et=4'
position 1 1 -3 1
et*position 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 -3 1;
contrast 'pos3 vs others in et<4'
position 3 3 -9 3
et*position 1 1 -3 1 1 1 -3 1 1 1 -3 1 0;
contrast 'et1 vs et2 in pos 2'
et 1 -1 0 0
et*position 0 1 0 0 0 -1 0 0 0 0 0 0 0 0 0 0 0;
contrast 'et1 vs others in pos 1'
et -3 1 1 1
et*position -3 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0;
contrast 'et1 vs others in pos 2'
et -3 1 1 1
et*position 0 -3 0 0 0 1 0 0 0 1 0 0 0 1 0  / 
lsmeans et*position / tdiff=control('1' '1');
lsmeans et*position / tdiff=control('1' '2');
lsmeans et*position / tdiff=control('1' '3');
lsmeans et*position / tdiff=control('1' '4');
run;

```

There are some differences in these statements compared to PROC MIXED. In the MODEL statement, all effects, fixed and random, must be listed on the right-hand side of the equal sign. PROC GLM does not distinguish between fixed and random effects computationally the way PROC MIXED does. The RANDOM statement causes the expected mean squares to be listed, but the TEST statement is required to obtain correct ANOVA F -statistics for the whole plot. Similarly, in the LSMEANS statement for ET, the whole-plot error term E=WAFER(ET) must be named explicitly or PROC GLM will use the residual error term; unlike MIXED, the RANDOM statement in GLM does not set up the appropriate variance structure upon which all standard errors and test statistics are based. Consequently, you must name the correct error term where possible. For cases that involve a linear combination of whole-plot and split-plot error, PROC GLM cannot compute the correct statistics because it has no programming option to define the needed error term. This is the primary reason that we **emphatically recommend against using PROC GLM to analyze split-plot experiments.**

We consider the PROC GLM output in parts. Output 4.10 shows the basic analysis of variance results from the MODEL and RANDOM statements.

Output 4.10 PROC GLM ANOVA: Selected Results for Semiconductor Experiment

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	9.32500833	0.40543514	3.65	0.0013
Error	24	2.66758333	0.11114931		
Corrected Total	47	11.99259167			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
et	3	3.11215833	1.03738611	9.33	0.0003
wafer(et)	8	4.27448333	0.53431042	4.81	0.0013
position	3	1.12889167	0.37629722	3.39	0.0345
et*position	9	0.80947500	0.08994167	0.81	0.6125

Source	Type III Expected Mean Square
et	Var(Error) + 4 Var(wafer(et)) + Q(et,et*position)
wafer(et)	Var(Error) + 4 Var(wafer(et))
position	Var(Error) + Q(position,et*position)
et*position	Var(Error) + Q(et*position)

*Tests of Hypotheses for Mixed Model Analysis of Variance**Dependent Variable: resistance*

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	et	3	3.112158	1.037386	1.94	0.2015
	Error: MS(wafer(et))	8	4.274483	0.534310		

* This test assumes one or more other fixed effects are zero.

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	wafer(et)	8	4.274483	0.534310	4.81	0.0013
*	position	3	1.128892	0.376297	3.39	0.0345
	et*position	9	0.809475	0.089942	0.81	0.6125
	Error: MS(Error)	24	2.667583	0.111149		

* This test assumes one or more other fixed effects are zero.

Comments

The “MODEL” source of variation in the first ANOVA table is the sum of all sums of squares other than split-plot error. As such, it contains no information of interest. This is also true of the Type III ANOVA, whose *F*-statistics are all computed using the split-plot error mean square regardless of whether this is appropriate or not. The “Type III Expected Mean Squares” are identical to those obtained by the METHOD=TYPE3 option in PROC MIXED. The tables that

follow them contain the F -statistics consistent with the expected mean squares for the sources of variation named in the MODEL statement. These results are identical to the Type III results in PROC MIXED and are correct.

Least-Squares Means

Output 4.11 shows the results for the LSMEANS statements.

Output 4.11 Selected LSMEANS Results for the Semiconductor Experiment

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III MS for wafer(et) as an Error Term

et	resistance LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	5.62583333	0.21101154	<.0001	1
2	5.96583333	0.21101154	<.0001	2
3	6.08750000	0.21101154	<.0001	3
4	6.33250000	0.21101154	<.0001	4

Least Squares Means for Effect et t for H0: LSMean(i)=LSMean(j) / Pr > t				
Dependent Variable: resistance				
i/j	1	2	3	4
1		-1.13935 0.2875	-1.54706 0.1604	-2.36806 0.0454
2	1.139351 0.2875		-0.40771 0.6942	-1.22871 0.2541
3	1.547061 0.1604	0.407709 0.6942		-0.821 0.4354
4	2.368064 0.0454	1.228712 0.2541	0.821003 0.4354	

position	resistance LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	6.02083333	0.09624158	<.0001	1
2	6.13416667	0.09624158	<.0001	2
3	5.74750000	0.09624158	<.0001	3
4	6.10916667	0.09624158	<.0001	4

Least Squares Means for Effect position t for H0: LSMean(i)=LSMean(j) / Pr > t					
Dependent Variable: resistance					
i/j	1	2	3	4	
1		-0.83268 0.4132	2.008236 0.0560	-0.649 0.5225	
2	0.832683 0.4132		2.84092 0.0090	0.18368 0.8558	
3	-2.00824 0.0560	-2.84092 0.0090		-2.65724 0.0138	
4	0.649003 0.5225	-0.18368 0.8558	2.65724 0.0138		

et*position Effect Sliced by et for resistance					
et	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	3	0.313092	0.104364	0.94	0.4373
2	3	0.274825	0.091608	0.82	0.4934
3	3	0.504958	0.168319	1.51	0.2362
4	3	0.845492	0.281831	2.54	0.0807

et*position Effect Sliced by position for resistance					
position	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	3	0.846292	0.282097	2.54	0.0805
2	3	2.075092	0.691697	6.22	0.0028
3	3	0.182958	0.060986	0.55	0.6539
4	3	0.817292	0.272431	2.45	0.0880

Comments

The output for the ET least-squares means is correct. Notice, however, that you cannot obtain the standard error for the difference, which is the one statistic that *should* be published in conjunction with ET mean comparisons. PROC GLM can compute a standard error of a difference using the ESTIMATE statement, but in this case it will use the MS for split-plot error, which is wrong. PROC GLM's ESTIMATE statement has no option to override the default, so you cannot obtain the correct standard error of a difference. The POSITION main effect mean estimates are correct, but the standard errors are wrong (they should be 0.1345, not 0.0962). The standard error is a linear combination of whole-plot and split-plot error; there are no statements in PROC GLM to do this. As with ET, you also cannot obtain the standard error of a difference. The *t*-statistics and their *p*-values comparing the POSITION means are correct because in this case, the GLM default using the split-plot error is the correct error term for testing the split-plot main effect.

The SLICE results comparing POSITION simple effects for given ET levels are correct, since they use split-plot error as a test term, but the SLICE results comparing ET simple effects for a given position are incorrect. For example, the test given position 1 should have an *F*-value of

1.30, but PROC GLM computes 2.54. This is because the correct error term involves both whole-plot and split-plot error, which GLM cannot accommodate.

The Dunnett results are not shown here, but they are similarly affected by the split-plot and whole-plot error issues just discussed. In general, you can compute correct Dunnett tests for main effects, provided you specify the whole-plot error term for testing whole-plot main effects, but you **cannot** use PROC GLM for testing simple effects as was shown with PROC MIXED.

Estimates and Contrasts

Output 4.12 shows results for the ESTIMATE statements. Because the CONTRAST results are similar in terms of PROC GLM versus PROC MIXED issues, they are not shown here.

Output 4.12 ESTIMATE Results Using PROC GLM for Semiconductor Experiment

Parameter	Estimate	Standard Error	t Value	Pr > t
pos 1 vs 3 in et4	0.44333333	0.27221230	1.63	0.1164
pos 1 vs 3 in et1	0.06000000	0.27221230	0.22	0.8274
pos3 vs others in et=4	0.58111111	0.22226041	2.61	0.0152
pos3 vs others in et<4	0.26037037	0.12832211	2.03	0.0537
et1 vs et2 in pos 2	-0.73666667	0.27221230	-2.71	0.0123
et1 vs others in pos 1	0.54333333	0.22226041	2.44	0.0222
et1 vs others in pos 2	0.91222222	0.22226041	4.10	0.0004

Comments

The estimated differences shown in Output 4.12 are all correct, but in many cases the standard errors and hence the test statistics are not. Specifically, the simple effects among POSITION means for given ET are correct, but simple effect tests between ET levels for a given POSITION are incorrect. For example, the correct standard error for the ‘et1 vs et2 in pos 2’ difference is 0.3803, not 0.2722 as shown by PROC GLM. Similarly, the two ‘et1 vs others’ given POSITION 1 and 2 should have a standard error of 0.3105, not 0.2223. These standard errors require split-plot and whole-plot components, which GLM cannot do. For whole-plot main effect contrasts, you can obtain the correct test statistics using the CONTRAST statement, which allows for an optional error term, but you cannot obtain the correct standard error of a difference with the ESTIMATE statement in PROC GLM.

4.6 Example: Type x Dose Response

This example involves data from an experiment to compare the response of two plant varieties to increasing amounts of pesticide designed to protect the plants against disease. The experiment was conducted in a greenhouse. Plants were placed on five benches, which were used as blocks to account for local variation in the greenhouse. Each bench was divided into four sections where, for convenience, levels of the pesticide were applied. Each section had two plants, one of each variety. Data Set 4.6, “Variety-Pesticide Evaluation,” in Appendix 2, “Data Sets,” contains the data for this experiment. The variable TYPE denotes the variety, S for “susceptible” (a genotype susceptible to the disease) and R for “resistant”; the variable DOSE refers to the amount of pesticide, and the variable Y is the plant response.

The design is a split plot with the whole plot conducted in randomized complete blocks, the layout shown in Figure 4.1.e. DOSE is the whole-plot factor; TYPE is the split-plot factor. Thus, a model for this experiment is

$$Y_{ijk} = \mu_{ij} + r_k + w_{ik} + e_{ijk}$$

where

Y_{ijk} is the observation on the i^{th} dose, j^{th} variety, and k^{th} block

μ_{ij} is the ij^{th} dose \times variety mean

r_k is the k^{th} block effect, assumed iid $N(0, \sigma_R^2)$

w_{ik} is the ik^{th} whole-plot (block \times dose) effect, assumed iid $N(0, \sigma_W^2)$

e_{ijk} is the ijk^{th} split-plot error effect, assumed iid $N(0, \sigma^2)$

The example analyses that follow show several ways to model the dose \times variety means, μ_{ij} . The next section starts with the conventional partition into main effects and interactions,

$$\mu_{ij} = \mu + \delta_i + \tau_j + (\delta\tau)_{ij}$$

where

δ_i is the i^{th} DOSE main effect

τ_j is the j^{th} variety (TYPE) main effect

$(\delta\tau)_{ij}$ is the ij^{th} DOSE \times TYPE interaction effect

4.6.1 PROC MIXED Analysis of DOSE \times TYPE Main Effects and Interactions

The following SAS statements, consistent with Table 4.2, obtain the main effect and interaction analysis of the variety evaluation data. Outputs 4.13 and 4.14 show the results.

```
proc mixed data=variety_eval;
  class block type dose;
  model y = type|dose / ddfm=satterth;
  random block block*dose;
  lsmeans type*dose / slice=(type dose);
run;
```

Output 4.13 Variance Estimates, Main Effect, and Interaction Tests for Variety Evaluation

Covariance Parameter Estimates	
Cov Parm	Estimate
block	2.0735
block*dose	4.5132
Residual	4.3189

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
type	1	16	2.78	0.1151
dose	3	12	13.63	0.0004
type*dose	3	16	2.29	0.1176

Interpretation

The “Covariance Parameter Estimates” table gives the estimates of the block, whole-plot error, and split-plot error variances components, σ_R^2 , σ_W^2 , and σ^2 , respectively. In this case, the variance component estimates are all positive. Section 4.7 contains a brief discussion of issues you should consider if either the block (σ_R^2) or whole-plot error (σ_W^2) variance component estimate is zero. The “Type 3 Tests of Fixed Effects” table gives the tests for TYPE \times DOSE interactions and TYPE and DOSE main effects. As with any factorial experiment, you should consider the tests in that order: interaction first. In this case, the test for interaction has a *p*-value of 0.1176. In the next section we consider this test in more detail: because there are multiple degrees of freedom (i.e., 3 numerator d.f.) associated with this test, important interaction effects with 1 or 2 d.f. may be masked. For the moment, assuming no interaction effect, the tests of DOSE (H_0 : all $\delta_i = 0$ assuming no interaction) and TYPE (H_0 : all $\tau_j = 0$ assuming no interaction) suggest that the DOSE effect is statistically significant but the TYPE effect is not.

Ordinarily, if DOSE \times TYPE effects are assumed to be zero, the SLICE results would not be of interest. However, for the sake of argument, Output 4.14 shows the SLICE results.

Output 4.14 Test of SLICEs for Variety Evaluation Data

Tests of Effect Slices						
Effect	type	dose	Num DF	Den DF	F Value	Pr > F
type*dose	r		3	19.5	8.12	0.0010
type*dose	s		3	19.5	13.58	<.0001
type*dose		1	1	16	5.00	0.0399
type*dose		2	1	16	4.03	0.0618
type*dose		4	1	16	0.02	0.8810
type*dose		8	1	16	0.58	0.4578

Note that the test of DOSE simple effects for given variety TYPE occurs across whole-plot experimental units and thus involves both whole-plot and split-plot error. Hence the denominator degrees of freedom (19.5) reflect the use of Satterthwaite’s approximation. For the TYPE simple effects for given DOSE levels, you can see that there appear to be large differences among varieties at low DOSE levels (i.e., low levels of pesticide protection from disease) but no statistically significant differences at higher DOSE levels. This is exactly what one would expect when comparing a resistant variety to a susceptible variety, and it suggests that we should rethink the test for interaction.

The first step is to examine an interaction plot. You can obtain an interaction plot by adding the following SAS statement to the PROC MIXED program given above:

```

ods output lsmeans=typedosemeans;
run;

data to_plot;
  set typedosemeans;
  logdose=log2(dose);

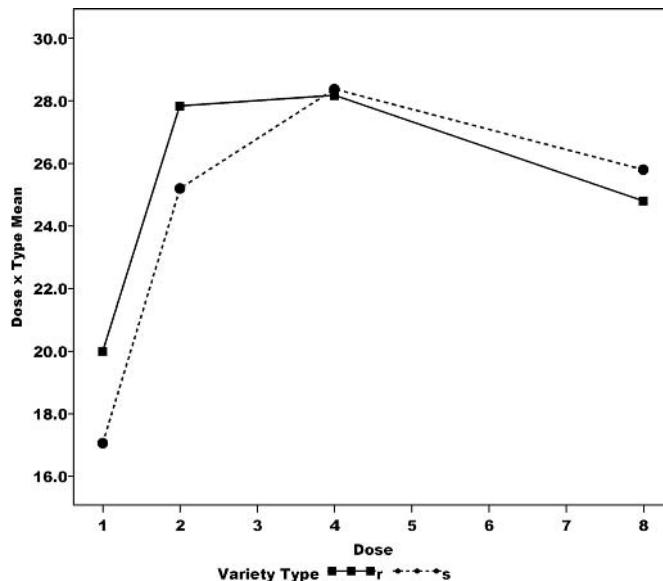
axis1 value=(font=swiss h=2)
      label=(angle=90 f=swiss h=2 'Dose x Type Mean');
axis2 value=(font=swiss h=2)
      label=(f=swiss h=2 'Dose');
axis3 value=(font=swiss h=2)
      label=(f=swiss h=2 'LogDose');
legend1 value=(font=swiss h=2)
      label=(f=swiss h=2 'Variety Type');
symbol1 color=black interpol=join line=1 value=square;
symbol2 color=black interpol=join line=2 value=circle;

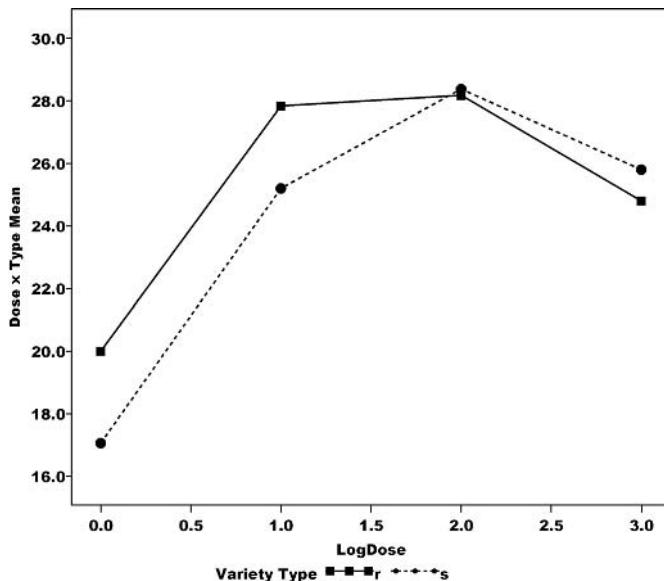
proc gplot data=to_plot;
  plot estimate*logdose=type/haxis=axis3
        vaxis=axis1
        legend=legend1;
  plot estimate*dose    =type/haxis=axis2
        vaxis=axis1
        legend=legend1;
run;

```

The ODS statement saves the least-squares means to a new data set, in this case called TYPEDOSEMEANS. The AXIS and SYMBOL statements are optional; they are used by SAS/GPGRAPH to make the interaction plot more readable. A DATA step, creating the data set TO_PLOT, is added here to allow plotting mean response to Log₂(dose) as well as unadjusted DOSE level. The reason for doing this will be clear in subsequent analysis discussed below. Figure 4.2 shows the interaction plots over DOSE and LOGDOSE.

Figure 4.2 Interaction Plots: Means over DOSE and LOGDOSE by TYPE





You can see from the plots that the susceptible TYPE has noticeably lower mean response at low levels of DOSE. You can also see that the resistant TYPE attains maximum mean response at approximately DOSE level 2, whereas the susceptible does not attain maximum response until DOSE level 4. The mean response of both varieties decreases somewhat when DOSE level increases from 4 to 8, but the decrease is somewhat less for the susceptible variety. You can also see that the response over LOGDOSE appears to have more of a parabolic shape than the response over DOSE, suggesting that a quadratic regression model in LOGDOSE might fit better. The next section pursues this idea.

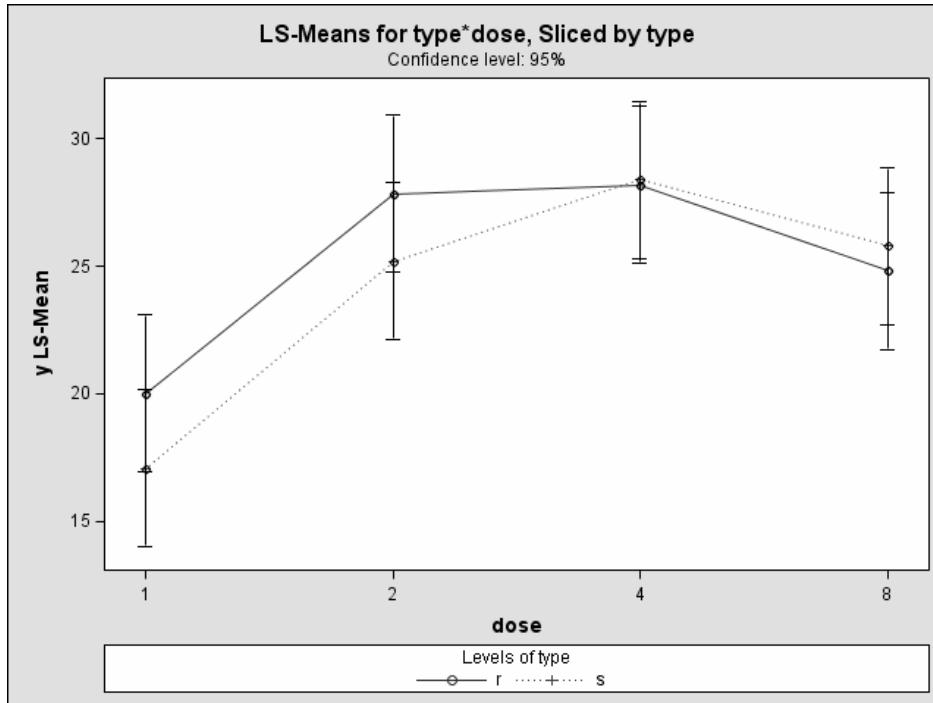
You can also obtain plots of least-squares means with the GLIMMIX procedure and ODS Graphics. The following statements produce the plot in Figure 4.3:

```

ods html;
ods graphics on;
ods select MeanPlot;
proc glimmix data=variety_eval;
  class block type dose;
  model y = type|dose / ddfm=satterth;
  random block block*dose;
  lsmeans type*dose / plot=meanplot(sliceby=type join cl);
run;
ods graphics off;
ods html close;

```

The PLOT=MEANPLOT option of the LSMEANS statement requests a plot of the interaction least-squares means. The sub-options of the MEANPLOT option control the layout of the plot. The SLICEBY=TYPE sub-option specifies the effect whose means are distinguished in the plot. The effects not involved in the SLICEBY= effect determine the levels for the horizontal axis. The JOIN option connects the least-squares means that belong to the same type, and the CL option adds (95%) confidence limits. Note, however, that the DOSE variable is listed in the CLASS statement. Hence, it is treated as a classification variable, which produces equal spacing of the ticks on the horizontal axis. Figure 4.2, on the other hand, treats the DOSE variable as continuous.

Figure 4.3 Interaction Plot from PROC GLIMMIX

4.6.2 Regression Analysis over DOSE by TYPE

The interaction plots in Figures 4.2 and 4.3 suggest that one way to characterize the variety evaluation data is to look at polynomial regression effects of dose for each type.

The “traditional” approach to regression analysis in the context of analysis of variance uses orthogonal polynomial contrasts. For completeness, this method is shown below. However, instead of orthogonal polynomials, we strongly recommend the “direct regression approach” shown later in this section. It is easier to implement and provides a more complete analysis. Orthogonal polynomials can be very tedious, especially in conjunction with factorial treatment structures, and needed statistics can be very difficult to obtain. These difficulties largely disappear when you use the direct regression approach. For additional background refer to Littell et al. (2002, Ch. 7).

Orthogonal Polynomial Analysis

Use the following SAS statements to compute the PROC MIXED regression analysis with orthogonal polynomial contrasts:

```
proc mixed data=variety_eval;
  class block type dose;
  model y = dose|type / ddfm=satterth;
  random block block*dose;
  contrast 'dose linear'    dose -11 -7   1 17;
  contrast 'dose quad'      dose  20 -4 -29 13;
  contrast 'dose cubic'     dose -8 14 -7  1;
  contrast 'type x linear'  dose*type -11 -7   1 17 11   7 -1 -17;
  contrast 'type x quad'    dose*type  20 -4 -29 13 -20  4 29 -13;
  contrast 'type x cubic'   dose*type -8 14 -7  1   8 -14  7 -1;
run;
```

The CONTRAST statements allow you to test polynomial regression, specifically the linear, quadratic, and cubic main effects and their interaction with TYPE. The coefficients reflect the geometric spacing of the dose levels 1, 2, 4, and 8. You can obtain these coefficients using the ORPOL function in PROC IML. Output 4.15 shows the results.

Output 4.15 Orthogonal Polynomial Results for DOSE Effect in Variety Evaluation Data

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
dose linear	1	12	7.74	0.0166
dose quad	1	12	27.01	0.0002
dose cubic	1	12	6.14	0.0291
type x linear	1	16	5.82	0.0282
type x quad	1	16	0.68	0.4205
type x cubic	1	16	0.36	0.5582

Interpretation

The TYPE \times LINEAR contrast indicates that the linear component of the regression over dose levels differs between the resistant and susceptible variety type. This is consistent with the plot over DOSE in Figure 4.2 (4.3) showing a greater increase over dose level for the susceptible TYPE. There is no evidence that the quadratic or cubic components differ by TYPE, but there is strong evidence ($p = 0.0291$) that a quadratic regression would not fit that data and that a cubic (or possibly something else—with only 1 d.f. beyond quadratic you cannot be sure) is required.

Alternatively, you could fit $\text{Log}_2(\text{DOSE})$, converting the dose levels to equally spaced levels 0, 1, 2, and 3. To do this, substitute the following CONTRAST statements:

```
contrast 'logdose linear' dose -3 -1 1 3;
contrast 'logdose quad' dose 1 -1 -1 1;
contrast 'logdose cubic' dose -1 3 -3 1;
contrast 'type x linear' dose*type -3 -1 1 3 3 1 -1 -3;
contrast 'type x quad' dose*type 1 -1 -1 1 -1 1 1 -1;
contrast 'type x cubic' dose*type -1 3 -3 1 1 -3 3 -1;
```

These coefficients are the standard orthogonal polynomial coefficients (given in most statistical methods textbooks) for four levels. The results appear in Output 4.16.

Output 4.16 Orthogonal Polynomial Results for Log-Dose in Variety Evaluation Data

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
logdose linear	1	12	18.25	0.0011
logdose quad	1	12	22.54	0.0005
logdose cubic	1	12	0.08	0.7780
type x linear	1	16	6.22	0.0240
type x quad	1	16	0.04	0.8515
type x cubic	1	16	0.61	0.4472

The results are similar except that there is now no evidence of lack of fit for a quadratic regression over Log(Dose) levels. The *p*-value for the “logdose cubic” term, which tests lack of fit from quadratic, is 0.7780.

The next logical step would be to explore the shape of the regression equations over Log DOSE for each type. You can do this with the following PROC MIXED statements:

```
proc mixed data=variety_eval;
  class block type dose;
  model y = type dose(type) / ddfm=satterth;
  random block block*dose;
  contrast 'linear LogD' type R' dose(type) -3 -1 1 3 0;
  contrast 'quad LogD' type R' dose(type) 1 -1 -1 1 0;
  contrast 'cubic LogD' type R' dose(type) -1 3 -3 1 0;
  contrast 'linear LogD' type S' dose(type) 0 0 0 -3 -1 1 3;
  contrast 'quad LogD' type S' dose(type) 0 0 0 1 -1 -1 1;
  contrast 'cubic LogD' type S' dose(type) 0 0 0 -1 3 -3 1;
run;
```

This model exploits the fact that the sum of SS(DOSE) and SS(DOSE \times TYPE) is equal to the sum of squares for the nested effect of dose within type, i.e., SS[DOSE(TYPE)]. That is, $\mu_{ij} = \mu + \tau_j + \delta(\tau)_{j(i)}$, where $\delta(\tau)_{j(i)}$ is the nested effect of dose within type, is a valid alternative for modeling the DOSE \times TYPE means, with $\delta(\tau)_{ij}$ replacing $\delta_i + (\delta\tau)_{ij}$ in the model. Note that there is no need for the DOSE main effect to appear in the MODEL statement in order for BLOCK \times DOSE, the term that identifies whole-plot error, to remain in the RANDOM statement. The CONTRAST statements are simply the linear, quadratic, and cubic orthogonal polynomial contrasts for each level of TYPE, one at a time, with the coefficients for the other type set to zero. The results appear in Output 4.17.

Output 4.17 Order of Polynomial Regression Fit by TYPE

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
linear LogD type R	1	19.5	6.15	0.0224
quad LogD type R	1	19.5	17.82	0.0004
cubic LogD type R	1	19.5	0.40	0.5322
linear LogD type S	1	19.5	24.47	<.0001
quad LogD type S	1	19.5	16.26	0.0007
cubic LogD type S	1	19.5	0.02	0.8943

You can see from this output that a quadratic regression over log(DOSE) for both cultivar types is appropriate. Both “quad LogD” contrasts (given type R or S) are statistically significant ($p = 0.0004$ and 0.0007 , respectively, for type R and S), and both the “cubic LogD” contrasts are nonsignificant ($p = 0.5322$ for type S, $p = 0.8943$ for type R).

Direct Regression Model

Output 4.17 suggests that μ_{ij} can be modeled in yet a third way:

$$\mu_{ij} = \mu + \delta_i + (\beta_L + \Delta_{Li})[\log_2(\text{dose})] + (\beta_Q + \Delta_{Qi})[\log_2(\text{dose})]^2$$

where

β_L is the linear slope coefficient for the regression over Log₂(DOSE)

Δ_{Li} is the change in slope for the i^{th} type

β_Q is the quadratic coefficient for the regression over Log₂(DOSE)

Δ_{Qi} is the change in quadratic coefficient for the i^{th} type

You can estimate this model with the following SAS statements:

```
proc mixed data=variety_eval;
  class block type dose;
  model y=type|logdose|logdose / ddfm=satterth
                                solution htype=1;
  /* model y=type|logdose|logdose|logdose/ddfm=satterth
                                solution htype=1; */
  random block block*dose;
run;
```

You have to create the variable LOGDOSE=LOG2(DOSE) in the DATA step. You define DOSE as a class variable in the CLASS statement in order to define whole-plot error (BLOCK \times DOSE) in the RANDOM statement. However, you treat LOGDOSE as a regression variable, so it is not included in the CLASS statement. The vertical bars in the effects specification of the MODEL statement cause MIXED to compute all possible interactions of the variables in the list. It is equivalent to the statement

```
model y = type logdose type*logdose logdose*logdose
          type*logdose*logdose;
```

In the MODEL statement, TYPE corresponds to δ_i , LOGDOSE to β_L , TYPE*LOGDOSE to Δ_{L_i} , LOGDOSE*LOGDOSE to β_Q , and TYPE*LOGDOSE*LOGDOSE to Δ_{Q_i} . The SOLUTION option requests a listing of the fixed effects estimates. The HTYPE=1 option causes the hypotheses to be tested in sequential order (like TYPE 1 SS in PROC GLM). You want to do this because the default TYPE 3 tests adjust effects mentioned first in the model (e.g., linear effects) by terms mentioned later in the model (e.g., quadratic effects). In fitting polynomial regression models, you want tests that enter the model terms sequentially, and hence you must override the default to obtain the needed hypothesis tests. Finally, the SAS statements contain an alternative model (commented out). This model includes the cubic effects. You can run this model as an alternative to the orthogonal polynomials for testing “type*linear,” “type*quad,” and “type*cubic” shown above. Using the HTYPE=1 option, the results are equivalent.

Output 4.18 shows the results.

Output 4.18 Regression of LOGDOSE on Y by TYPE Using Direct Regression

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
type	1	17	2.84	0.1101
logdose	1	13	19.63	0.0007
logdose*type	1	17	6.37	0.0219
logdose*logdose	1	13	24.25	0.0003
logdose*logdose*type	1	17	0.04	0.8497

Solution for Fixed Effects						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		17.0200	1.4204	19.6	11.98	<.0001
type	r	3.1690	1.2662	17	2.50	0.0228
type	s	0
logdose		10.9800	2.0181	21.4	5.44	<.0001
logdose*type	r	-1.0910	2.0334	17	-0.54	0.5985
logdose*type	s	0
logdose*logdose		-2.6800	0.6447	21.4	-4.16	0.0004
logdose*logdose*type	r	-0.1250	0.6495	17	-0.19	0.8497
logdose*logdose*type	s	0

Interpretation

The “Type 1 Test of Fixed Effects” gives tests for the various model effects. The row labeled “logdose*logdose*type” tests $H_0: \Delta_{Q_i} = 0$ for all levels of i . Similarly, the row labeled “logdose*logdose” tests $H_0: \beta_Q = 0$, “logdose*type” tests $H_0: \Delta_{L_i} = 0$ for all i , and “logdose” tests $H_0: \beta_L = 0$. These tests suggest that while two types have different linear slopes (i.e., the

Δ_{L_i} terms are nonzero), they share a common quadratic LOGDOSE effect (that is, the model term $\beta_Q + \Delta_{Qi}$ can be reduced to β_Q only).

The “Solutions for Fixed Effects” table allows you to estimate the regression equation over LOGDOSE for each TYPE. For example, the intercept for TYPE R is obtained by adding the estimate in the “Intercept” row (which corresponds to $\hat{\mu}$ in the model) and the estimate of “type r” (which corresponds to δ_R in the model equation). The intercept for the TYPE R regression equation is thus $17.020 + 3.169 = 20.189$. Similarly, the slope for TYPE R is $\hat{\beta}_L + \hat{\Delta}_{LR}$, the sum of the “logdose” and “logdose*type R” estimates, $10.98 - 1.091 = 9.889$, and the quadratic coefficient is $\hat{\beta}_Q + \hat{\Delta}_{QR}$, the sum of the “logdose*logdose” and “logdose*logdose*type R” estimates, $-2.86 - 0.125 = -2.805$. That is, for TYPE R, the regression equation is

$$\hat{\mu}_{Rj} = 20.189 + 9.889 \times \log_2(\text{dose}_j) - 2.805 \times [\log_2(\text{dose}_j)]^2$$

You can add the following ESTIMATE statements to the direct regression PROC MIXED program to obtain the regression parameters:

```
estimate 'B_0' for type R' intercept 1 type 1 0;
estimate 'B_Lin' for type R' logdose 1 logdose*type 1 0;
estimate 'B_quad' for type R' logdose*logdose 1
                           logdose*logdose*type 1 0;
estimate 'B_0' for type S' intercept 1 type 0 1;
estimate 'B_Lin' for type S' logdose 1 logdose*type 0 1;
estimate 'B_quad' for type S' logdose*logdose 1
                           logdose*logdose*type 0 1;
```

Output 4.19 shows the results.

Output 4.19 Estimates of Regression Model by TYPE

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
B_0 for type R	20.1890	1.4204	19.6	14.21	<.0001
B_Lin for type R	9.8890	2.0181	21.4	4.90	<.0001
B_quad for type R	-2.8050	0.6447	21.4	-4.35	0.0003
B_0 for type S	17.0200	1.4204	19.6	11.98	<.0001
B_Lin for type S	10.9800	2.0181	21.4	5.44	<.0001
B_quad for type S	-2.6800	0.6447	21.4	-4.16	0.0004

An alternative, and generally easier to understand, regression model is

$$\mu_{ij} = \beta_{0i} + \beta_{Li} [\log_2(\text{dose})] + \beta_{Qi} [\log_2(\text{dose})]^2$$

where

$\beta_{0i} = \mu + \delta_i$ is the intercept for the i^{th} type

$\beta_{Li} = \beta_L + \Delta_{Li}$ is the slope for the i^{th} type

$\beta_{Qi} = \beta_Q + \Delta_{Qi}$ is the quadratic regression parameter for the i^{th} type

You can estimate this model using PROC MIXED as follows:

```
proc mixed data=variety_eval;
  class block type dose;
  model y = type logdose(type) ld_sq(type) / noint ddfm=satterth
                                             solution;
  random block block*dose;
  contrast 'equal quad by type?' ld_sq(type) 1 -1;
run;
```

In the MODEL statement, the effect LOGDOSE(TYPE) creates a linear regression parameter for each TYPE. Similarly, LD_SQ(TYPE) creates a quadratic parameter for each TYPE. Note that you must create a new variable for the quadratic effect: the MODEL statement will not accept the syntax LOGDOSE*LOGDOSE(TYPE). In this case, the new variable is named LD_SQ; you create it using the statement LD_SQ= LOGDOSE*LOGDOSE in a DATA step. The NOINT option suppresses the SAS default of inserting μ in the model automatically. With the NOINT option, TYPE produces the parameter β_{0i} rather than the default $\mu + \delta_i$. This makes the SOLUTION output easier to read directly to get the regression equation. The CONTRAST statement is equivalent to LOGDOSE*LOGDOSE*TYPE in the previous SAS program: it tests $H_0: \beta_{QR} = \beta_{QS}$, which is equivalent to testing $H_0: \Delta_{Qi} = 0$ in the previous model. That is, if H_0 is true, then both β_{QR} and β_{QS} can be set equal to β_Q , i.e., $\Delta_{Qi} = 0$. Output 4.20 shows the results.

Output 4.20 Direct Regression by TYPE Using Nested Model

Solution for Fixed Effects						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
type	r	20.1890	1.4204	19.6	14.21	<.0001
type	s	17.0200	1.4204	19.6	11.98	<.0001
logdose(type)	r	9.8890	2.0181	21.4	4.90	<.0001
logdose(type)	s	10.9800	2.0181	21.4	5.44	<.0001
ld_sq(type)	r	-2.8050	0.6447	21.4	-4.35	0.0003
ld_sq(type)	s	-2.6800	0.6447	21.4	-4.16	0.0004

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
equal quad by type?	1	17	0.04	0.8497

Interpretation

You can see that the regression estimates and their standard errors are identical to those obtained in the previous PROC MIXED program. The advantage of this approach is that you do not have to write ESTIMATE statements to make the SOLUTION output directly usable. You can also use the t -values to test the null hypotheses about each regression parameter. For example, the LD_SQ(TYPE) t -values test $H_0: \beta_{Qi} = 0$ for each TYPE. These are equivalent to

the orthogonal polynomial contrasts shown in the nested model previously, except that this approach is much easier. You do not need to look up the contrast coefficients, or compute them for unequally spaced DOSE levels, or enter the several CONTRAST statements required. Also note that the F -value for ‘equal quad by type?’ contrast is the same as the test for “logdose*logdose*type” in the previous direct regression program (Output 4.18). Thus, the nested program shown here is the most versatile for most analyses once you decide to model a quantitative factor such as DOSE by polynomial regression.

Extensions

The direct regression methods shown here are closely related to the **analysis of covariance**, which is considered in greater detail in Chapter 8. In many applications with quantitative factors, **nonlinear regression** models are more appropriate, provide a better fit, or provide more interpretable parameter estimates than polynomial regression. These are considered in Chapter 15.

The Final Fit

Returning to the variety evaluation data, because the analysis indicated that the quadratic regression coefficients should be considered equal, a parsimonious model would be

$$\mu_{ij} = \beta_{0i} + \beta_{Li} [\log_2(\text{dose})] + \beta_Q [\log_2(\text{dose})]^2$$

That is, the same quadratic parameter, β_Q , is used for both types but separate intercepts, β_{0i} , and linear parameters, β_{Li} , are retained. Use the following SAS program to estimate the final model:

```
proc mixed data=variety_eval;
  class block type dose;
  model y=type logdose(type) ld_sq /noint ddfm=satterth solution;
  random block block*dose;
run;
```

The results are shown in Output 4.21.

Output 4.21 Final Regression Model Parameter Estimates for Variety Evaluation Data

Solution for Fixed Effects						
Effect	type	Estimate	Standard Error	DF	t Value	Pr > t
type	r	20.2515	1.3770	17.8	14.71	<.0001
type	s	16.9575	1.3770	17.8	12.31	<.0001
logdose(type)	r	9.7015	1.7660	13.7	5.49	<.0001
logdose(type)	s	11.1675	1.7660	13.7	6.32	<.0001
ld_sq		-2.7425	0.5569	13	-4.92	0.0003

For example, from the output, the regression model for TYPE R is

$$\hat{\mu}_{Rj} = 20.2515 + 9.7015 \times \log_2(\text{dose}_j) - 2.7425 [\log_2(\text{dose}_j)]^2$$

4.7 Example: Variance Component Estimates Equal to Zero

The semiconductor and variety evaluation data both had whole-plot error variance estimates greater than zero. As shown with the semiconductor data, when the variance estimates are all positive, mixed model inference on data with split-plot features is identical to standard analysis of variance methods. However, when one or more of the variance component estimates is negative, then there are issues in mixed model inference that require extra care. This happens, for example, when the whole-plot error mean square is less than the split-plot error mean square. This example illustrates the main issues.

The example is a trial involving laboratory mice. There are four different housing conditions (denoted CONDITION 1, 2, 3, and 4) and three feeding regimes (denoted DIET: “restricted,” “normal,” and “supplement”) under investigation. Two conditions can be handled in a single CAGE unit. Within each condition at each cage unit, mice can be separated into three diet groups. Thus, CAGE is a blocking factor, CONDITION is the whole-plot treatment factor, CAGE \times CONDITION combinations are the whole-plot experimental units, and DIET is the split-plot factor. The layout is similar to that of Figure 4.1.e, except that the whole plots form an incomplete block design rather than a randomized complete block design. The data are given in Data Set 4.7, “Mouse Condition—Diet Experiment,” in Appendix 2, “Data Sets.”

Following Table 4.2, a model for these data is

$$Y_{ijk} = \mu_{ij} + c_k + w_{ik} + e_{ijk}$$

where

μ_{ij} is the mean response for CONDITION i and DIET j

c_k is the effect of the k^{th} cage unit, assumed iid $N(0, \sigma_c^2)$

w_{ik} is the effect of the ik^{th} whole-plot e.u. (cage \times condition) assumed iid $N(0, \sigma_w^2)$

e_{ijk} is the effect of the ijk^{th} split-plot e.u. effect, assumed iid $N(0, \sigma^2)$

4.7.1 Default Analysis Using PROC MIXED

Following the model and the approach used in the previous examples in this chapter, you obtain the basic analysis of these data using the following SAS statements:

```
proc mixed data=mice;
  class cage condition diet;
  model gain = condition|diet / ddfm=satterth;
  random cage cage*condition;
run;
```

Output 4.23 shows the results.

Output 4.23 Standard PROC MIXED Analysis of Mouse Data

Covariance Parameter Estimates	
Cov Parm	Estimate
cage	3.0373
cage*condition	0
Residual	27.8431

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	23.6	2.71	0.0677
diet	2	20.2	0.93	0.4090
condition*diet	6	20.2	1.73	0.1661

Comments

First, note that the estimate for the whole-plot error (CAGE*CONDITION) variance component is 0. If you look in the SAS log,² you will see a note:

NOTE: Estimated G matrix is not positive definite.

The note is followed by some additional commentary regarding the degrees of freedom. The degrees-of-freedom comment refers to the denominator degrees of freedom in the “Type 3 Tests of Fixed Effects” output. Output 4.24 shows the analysis of variance of these data, which you can obtain using either PROC GLM or PROC MIXED as shown earlier in this chapter (Example: A Split-Plot Semiconductor Experiment). Using analysis of variance, there are 3 degrees of freedom for the whole-plot error, instead of 23.6 shown in Output 4.23, and the ANOVA degrees of freedom for the split-plot error is 16, not 20.2 as shown above. Also, the *F*-value for the CONDITION main effect is 2.71 using the standard PROC MIXED approach, whereas the analysis of variance *F*, the ratio of MS(CONDITION) to MS(CAGE*CONDITION), is 3.95.

You can drop the DDFM=SATTERTH option in PROC MIXED to make the degrees of freedom agree with the analysis of variance—and, more importantly, with the structure of the design—but the *F*-value is still 2.71. Obviously, there is a discrepancy between the analysis of variance and the mixed model results you get with PROC MIXED. What is going on? Which should you use?

Output 4.24 Analysis of Variance for Mouse Data—PROC GLM

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cage	5	198.277778	39.655556	2.73	0.2185
condition	3	171.666667	57.222222	3.95	0.1446
Error	3	43.500000	14.500000		

² If you use the GLIMMIX procedure, the text of the note regarding the G matrix not being positive definite is also displayed in the regular output, following the “Iteration History” and “Convergence Status” tables.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Error: MS(cage*condition)					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cage*condition	3	43.500000	14.500000	0.46	0.7144
diet	2	52.055556	26.027778	0.82	0.4561
condition*diet	6	288.388889	48.064815	1.52	0.2333
Error: MS(Error)	16	504.888889	31.555556		

The primary issue with the default PROC MIXED analysis is the REML estimation procedure for variance components. By definition, a variance component cannot be negative, so when the REML computing algorithm obtains a negative solution, it is set to 0. In theory, this could happen to either the CAGE variance, σ_C^2 , or the whole-plot error variance, σ_W^2 . Setting either estimate to zero has a ripple effect on all subsequent analysis.

4.7.2 Recommended Alternative: Analysis with NOBOUND or METHOD=TYPE3 Options

Stroup and Littell (2002) discussed the impact of negative variance component estimates on tests of fixed effects in mixed model analysis. Basically, it is easy to show by simulation that if the block variance (in this case CAGE) is set to zero, the whole-plot error variance tends to be underestimated, which in turn means that the whole-plot F -value is overestimated. This results in an inflated Type I error rate. On the other hand, what you see for the mouse data in this example is the impact of a negative whole-plot error variance estimate. If the whole-plot error variance is set to zero, the whole-plot main effect (CONDITION) F -statistic is in essence based on MS(split-plot). This results in an underestimate of the F -statistic. At the same time, the split-plot and whole-plot error degrees of freedom are pooled, with some additional problems with the Satterthwaite procedure, which assumes positive variance component estimates. The net result is a tendency for the Type II error rate to be inflated, although this depends on the extent to which whole-plot denominator degrees of freedom are affected.

On the other hand, if you do **not** set the negative variance component estimate to zero, but allow it to remain negative, you get better control over Type I error and, for cases of negative whole-plot error variance estimates, greater power. Therefore, this is the recommended procedure. You can override the PROC MIXED set-to-zero default in one of two ways. You can either use the NOBOUND option in the PROC MIXED statement, i.e.,

```
proc mixed nobound;
```

Or you can use the METHOD=TYPE3 option. The NOBOUND option removes the lower bound of 0 from the variance components. The METHOD=TYPE3 causes MIXED to estimate variance components based on the expected mean squares instead of computing REML solutions. For unbalanced data, such as incomplete block designs, the results are not equal, but they are usually close unless the data are severely unbalanced. Output 4.25 shows the results for NOBOUND and METHOD=TYPE3.

Output 4.25 NOBOUND and METHOD=TYPE3 Results Overriding Set-to-Zero Default

NOBOUND

Covariance Parameter Estimates	
Cov Parm	Estimate
cage	5.0292
cage*condition	-6.2415
Residual	31.5567

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	4.72	4.31	0.0798
diet	2	16	0.82	0.4561
condition*diet	6	16	1.52	0.2334

Method=Type3

Covariance Parameter Estimates	
Cov Parm	Estimate
cage	5.2407
cage*condition	-5.6852
Residual	31.5556

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	3.73	3.70	0.1271
diet	2	15.9	0.82	0.4562
condition*diet	6	15.9	1.52	0.2336

Comments

These results differ slightly. The CAGE variance estimates are 5.0292 for NOBOUND versus 5.2407 for TYPE3. The F -values for CONDITION are 5.16 and 4.61, respectively. Neither is exactly equal to the analysis of variance ($F = 3.95$), but they are close. From simulation studies, there is no evidence suggesting any advantage for analysis of variance, unbounded REML, or METHOD=TYPE3 estimation. Clearly, however, you want to use either NOBOUND or METHOD=TYPE3 with PROC MIXED to estimate or test simple effects, slices, and other quantities that GLM-based analysis of variance cannot compute.

The degrees of freedom in Output 4.25 use the DDFM=KR option rather than DDFM=SATTERTH. The KR option uses the procedure developed by Kenward and Roger (1997). Chapter 5 introduces the KR option in more detail. Because the KR option was specifically intended to account for unbalanced data in conjunction with multiple random effects or any model with correlated errors (balanced or not), it is recommended for all such applications.

4.7.3 Conceptual Alternative: Negative Variance or Correlation?

Using the NOBOUND or METHOD=TYPE3 option presents a dilemma when you report the results. By definition, a variance must be nonnegative, yet you must report a negative variance estimate in order to control type I error or preserve power. Many authors and journal editors object to such an apparent contradiction. To address these objections, a conceptual alternative can help.

In the variance components model used for the mouse data in the previous sections, the ratio

$$\frac{\sigma_w^2}{\sigma_w^2 + \sigma^2}$$

is termed the **intraclass correlation**. That is, the whole-plot experimental units acts as blocks with respect to the split-plot experimental units. Observations within the same block are correlated. The intraclass correlation is a measure of this correlation.

Because variance components must be nonnegative, variance component models implicitly assume that the intraclass correlation is nonnegative. However, there is no conceptual reason why correlation among observations within the same whole-plot experimental unit *must* be nonnegative. In fact, in many practical situations, there are interference or competition effects among adjacent experimental units that manifest themselves in negative correlation. What appears as a negative variance component estimate in the analyses from the previous sections may in fact be negative interclass correlation. In such cases, you need a more realistic model.

Compound symmetry is an alternative model of intraclass correlation that allows for negative correlation. Define $\mathbf{e}_{ik} = [e_{i1k} \ e_{i2k} \ \dots \ e_{iBk}]$ as the vector of split-plot error terms for the ik^{th} whole-plot experimental unit. In the compound symmetry model, $\text{Var}[\mathbf{e}_{ik}]$ is defined as

$$\sigma_{CS}^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

where σ_{CS}^2 is the variance of a split-plot error and ρ is the correlation between any pair of split-plot errors. The model used in the previous sections is a special case, with $\sigma_{CS}^2 = \sigma_w^2 + \sigma^2$ and $\rho = \sigma_w^2 / (\sigma_w^2 + \sigma^2)$. The model can be more general.

The compound symmetry model is one of several types of mixed models that allow for **correlated errors**. They are used extensively in **repeated measures** analysis, introduced in detail in Chapter 5, and also in **spatial** statistics (Chapter 11).

You can fit the compound symmetry model using the following PROC MIXED statements:

```
proc mixed data=mice;
  class cage condition diet;
  model gain=condition|diet/ddfm=kr;
  random cage;
  repeated / type=cs subject=cage*condition r rcorr;
run;
```

The REPEATED statement replaces RANDOM CAGE*CONDITION. In general, you use the REPEATED statement to specify correlated error structures. The TYPE=CS option specifies the covariance structure as being of the compound-symmetric type. The SUBJECT=CAGE*CONDITION option specifies the unit within which observations are correlated, in this case the whole-plot experimental unit. The options R and RCORR request listings of the covariance and correlation matrices, respectively, within a given CAGE \times CONDITION unit. Output 4.26 shows the results.

Output 4.26 Compound Symmetry Analysis of Mouse Data

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
cage		5.0293
CS	cage*condition	-6.2417
Residual		31.5569

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	4.72	4.31	0.0798
diet	2	16	0.82	0.4561
condition*diet	6	16	1.52	0.2334

Estimated R Matrix for cage*condition 1 1			
Row	Col1	Col2	Col3
1	25.3152	-6.2417	-6.2417
2	-6.2417	25.3152	-6.2417
3	-6.2417	-6.2417	25.3152

Estimated R Correlation Matrix for cage*condition 1 1			
Row	Col1	Col2	Col3
1	1.0000	-0.2466	-0.2466
2	-0.2466	1.0000	-0.2466
3	-0.2466	-0.2466	1.0000

Interpretation

The “Covariance Parameter Estimates” table shows estimates identical to the NOBOUND results, apart from some relabeling and minor rounding discrepancies between computations based on the RANDOM statement and those based on the REPEATED statement. “CS” corresponds to the estimate of CAGE*CONDITION variance, $\hat{\sigma}_w^2$ in the NOBOUND output. Here, you interpret it as the covariance between pairs of split-plot experimental units. The row labeled “Residual” corresponds to the estimate of the split-plot error, $\hat{\sigma}^2$ in the NOBOUND output. The “Type 3 Tests of Fixed Effects” table is identical to Output 4.25. For the “Estimated R Matrix,” note that the terms on the diagonal are $\hat{\sigma}_{CS}^2 = \hat{\sigma}_w^2 + \hat{\sigma}^2 = 31.5569 - 6.2417 = 25.3152$, and the off-diagonal terms correspond to $\hat{\rho}$. In the “Estimated R Correlation Matrix” the off-diagonal terms are estimated interclass correlations,

$$\hat{\rho} = \frac{\hat{\sigma}_w^2}{\hat{\sigma}_w^2 + \hat{\sigma}^2} = \frac{-6.2417}{25.3152} = -0.2466$$

In general, the statements

```
random whole-plot error;
```

and

```
repeated / type=cs subject=whole-plot error;
```

produce identical analyses, and are in fact identical models, provided the whole-plot variance component estimate is positive. If the NOBOUND option is used with the RANDOM statement, these two statements will also yield identical results. The main difference is that the compound symmetry analysis obtained using the REPEATED statement lends itself to a more palatable, and often more physically accurate, interpretation.

4.8 More on PROC GLM Compared to PROC MIXED: Incomplete Blocks, Missing Data, and Estimability

Section 4.5 used the semiconductor data example to illustrate several problems with PROC GLM compared to PROC MIXED as a tool for analyzing multiple-error term factorial designs. The mouse data analysis in Section 4.7 reveals another problem, specific to the fact that the mouse data are unbalanced. In this case, the incomplete-block whole plot creates the unbalance, but unbalance also occurs when experiments are set up with complete-block whole plots but then data are lost or missing for various reasons. The problem illustrated in this section is especially serious if data from entire whole plots are missing.

Section 4.7.1 showed the analysis of variance output from PROC GLM for the mouse data. If you pursue this approach, you would logically want to include inference on the various main effect or simple effect means, and perhaps other contrasts as discussed in earlier examples. You could use the following SAS statements, which are not exhaustive, but are sufficient to illustrate why you do not want to use PROC GLM to analyze this type of data.

```
proc glm data=mice;
  class cage condition diet;
  model gain=condition|diet cage cage*condition;
  lsmeans condition diet condition*diet;
run;
```

Output 4.27 shows the results.

Output 4.27 PROC GLM Output of Least-Squares Means for Mouse Data

condition	gain LSMEAN
1	Non-est
2	Non-est
3	Non-est
4	Non-est

diet	gain LSMEAN
normal	57.9166667
restrict	55.5000000
suppleme	58.1666667

condition	diet	gain LSMEAN
1	normal	Non-est
1	restrict	Non-est
1	suppleme	Non-est
2	normal	Non-est
2	restrict	Non-est
2	suppleme	Non-est
3	normal	Non-est
3	restrict	Non-est
3	suppleme	Non-est
4	normal	Non-est
4	restrict	Non-est
4	suppleme	Non-est

You can see that no least-squares means appear for the CONDITION or CONDITION \times DIET effects. Instead, all you see is “Non-est.” This means that PROC GLM has declared the CONDITION and CONDITION \times DIET least-squares means to be **nonestimable**. However, these means *should* be estimable, so what is the problem?

Referring to Section 4.3.2, least-squares means are defined as linear combinations of the fixed effects. Specifically, the CONDITION main effect least-squares means are defined as

$$\mu + \alpha_i + \frac{1}{3} \sum_j \beta_j + \frac{1}{3} \sum_j (\alpha\beta)_{ij}$$

where α and β refer to CONDITION and DIET, respectively. The CONDITION*DIET combination means are defined as $\mu + a_i + \beta_j + (\alpha\beta)_{ij}$. Note that these definitions do not include random effects. In order for PROC GLM and PROC MIXED to compute least-squares means, they must be **estimable**, a linear model criterion meaning that the least-squares mean must

correspond to a linear combination of expected values of observations in the data set. The PROC MIXED-based least-squares means for CONDITION and CONDITION \times DIET correspond to $\mu_{i\bar{}}$ and μ_{ij} respectively, so there is no problem.

On the other hand, PROC GLM is not set up to distinguish between fixed and random effects. Instead, it defines the LSMEANS for CONDITION as

$$\mu + \alpha_i + \frac{1}{3} \sum_j \beta_j + \frac{1}{3} \sum_j (\alpha\beta)_{ij} + \frac{1}{6} \sum_k c_k + \frac{1}{6} \sum_k w_{ik}$$

and for CONDITION \times DIET as

$$\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \frac{1}{6} \sum_k c_k + \frac{1}{6} \sum_k w_{ik}$$

including *all* effects in the model statement, regardless of whether they are fixed or random. However, three of the w_{ik} (CAGE \times CONDITION combinations) are not observed. Thus, PROC GLM sets up the *estimate* of the CONDITION least-squares mean as

$$\hat{\mu} + \hat{\alpha}_i + \frac{1}{3} \sum_j \hat{\beta}_j + \frac{1}{3} \sum_j (\hat{\alpha}\hat{\beta})_{ij} + \frac{1}{6} \sum_k \hat{c}_k + \frac{1}{3} \sum_k I_{ik} \hat{w}_{ik}$$

where I_{ik} is equal to 1 if the ik^{th} CAGE \times CONDITION combination is included in the design and 0 otherwise. However, this term fails the **estimability** criterion. That is, it cannot be expressed in terms of the expected values of the observations in the data set. The CONDITION \times DIET least-squares mean is similarly affected by the absent CAGE \times CONDITION combination.

The problem with the PROC GLM approach is that it applies the estimability criterion inappropriately. There is no way to properly specify the model and get GLM to compute least-squares means for any treatment effects involving the whole-plot factor unless it is present in all whole-plot blocks. The only way PROC GLM will find estimable least-squares means is to drop CAGE \times CONDITION from the MODEL statement—that is, drop the whole-plot error term from the analysis. If you do this, you will get results from the LSMEANS statement, but they will be incorrect for two reasons. First, they will not be adjusted for the effect of incomplete blocks. Second, all standard errors and tests will be based on an “error term” that pools MS(whole-plot error) with MS(split-plot error), whatever such a term estimates. In other words, you cannot use PROC GLM to analyze experiments with split-plot features if they include incomplete block structure or missing whole plots. This is another reason why we recommend the use of PROC MIXED and discourage the use of PROC GLM for all experiments with split-plot structure. This includes repeated measures experiments, to be discussed in Chapter 5.

4.9 Summary

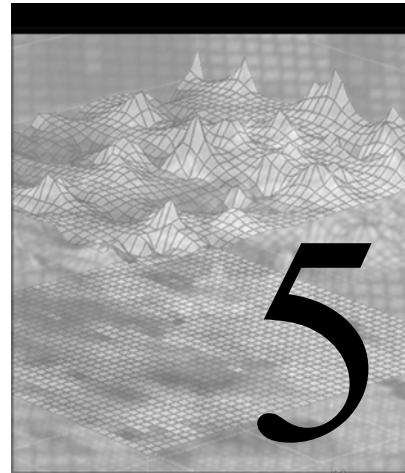
This chapter presented an extensive set of material required for analyzing factorial treatment structures with mixed models. It began by distinguishing between treatment structure and experiment structure, defining the factorial treatment structure and emphasizing that this treatment structure can be implemented using a variety of experiment structures. These range from simple completely randomized designs to complex split-plot layouts.

Next, the chapter showed how to associate the various layouts with an appropriate mixed model. It then presented the various terms of interest in factorial experiments and how their standard errors and test statistics are affected by the experiment layout and the associated mixed model.

Three examples illustrated how to use SAS to analyze factorial data. The examples all have split-plot structures to illustrate the full range of mixed model issues, but the model and mean comparison methods shown could be used with any factorial structure. The main requirement is to adjust the RANDOM statement so that it accurately reflects the experiment layout. The first example (Section 4.4) is a standard two-factor factorial with qualitative factors (i.e., the factor levels are categories). The second example (Section 4.6) illustrates a two-factor factorial with factors having quantitative levels. The third example (Section 4.7) shows the inference problems that result when one or more variance component estimates are negative and what to do when this happens.

Finally, this chapter compared PROC GLM and PROC MIXED as tools for analyzing experiments with split-plot features. The discussions focused on shortcomings of PROC GLM for this type of analysis, and why PROC MIXED should be the tool of choice. Section 4.5 presented the major issues, using Example 1 to illustrate. Section 4.7.4 presented additional problems with PROC GLM relative to PROC MIXED specific to unbalanced designs with multiple error terms.

Although this chapter is long, it is also merely introductory. Subsequent chapters expand on the range of analyses with factorial treatment design and layouts with split-plot features. Chapter 5 covers repeated measures experiments, where treatment and time form a factorial structure. Chapter 6 includes a section on multi-location experiments, an important and often misunderstood treatment structure with treatment and location as the factors. Later chapters consider split-plot experiments with categorical or other non-normally distributed response variables (Chapter 14, “Generalized Linear Mixed Models”), split-plot experiments similar to Example 2 but with nonlinear models to characterize the quantitative factor effect (Chapter 15, “Nonlinear Mixed Models”). Finally, case studies in Chapter 16 show more advanced split-plot structures: one with no replication, one in which the whole plot uses a central-composite response-surface design, and one in which the correlation is among whole-plot experimental units.



Analysis of Repeated Measures Data

5.1	Introduction	160
5.1.1	Basic Concepts of Repeated Measures	160
5.1.2	Types of Repeated Measures Analyses.....	161
5.1.3	A Statistical Model for Repeated Measures	161
5.2	Example: Mixed Model Analysis of Data from Basic Repeated Measures Design	163
5.2.1	Using the REPEATED Statement in PROC MIXED	165
5.2.2	Comparing Results from Two Covariance Structures.....	171
5.3	Modeling Covariance Structure	174
5.3.1	Some Candidate Covariance Structures.....	174
5.3.2	Selecting an Appropriate Covariance Model	177
5.3.3	Reassessing the Covariance Structure with a Means Model Accounting for Baseline Measurement.....	186
5.3.4	PROC MIXED Analysis of FEV1 Data	187
5.3.6	Inference on Treatment and Time Effects of FEV1 Data Using PROC MIXED.....	190
5.4	Example: Unequally Spaced Repeated Measures.....	198
5.5	Summary	202

5.1 Introduction

The term **repeated measures** refers to data sets with multiple measurements of a response variable on the same experimental unit. In most applications, the multiple measurements are made over a period of time. An example is growth curve data, such as monthly weight measurements of babies for the first year of their lives. Another example is drug effects data, such as measurements of pulse or respiration on patients following administration of a drug. But repeated measures can also refer to multiple measurements over space, such as thicknesses of the vertebrae of animals. In a general sense, any data that are measured repeatedly over time or space are repeated measures data. Most of this chapter uses the term in the more traditional sense, referring to sequences of measurements on experimental units in a designed experiment, sample survey, or retrospective study.

5.1.1 Basic Concepts of Repeated Measures

A commonly occurring repeated measures study consists of a completely randomized experimental design with data collected in a sequence of equally spaced time points from each experimental unit. Much of the development of repeated measures methodology occurred in the area of human psychology. As a result, the experimental units are often called **subjects**. But “subject” could refer to an animal, a laboratory sample, or a piece of industrial equipment.

In this basic setup of a completely randomized design with repeated measures, there are two factors, **treatments** and **time**. In this sense, all repeated measures experiments are factorial experiments. Treatment is called the **between-subjects** factor because levels of treatment can change only between subjects; all measurements on the same subject will represent the same treatment. Time is called a **within-subjects** factor because different measurements on the same subject are taken at different times. In repeated measures experiments, interest centers on (1) how treatment means differ, (2) how treatment means change over time, and (3) how differences between treatment means change over time. In other words, is there a **treatment main effect**, is there a **time main effect**, and is there a **treatment-by-time interaction**? These are the types of questions we may want to ask in any two-factor study. Ordinarily, the interaction would be the first question to investigate.

There is nothing peculiar about the objectives of a repeated measures study. What makes repeated measures data analysis distinct is the **covariance structure** of the observed data. In randomized blocks designs, treatments are randomized to units within a block. This makes all observations within a given block equally correlated. But, in repeated measures experiments, two measurements taken at adjacent time points are typically more highly correlated than two measurements taken several time points apart. Effort is usually needed at the beginning of the statistical analysis to assess the covariance structure of the data. Several pages in Sections 5.2.1 and 5.2.2 are devoted to modeling the covariance structure of the example data. Modeling an appropriate covariance structure is essential so that valid inference in the form of tests of hypotheses and confidence intervals can be made about the treatment means.

There are similarities between repeated measures experiments and split-plot experiments (see Chapter 4). The treatment factor in a repeated measures experiment corresponds to the main-plot factor in a split-plot experiment. The time factor in repeated measures corresponds to the sub-plot factor. In other words, the between-subjects factor corresponds to the main-plot factor, and the within-subjects factor corresponds to the sub-plot factor. The experimental units to which the treatments are assigned in the repeated measures experiment are analogous to main-plot units in the split-plot experiment, and the experimental units **at particular times** correspond to sub-plot units. However, in a genuine split-plot experiment, levels of the sub-plot factor are randomly assigned to sub-plot units within main-plot units. Consequently, responses

from different sub-plot units in the same main-plot unit are equally correlated with each other. But in repeated measures experiments, responses from points close in time are usually more highly correlated than responses from points far apart in time. Therefore, special methods of analysis are usually needed to accommodate the correlation structure of the repeated measures.

5.1.2 Types of Repeated Measures Analyses

Three general types of statistical analyses are most commonly used for repeated measures. One method treats repeated measures data as having come from a split-plot experiment. This method, often called a **univariate analysis of variance**, can be implemented in SAS using PROC GLM with the RANDOM statement, as was discussed in Chapter 2. Another method applies **multivariate and univariate analysis methods to linear transformations** of the repeated measures. The linear transformations can be means, differences between responses at different time points, slopes of regression curves, etc. These techniques are invoked by the REPEATED statement in PROC GLM, and are illustrated in Littell, Stroup, and Freund (2002). The third method applies **mixed model methods with special parametric structure on the covariance matrices**. This type of methodology has been computationally feasible only in recent years. It is applied in PROC MIXED, typically using the REPEATED statement, and is illustrated in this book.

As noted in the introduction to this chapter, mixed model analysis involves two stages: First, estimate the covariance structure. Second, assess treatment and time effects using generalized least squares with the estimated covariance. Littell, Pendergast, and Natarajan (2000) break these stages further into a four-step procedure for mixed model analysis:

- Step 1: Model the mean structure, usually by specification of the fixed effects.
- Step 2: Specify the covariance structure, between subjects as well as within subjects.
- Step 3: Fit the mean model accounting for the covariance structure.
- Step 4: Make statistical inference based on the results of step 3.

Other authors, such as Diggle (1988) and Wolfinger (1993a), recommend similar model-fitting and inference processes. The following sections show you how to use PROC MIXED to implement these four steps.

5.1.3 A Statistical Model for Repeated Measures

Consider the experimental situation described in Section 5.1.1. Subjects are randomly assigned to a treatment factor, and measurements are made at equally spaced times on each subject. Let Y_{ijk} denote the measurement at time k on the j^{th} subject assigned to treatment i .

Model

A statistical model for repeated measures data is

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk} \quad (5.1)$$

where

$\mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}$ is the mean for treatment i at time k , containing effects for treatment, time, and treatment \times time interaction

e_{ijk} is the random error associated with the measurement at time k on the j^{th} subject that is assigned to treatment i

Equation (5.1) is the same as the model equation for a standard factorial experiment with main effects of treatment and time, and treatment \times time interaction. The distinguishing feature of a repeated measures model is the variance and covariance structure of the errors, e_{ijk} . Although treatments were randomly assigned to subjects, the levels of the repeated measures factor, in this case time, is not randomly assigned to units within subjects. Thus we cannot reasonably assume that the random errors e_{ijk} for the same subject are independent. Instead, we assume that errors for *different* subjects are independent, giving

$$\text{Cov}[e_{ijk}, e_{ij'l}] = 0 \text{ if either } i \neq i' \text{ or } j \neq j' \quad (5.2)$$

Also, since measurement on the same subject are over a time course, they may have different variances, and correlations between pairs of measurements may depend on the length of the time interval between the measurements. Therefore, in the most general setting, we only assume

$$\text{Var}[e_{ijk}] = \sigma_k^2 \text{ and } \text{Cov}[e_{ijk}, e_{ijk'}] = \sigma_{kk'} \quad (5.3)$$

In other words, we allow that the variance of e_{ijk} depends on the measurement time k , and the covariance between the errors at two times, k and k' , for the same subject, depends on the times. In most cases, the model for the covariance can be expressed according to some structure involving fewer parameters. If we express the vector of observations on subject j in treatment i as $\mathbf{Y}_{ij} = [Y_{ij1}, \dots, Y_{ijT}]'$, then we have $\text{Var}[\mathbf{Y}_{ij}] = \boldsymbol{\Sigma}$, where the element in row k and column k' is $\sigma_{kk'}$. This assumes that the covariance matrix $\boldsymbol{\Sigma}$ is the same for all subjects. If we stack the \mathbf{Y}_{ij} vectors into a single vector $\mathbf{Y} = [\mathbf{Y}'_{11}, \mathbf{Y}'_{12}, \dots, \mathbf{Y}'_{nT}]'$, then $\text{Var}[\mathbf{Y}] = \mathbf{V}$ is block diagonal with $\boldsymbol{\Sigma}$ along the diagonal. We can write $\boldsymbol{\Sigma}$ as

$$\text{Var}[\mathbf{Y}] = \mathbf{I}_k \otimes \boldsymbol{\Sigma} \quad (5.4)$$

where \mathbf{I}_k is an identity matrix of dimension equal to the number of subjects.

In some situations, it is advantageous to include a between-subjects random effect to give the model

$$Y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ijk} \quad (5.5)$$

where b_{ij} is a random effect for subject j assigned to treatment i , and ε_{ijk} is an error with covariance matrix \mathbf{R} with a parametric structure. The covariance matrix of $\mathbf{Y}_{ij} = [Y_{ij1}, \dots, Y_{ijT}]'$ becomes

$$\boldsymbol{\Sigma} = \text{Var}[\mathbf{Y}_{ij}] = \sigma_b^2 \mathbf{J} + \mathbf{R} \quad (5.6)$$

where \mathbf{J} is a matrix of ones. Equation (5.6) shows the two aspects of covariance between measures on the same subject. The part $\sigma_b^2 \mathbf{J}$ represents the covariance due to the fact that the measures are on the same subject, and \mathbf{R} represents the contribution to the covariance due to the proximity of the measurements.

Model (5.5) is similar to the split-plot-type models encountered in Chapter 4, where b_{ij} corresponds to whole-plot errors and ε_{ijk} corresponds to sub-plot errors. The distinction is that the ε_{ijk} , which would correspond to the sub-plot errors, cannot necessarily be assumed independent and identically distributed because the time effects may not be equally correlated within subjects.

Repeated measures data can be analyzed using mixed model methods based on generalized least squares and maximum likelihood. In matrix notation, the GLS model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (5.7)$$

where \mathbf{y} is the vector of observed data, \mathbf{X} is a matrix of known constants, $\boldsymbol{\beta}$ is a vector of fixed but unknown parameters, and \mathbf{e} is a vector of random errors with covariance matrix $\text{Var}[\mathbf{e}] = \mathbf{V}$. Then the expectation of the observation vector is $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ and its variance is $\text{Var}[\mathbf{Y}] = \mathbf{V}$. The GLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (5.8)$$

The covariance matrix of $\hat{\boldsymbol{\beta}}$ is

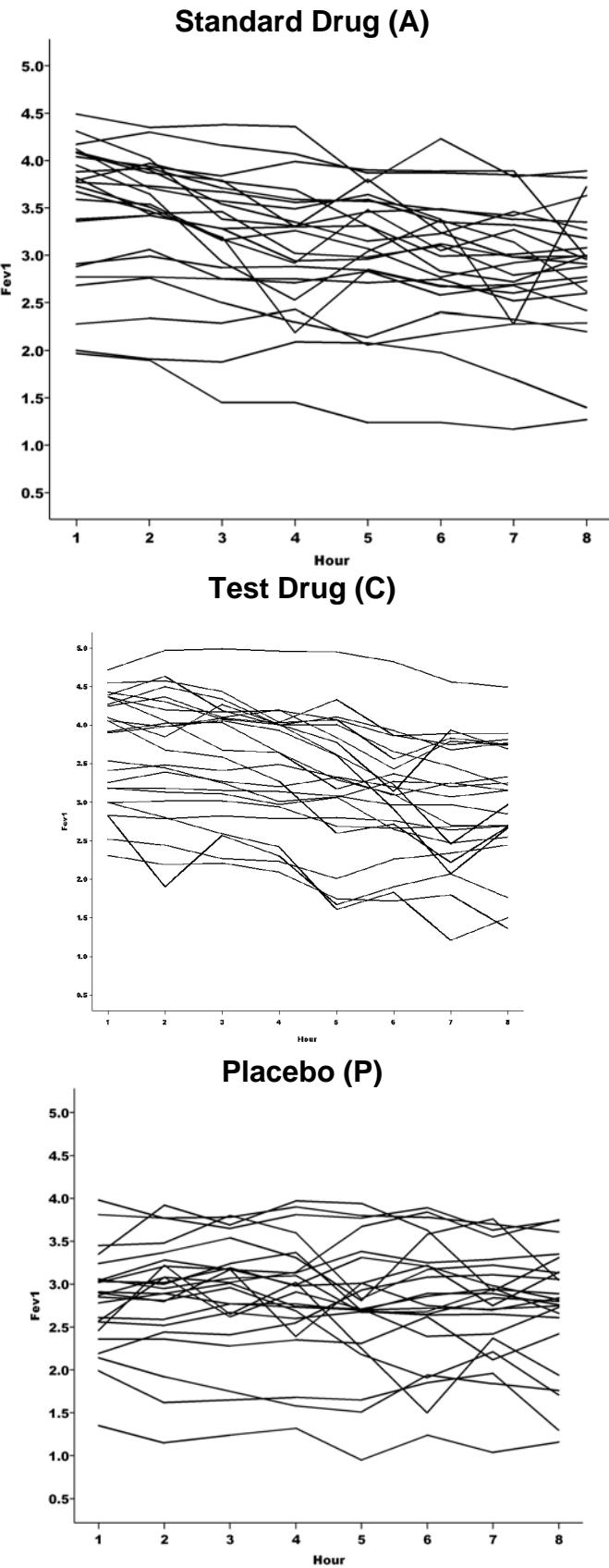
$$\text{Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \quad (5.9)$$

In the situation of this section, $\mathbf{V}^{-1} = \mathbf{I}_k \otimes \boldsymbol{\Sigma}^{-1}$, which shows more directly why it is necessary to model the form of $\boldsymbol{\Sigma}$, the covariance matrix for an individual subject.

It is apparent from equations (5.8) and (5.9) that the GLS estimator and its covariance matrix depend on the covariance matrix $\text{Var}[\mathbf{Y}] = \mathbf{V}$. This presents a technical difficulty that will be discussed in later chapters.

5.2 Example: Mixed Model Analysis of Data from Basic Repeated Measures Design

This repeated measures example is from Littell, Pendergast, and Natarajan (2000). It is also used in Littell, Stroup, and Freund (2002). The data appear as Data Set 5.2, “Respiratory Ability,” in Appendix 2, “Data Sets.” A pharmaceutical company examined effects of three drugs on respiratory ability of asthma patients. Treatments were a standard drug (A), a test drug (C), and a placebo (P). The drugs were randomly assigned to 24 patients each. The assigned treatment was administered to each patient, and a standard measure of respiratory ability called FEV1 was measured hourly for 8 hours following treatment. FEV1 was also measured immediately prior to administration of the drugs. Figure 5.1 shows profile plots for each patient.

Figure 5.1 Plots of the FEV1 Profiles by Time and Patient

The profile plots indicate higher FEV1 values at hour 1 in the active drugs A and C than in the placebo P. Response profiles in A and C appear to decline from hour 1 to hour 8, whereas the profiles in P are basically flat, except for random disturbance. Some patients have FEV1 measures that are consistently higher or lower than those of other patients, indicating the presence of a patient random effect. This means that two measures on the same patient are correlated simply because they have the patient effect in common. Also, for a given patient, consecutive measures are more highly correlated than measures several hours apart, although this is not readily apparent from the profile plots.

As noted above, there are often two aspects of covariance structure in the errors. First, two FEV1 measures on the same subject are likely to be more nearly the same than two measures on different subjects. Thus, measures on the same subject are usually positively correlated simply because they share common effects from that subject. This is the same phenomenon possessed by measures on the same whole-plot unit in a split-plot experiment. Second, two FEV1 measures made close in time on the same subject are likely to be more highly correlated than two measures made far apart in time. This feature distinguishes repeated measures covariance structure from split-plot covariance structure. In a split-plot experiment, levels of the sub-plot factor are randomized to sub-plot units within whole-plot units, resulting in equal correlation between all pairs of measures in the same whole-plot unit.

Figure 5.2 Plot of the FEV1 Means by Time for Each Drug

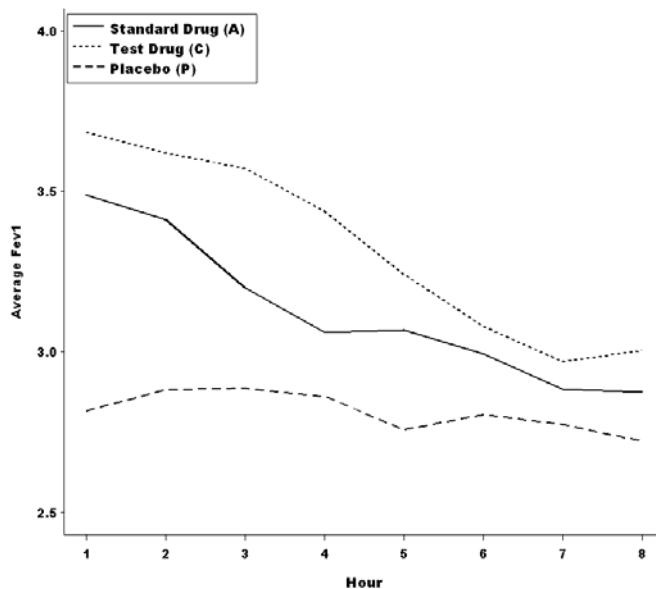


Figure 5.2 shows the FEV1 means for each drug over the eight levels of HOUR. This plot shows trends in the drug mean profiles. The means for drug P do not change much from hour 1 to hour 8. But means for drugs A and C decrease substantially, with the means for A being somewhat less than the means for C. The decreasing trend is more obvious in Figure 5.2 than in Figure 5.1 because of the difference in scale of the FEV1 axis.

5.2.1 Using the REPEATED Statement in PROC MIXED

In previous chapters we used the RANDOM statement in PROC MIXED to model covariance in data. But covariance structure of repeated measures often requires techniques more easily obtained with the REPEATED statement. In this section, we illustrate the basic aspects of the REPEATED statement. In subsequent sections of this chapter we will demonstrate several types

of covariance structures and how to implement them in PROC MIXED. Sometimes it is advantageous to use the REPEATED and RANDOM statements simultaneously.

The factors DRUG and HOUR in the FEV1 example are considered fixed. Therefore, these effects will appear in the MODEL statement. All aspects of random variation will be incorporated into a covariance structure associated with individual patients. The vector of data for each PATIENT is considered to have covariance matrix Σ . The form of Σ can be specified in the REPEATED statement.

The basic syntax of the REPEATED statement is as follows:

```
REPEATED variable name of repeated measures factor /  
    subject=combinations of variable names  
        defining sets of repeated measures  
    type=name of covariance structure;
```

Program

In this example, HOUR is the repeated measures factor. The sets of repeated measures correspond to the individual patients. In the first illustration, no particular structure is imposed; that is, the covariance is “unstructured.” The PROC MIXED statements to fit the fixed effects of DRUG, HOUR, and DRUG \times HOUR interaction, with unstructured covariance matrix Σ for each patient, are as follows:

```
proc mixed data=fevluni;  
  class drug patient hour;  
  model fev1=drug hour drug*hour;  
  repeated hour / subject=patient(drug) type=un r rcorr;  
run;
```

Except for the REPEATED statement, these statements were illustrated in earlier chapters. The REPEATED statement is used to define covariance matrix of the data vector conditional on random effects, $\text{Var}[\mathbf{Y}|\mathbf{u}] = \mathbf{R}$. Recall that in the mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, the random effects \mathbf{u} have covariance matrix \mathbf{G} and the errors \mathbf{e} have covariance matrix \mathbf{R} . Consequently, the variances of the conditional and marginal distributions are $\text{Var}[\mathbf{Y}|\mathbf{u}] = \mathbf{R}$ and $\text{Var}[\mathbf{Y}] = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, respectively. (We frequently denote the marginal variance as \mathbf{V} .) In a model without random effects ($\mathbf{u} = \mathbf{0}$), the marginal and conditional variances are identical. In other words, the PROC MIXED statements above—because they have a REPEATED but not a RANDOM statement—model the covariance matrix of the data vector \mathbf{Y} directly. In this chapter we use Σ to denote that portion of \mathbf{R} that corresponds to an individual subject.

The important elements of the REPEATED statement are as follows:

1. The effect listed before the option slash (/). Here, HOUR is listed as this effect. The levels of the effect define the rows and columns of the matrix Σ in equation (5.5).
2. The SUBJECT= effect defines the effect whose levels identify observations belonging to the same subject. In the example, all observations that share the same levels of the PATIENT(DRUG) effect represent a single subject. Observations from different subjects are independent.
3. The TYPE= option determines the covariance structure in Σ .

In addition to these three essential elements—which are almost always used—there are several options available. The R and RCORR options are used here. The R option requests that PROC MIXED display the estimate of the **R** matrix (more precisely, the Σ matrix for the first subject). The RCORR option requests the correlation matrix, which is obtained by computing correlations from the elements of the **R** matrix (a covariance matrix).

Selected results from these statements and options appear in Outputs 5.1–5.6.

Output 5.1 Mixed Model Analysis of Repeated Measures Using the REPEATED Covariance Parameter Estimates—Unstructured

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	PATIENT(DRUG)	0.4541
UN(2,1)	PATIENT(DRUG)	0.4587
UN(2,2)	PATIENT(DRUG)	0.5163
UN(3,1)	PATIENT(DRUG)	0.4441
UN(3,2)	PATIENT(DRUG)	0.4808
UN(3,3)	PATIENT(DRUG)	0.4923
UN(4,1)	PATIENT(DRUG)	0.4154
UN(4,2)	PATIENT(DRUG)	0.4688
UN(4,3)	PATIENT(DRUG)	0.4687
UN(4,4)	PATIENT(DRUG)	0.4938
UN(5,1)	PATIENT(DRUG)	0.4349
UN(5,2)	PATIENT(DRUG)	0.4943
UN(5,3)	PATIENT(DRUG)	0.4843
UN(5,4)	PATIENT(DRUG)	0.4837
UN(5,5)	PATIENT(DRUG)	0.5779
UN(6,1)	PATIENT(DRUG)	0.3934
UN(6,2)	PATIENT(DRUG)	0.4254
UN(6,3)	PATIENT(DRUG)	0.4263
UN(6,4)	PATIENT(DRUG)	0.4179
UN(6,5)	PATIENT(DRUG)	0.4945
UN(6,6)	PATIENT(DRUG)	0.4906
UN(7,1)	PATIENT(DRUG)	0.3562
UN(7,2)	PATIENT(DRUG)	0.3992
UN(7,3)	PATIENT(DRUG)	0.4021
UN(7,4)	PATIENT(DRUG)	0.4023
UN(7,5)	PATIENT(DRUG)	0.4643
UN(7,6)	PATIENT(DRUG)	0.4454
UN(7,7)	PATIENT(DRUG)	0.4994
UN(8,1)	PATIENT(DRUG)	0.3840

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(8,2)	PATIENT(DRUG)	0.4257
UN(8,3)	PATIENT(DRUG)	0.4256
UN(8,4)	PATIENT(DRUG)	0.4251
UN(8,5)	PATIENT(DRUG)	0.4950
UN(8,6)	PATIENT(DRUG)	0.4632
UN(8,7)	PATIENT(DRUG)	0.4496
UN(8,8)	PATIENT(DRUG)	0.5031

Output 5.2 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Estimated Covariance Matrix—Unstructured

Estimated R Matrix for PATIENT(DRUG) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.454	0.458	0.444	0.415	0.434	0.393	0.356	0.384
2	0.458	0.516	0.480	0.468	0.494	0.425	0.399	0.425
3	0.444	0.480	0.492	0.468	0.484	0.426	0.402	0.425
4	0.415	0.468	0.468	0.493	0.483	0.417	0.402	0.425
5	0.434	0.494	0.484	0.483	0.577	0.494	0.464	0.495
6	0.393	0.425	0.426	0.417	0.494	0.490	0.445	0.463
7	0.356	0.399	0.402	0.402	0.464	0.445	0.499	0.449
8	0.384	0.425	0.425	0.425	0.495	0.463	0.449	0.503

Interpretation

The “Covariance Parameter Estimates” table lists the estimates of the $8 \times 9/2 = 36$ covariance parameters. The “Estimated R Matrix for PATIENT(DRUG) 201 a” table lists the covariance matrix constructed from the covariance parameter estimates. In the unstructured case, the covariance matrix simply arranges the estimates above and below the main diagonal. In conventional matrix notation, this matrix is

$$\hat{\Sigma} = \begin{bmatrix} .454 & .489 & .444 & .415 & .435 & .393 & .356 & .384 \\ & .516 & .481 & .469 & .494 & .425 & .399 & .426 \\ & & .492 & .469 & .484 & .426 & .402 & .426 \\ & & & .494 & .484 & .418 & .402 & .425 \\ & & & & .578 & .495 & .464 & .495 \\ & & & & & .491 & .445 & .463 \\ & & & & & & .499 & .450 \\ & & & & & & & .503 \end{bmatrix}$$

This matrix reveals the essential feature of repeated measures data. The numbers on the main diagonal are the estimates of variances at hours 1 through 8, $\sigma_1^2, \dots, \sigma_8^2$. The estimates are similar, with all values between .454 and .578. It may be reasonable to assume that the variances at different times are equal, giving $\sigma_k^2 = \sigma^2$ for all k . The numbers off the diagonal are covariance estimates between FEV1 measures at two different hours of the same subject. For example, the number .489 is an estimate of σ_{12} , the covariance between measures at hours 1 and 2. These covariance estimates generally decrease as the length of the time interval (called the **lag**) increases. This manifests the fact that measures close in time have greater covariance than measures far apart in time. The “unstructured” covariance matrix has a large number of parameters, 36 in the case of 8 measures per patient.

The correlation matrix for the TYPE=UN covariance structure is shown in Output 5.3.

Output 5.3 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Estimated Correlation Matrix—Unstructured

Estimated R Correlation Matrix for PATIENT(DRUG) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.000	0.947	0.939	0.877	0.849	0.833	0.747	0.803
2	0.947	1.000	0.953	0.928	0.905	0.845	0.786	0.835
3	0.939	0.953	1.000	0.950	0.908	0.867	0.810	0.855
4	0.877	0.928	0.950	1.000	0.905	0.849	0.810	0.852
5	0.849	0.905	0.908	0.905	1.000	0.928	0.864	0.918
6	0.833	0.845	0.867	0.849	0.928	1.000	0.899	0.932
7	0.747	0.786	0.810	0.810	0.864	0.899	1.000	0.897
8	0.803	0.835	0.855	0.852	0.918	0.932	0.897	1.000

Interpretation

You can see the decreasing trend in correlation with increasing lag, for example, by panning along the entries for Row 1 or Col 1.

The covariance parameters are estimated using likelihood-based methods. In particular, the mixed procedure uses the REML method (Patterson and Thompson 1971) by default. This method obtains estimates of parameters by minimizing the likelihood of residuals from fitting the fixed effects portion of the model, or equivalently by minimizing the negative of the log likelihood. Goodness of fit statistics can be computed based on $-2 \text{ Res Log Likelihood}$. The values of $-2 \text{ Res Log Likelihood}$ and associated fit statistics are shown in Output 5.4.

Output 5.4 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Fit Statistics—Unstructured

Fit Statistics	
-2 Res Log Likelihood	197.5
AIC (smaller is better)	269.5
AICC (smaller is better)	274.6
BIC (smaller is better)	351.4

Interpretation

The “Fit Statistics” table displays indices of goodness of fit of the covariance structure. Akaike’s information criterion (AIC) is equal to $-2 \text{ Res Log Likelihood}$ plus twice the number of parameters in the covariance structure model (36) as a charge for estimating parameters. The AIC corrected (AICC) is a version of AIC that is adjusted for the effects of estimating parameters (Burnham and Anderson 1998) on the AIC itself. The Bayesian information criterion (BIC) is also based on $-2 \text{ Res Log Likelihood}$, but charges a heavier penalty for a large number of parameters. These will be utilized later in choosing among several covariance structures.

Output 5.5 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Null Model Likelihood Ratio Test—Unstructured

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
35	1066.49	<.0001

Interpretation

The “Null Model Likelihood Ratio Test” is a likelihood ratio test of whether the model with the specified covariance fits better than a model with errors—that is, with $\Sigma = \sigma^2 \mathbf{I}$. The p -value <.0001 shows that the $iid N(0, \sigma^2 \mathbf{I})$ model is clearly inadequate.

Output 5.6 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Tests of Fixed Effects—Unstructured

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
DRUG	2	69	3.60	0.0327
hour	7	69	13.72	<.0001
DRUG*hour	14	69	4.06	<.0001

Interpretation

The “Tests of Fixed Effects” table provides F -statistics based on GLS for testing the statistical significance of DRUG, HOUR, and DRUG \times HOUR.

5.2.2 Comparing Results from Two Covariance Structures

The ultimate objective of the statistical analysis is to make inference about the fixed effects of DRUG and HOUR, in the form of tests of hypotheses and confidence intervals for means and differences between means. But the inferential results about fixed effects depend on the specification of the covariance matrix of random effects. That is why it is important to appropriately model the covariance structure. To illustrate this point, we will run the same PROC MIXED program, except that we specify a different covariance structure, called **compound symmetry**. The essential features of a compound symmetry structure are: (1) equal variances at all times, $\text{Var}[Y_{ijk}] = \sigma^2$ for all k , and (2) equal covariance between observations on the same subject at all pairs of times, $\text{Cov}[Y_{ijk}, Y_{ijk'}] = \rho\sigma^2$.

The statements for specifying compound symmetric covariance are as follows:

```
proc mixed data=fev1uni;
  class drug patient hour;
  model fev1=drug hour drug*hour;
  repeated hour / sub=patient(drug) type=cs r rcorr;
run;
```

Partial output appears in Output 5.7 through Output 5.10.

Output 5.7 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Covariance Parameter Estimates and Estimated Covariance Matrix—Compound Symmetry

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
CS	PATIENT(DRUG)	0.4403
Residual		0.06313

Estimated R Matrix for PATIENT(DRUG) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.503	0.440	0.440	0.440	0.440	0.440	0.440	0.440
2	0.440	0.503	0.440	0.440	0.440	0.440	0.440	0.440
3	0.440	0.440	0.503	0.440	0.440	0.440	0.440	0.440
4	0.440	0.440	0.440	0.503	0.440	0.440	0.440	0.440
5	0.440	0.440	0.440	0.440	0.503	0.440	0.440	0.440
6	0.440	0.440	0.440	0.440	0.440	0.503	0.440	0.440
7	0.440	0.440	0.440	0.440	0.440	0.440	0.503	0.440
8	0.440	0.440	0.440	0.440	0.440	0.440	0.440	0.503

Interpretation

You see different results in the “Covariance Parameter Estimates” table when using TYPE=CS compared to TYPE=UN (compare Output 5.7 to Output 5.1). The number .4402, labeled “CS PATIENT(DRUG),” is the estimate of the covariance between two measures on the same patient $\text{Cov}[Y_{ijk}, Y_{ijk'}] = \rho\sigma^2$, where $\text{Var}[Y_{ijk}] = \sigma^2$. In most cases, $\rho > 0$, and $\rho\sigma^2$ is equivalent to a between-subjects variance component σ_S^2 . That is the case with this example, and we shall

refer to the parameter as σ_S^2 . The number .06313, labeled “Residual” in the output, is the estimate of the residual variance component. It is the variance of Y_{ijk} conditional on a patient, $\text{Var}[Y_{ijk}|i] = \sigma_B^2$. It follows that $\sigma^2 = \sigma_S^2 + \sigma_B^2$.

These parameters determine the covariance matrix Σ . The estimate of the matrix is

$$\hat{\Sigma} = \begin{bmatrix} .503 & .440 & .440 & .440 & .440 & .440 & .440 & .440 \\ .440 & .503 & .440 & .440 & .440 & .440 & .440 & .440 \\ .440 & .440 & .503 & .440 & .440 & .440 & .440 & .440 \\ .440 & .440 & .440 & .503 & .440 & .440 & .440 & .440 \\ .440 & .440 & .440 & .440 & .503 & .440 & .440 & .440 \\ .440 & .440 & .440 & .440 & .440 & .503 & .440 & .440 \\ .440 & .440 & .440 & .440 & .440 & .440 & .503 & .440 \\ .440 & .440 & .440 & .440 & .440 & .440 & .440 & .503 \end{bmatrix}$$

The compound symmetric covariance is highly structured with only two parameters, σ_S^2 and σ_B^2 . According to this structure, the covariance between any two measures on the same subject, $\text{Cov}[Y_{ijk}, Y_{ijk'}]$, is equal to σ_S^2 . Consequently, the correlation between any two measures on the same subject is equal to $\rho = \sigma_S^2 / (\sigma_S^2 + \sigma_B^2)$.

Output 5.8 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Fit Statistics—Compound Symmetry

Fit Statistics	
-2 Res Log Likelihood	396.6
AIC (smaller is better)	400.6
AICC (smaller is better)	400.6
BIC (smaller is better)	405.1

Interpretation

Generally, it is desirable to have a small number of parameters. But it is also important to adequately model the covariance structure. It is doubtful that the CS structure is adequate because of the decreasing trend of the covariance as a function of lag that was apparent in the unstructured covariance matrix. Inadequate modeling of the covariance structure may result in biased estimates of variances of estimates of fixed effects. The fit statistics in Output 5.8 compared to those in Output 5.4 indicate UN provides a superior fit. AIC, AICC, and BIC are all smaller for UN than CS. In addition, you can perform a likelihood ratio test based on the difference between –2 Res Log Likelihood for the two covariance structures, which is approximately distributed as chi-square with degrees of freedom equal to the difference between the numbers of parameters in UN and CS covariance structures. The difference between –2 Res Log Likelihood for the two models is $405.1 - 197.5 = 207.6$, with $36 - 2 = 34$ degrees of freedom. This is highly significant, verifying the superior fit of the UN structure compared to the CS structure.

Output 5.9 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Null Model Likelihood Ratio Test—Compound Symmetry

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	867.41	<.0001

Interpretation

This likelihood ratio test is comparing the fit of compound symmetric covariance structure with a model with independent errors—that is, with $\Sigma = \sigma^2 I$. The *p*-value <.0001 shows that the *iid* $N(0, \sigma^2 I)$ model does not fit as well as the compound symmetry model. However, this does not imply that compound symmetry is adequate.

Output 5.10 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Tests of Fixed Effects—Compound Symmetry

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
DRUG	2	69	3.60	0.0327
hour	7	483	38.86	<.0001
DRUG*hour	14	483	7.11	<.0001

CONTRAST and ESTIMATE statements can be used with the REPEATED statement in the same way you saw them used with the RANDOM statement in Chapter 4. Here are ESTIMATE statements that estimate the difference between drugs A and C at hour 1 and the difference between hours 1 and 2 in drug A:

```
estimate 'drug a-c at hour 1'
    drug 1 -1
    drug*hour 1 0 0 0 0 0 0 -1 0 0 0 0 0 0;
```

```
estimate 'hour 1-2 drug a'
    hour 1 -1 0 0 0 0 0 0
    drug*hour 1 -1 0 0 0 0 0 0;
```

Results from the two ESTIMATE statements for two PROC MIXED runs using TYPE=UN and TYPE=CS runs appear in Output 5.11 and Output 5.12, respectively.

Output 5.11 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Results of Estimate Statement—Unstructured

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
drug a-c at hour 1	-0.1962	0.1945	69	-1.01	0.3166
drug a hour 1-2	0.07667	0.04698	69	1.63	0.1073

Output 5.12 Mixed Model Analysis of Repeated Measures Using the REPEATED Statement: Results of Estimate Statement—Compound Symmetry

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
drug a-c at hour 1	-0.1962	0.2048	483	-0.96	0.3385
drug a hour 1-2	0.07667	0.07253	483	1.06	0.2910

The estimates are the same using either covariance structure. This is a result of having balanced data, so that the estimates are, in fact, simply differences between means. The estimate ‘drug a-b at hour 1’ is equal to $\bar{Y}_{a1} - \bar{Y}_{b1}$, and the estimate ‘hour 1-2 drug a’ is equal to $\bar{Y}_{a1} - \bar{Y}_{a2}$. The standard errors are different for the two different covariance structures, although they differ for different reasons.

The variance of $\bar{Y}_{a1} - \bar{Y}_{b1}$ is $2\sigma_1^2/n$, where $\sigma_1^2 = \text{Var}[Y_{ijk}]$ is the variance of an observation at hour 1. Unbiased estimates are obtained from using either UN or CS covariance structure. An assumption of homogeneous variances is reasonable for the FEV1 data.

But the variance of $\bar{Y}_{a1} - \bar{Y}_{a2}$ is $2(\sigma^2 - \sigma_{1,2})/n$, where $\sigma_{1,2} = \text{Cov}[Y_{j1}, Y_{j2}]$ is the covariance between measures at times 1 and 2 on the same subject. This is one of the covariance parameters that are fundamentally different depending on the covariance structure.

5.3 Modeling Covariance Structure

In this section we compare several candidate covariance structures for the FEV1 data, using both graphical and information criteria to compare the covariance structures. Doing so is more or less analogous to procedures often used in selection of regression models. In the regression situation one plots the data, observes the fits of candidate models to the data, and then uses a criterion of fit to assess the various models.

5.3.1 Some Candidate Covariance Structures

PROC MIXED allows you to choose from many covariance models—in other words, forms of Σ . The simplest model is the independent covariance model, where the within-subject error

correlation is zero, and hence $\Sigma = \sigma^2 \mathbf{I}$. The most complex is the unstructured covariance model, where within-subject errors for each pair of times have their own unique correlation. Thus

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1K} \\ \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2K} \\ \sigma_3^2 & \cdots & \sigma_{3K} \\ \ddots & & \vdots \\ \sigma_K^2 & & & & \end{bmatrix} \quad (5.10)$$

In some applications, the within-subject correlation is negligible. For example, in some agronomic and large-animal-nutrition trials, repeated measurements may occur at long enough intervals, such as monthly, that correlation is effectively zero relative to other variation. In such cases, the independence structure is acceptable. However, this should be checked before the data are analyzed assuming uncorrelated errors.

Correlation is present in most repeated measures data to some extent. However, correlation is usually not as complex as the unstructured model. The simplest model with correlation is compound symmetry, referred to as TYPE=CS in PROC MIXED syntax. The CS model is written

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ 1 & 1 & \rho & \cdots & \rho \\ 1 & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & & & & 1 \end{bmatrix} \quad (5.11)$$

It assumes that correlation is constant regardless of the lag between pairs of repeated measurements. Note that the split-plot model equation

$$Y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk}$$

can be expressed as

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ijk}$$

by replacing $\varepsilon_{ijk} = b_{ij} + e_{ijk}$. Assuming that the e_{ijk} distributed *iid* $N(0, \sigma^2)$ and that the b_{ij} are distributed *iid* $N(0, \sigma^2)$ induces compound symmetry covariance of the ε_{ijk} for the ij^{th} subject. More specifically, $\text{Var}[Y_{ijk}] = \sigma_B^2 + \sigma_S^2$ and $\text{Cov}[Y_{ijk}, Y_{ijk'}] = \sigma_B^2$ for $k \neq k'$, identical to the CS model with $\sigma^2 = \sigma_B^2 + \sigma_S^2$ and $\rho = \sigma_B^2 / (\sigma_S^2 + \sigma_B^2)$.

Typically, correlation between observations is a function of their lag in time: adjacent observations tend to be more highly correlated than observations farther apart in time. Several models may adequately describe such correlation. Perhaps the most commonly used is the **first-order autoregressive**, or **AR(1)**, model. For the AR(1) model,

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{K-1} \\ & 1 & \rho & \cdots & \rho^{K-2} \\ & & 1 & \cdots & \vdots \\ & & & \ddots & \rho \\ & & & & 1 \end{bmatrix} \quad (5.12)$$

The AR(1) model assumes that $e_{ijk} = \rho e_{ij,k-1} + s_{ijk}$, where $s_{ijk} \sim iid N(0, \sigma_s^2)$. It follows that $\sigma^2 = \sigma_s^2 / (1 - \rho^2)$. This helps explain why independent error models tend to underestimate within-subject variance when correlation among the errors is non-negligible.

Under the AR(1) model, correlation between adjacent within-subject errors is ρ , regardless of whether the pair of observations is the 1st and 2nd, 2nd and 3rd, or (K-1)st and Kth, whereas with the unstructured model, each pair has its own correlation. The correlation is ρ^d for any pair of errors d units apart, such as the 1st and 3rd. In general, errors d units apart have correlation ρ^d . Note that the AR(1) model requires estimates of just two parameters, σ^2 and ρ , whereas unstructured models require estimating $K+K(K-1)/2$ parameters.

The Toeplitz model is similar to the AR(1) model in the sense that pairs of within-subject errors separated by a common lag share the same correlation. However, errors d units apart have correlation ρ_d instead of ρ^d . Thus, for the Toeplitz model,

$$\Sigma = \sigma_0^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{K-1} \\ & 1 & \rho_1 & \cdots & \rho_{K-2} \\ & & 1 & \cdots & \vdots \\ & & & \ddots & \rho_1 \\ & & & & 1 \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1,K} \\ & \sigma_0^2 & \sigma_{21} & \cdots & \sigma_{2,K} \\ & & \sigma_0^2 & \cdots & \vdots \\ & & & \ddots & \sigma_{K-1,K} \\ & & & & \sigma_0^2 \end{bmatrix} \quad (5.13)$$

The Toeplitz model is less restrictive than the AR(1) model, but it requires K parameters ($\sigma_0^2, \rho_1, \dots, \rho_{K-1}$) instead of just two.

The AR(1) and Toeplitz models make sense when observations are equally spaced and the correlation structure does not change appreciably over time. A more general model that preserves the main features of these models, but allows for unequal spacing and change over time, is the first-order ante-dependence model, or ANTE(1). The model structure is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \cdots & \sigma_1\sigma_K\rho_1\rho_2\cdots\rho_{K-1} \\ \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \cdots & \sigma_2\sigma_K\rho_2\rho_3\cdots\rho_{K-1} \\ \sigma_3^2 & \cdots & \vdots & & \\ \ddots & \sigma_{K-1}\sigma_K\rho_{K-1} & & & \\ \sigma_K^2 & & & & \end{bmatrix} \quad (5.14)$$

You can see that the ANTE(1) model assumes that the variance among observations changes over time and that correlation between pairs of observations is the product of the correlations

between adjacent times between observations, so that correlation may change over time. The ANTE(1) model requires estimating $2K-1$ parameters.

Other structures can be modified to accommodate heterogeneous variances over time, including first-order autoregressive model and Toeplitz. The modified forms of these structures are denoted ARH(1) and TOEPH, respectively, in PROC MIXED. See the SAS/STAT documentation for a complete listing of the covariance structures available in PROC MIXED.

Still another covariance is derived from combining CS with AR(1). The CS component models the variation *between* subject means and the AR(1) component models the *within*-subject component, conditional on a particular subject. This structure is produced by model (5.6), which includes the random effect for subject, with AR(1) structure imposed on \mathbf{R} . Thus, it can be termed AR(1)+RE. It has the matrix form

$$\Sigma = \sigma_B^2 \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ & 1 & 1 & \cdots & 1 \\ & & 1 & \cdots & 1 \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} + \sigma_S^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{K-1} \\ & 1 & \rho & \cdots & \rho^{K-2} \\ & & 1 & \cdots & \vdots \\ & & & \ddots & \rho \\ & & & & 1 \end{bmatrix} \quad (5.15)$$

where σ_B^2 is the between-subjects variance and σ_S^2 is the variance conditional on a subject.

Section 5.3.2 presents methods for selecting an appropriate covariance model—that is, a model that adequately accounts for within-subject correlation but does not require estimating an excessive number of covariance parameters. As with all model selection activities, evaluating covariance structure should not be a purely statistical exercise. You should first rule out covariance structures that clearly make no sense in the context of a given data set. For example, AR(1) or TOEP models are generally inappropriate if the times of observation are not equally spaced either chronologically or in terms of some meaningful criterion such as biological stage of development.

5.3.2 Selecting an Appropriate Covariance Model

You need an appropriate covariance model in order to draw accurate conclusions from repeated measures data. If you ignore important correlation by using a model that is too *simple*, you risk increasing the Type I error rate and underestimating standard errors. If the model is too *complex*, you sacrifice power and efficiency. Guerin and Stroup (2000) documented the effects of various covariance modeling decisions using PROC MIXED for repeated measures data. Their work supports the idea that repeated measures analysis is robust as long as the covariance model used is approximately correct. This is the philosophy underlying this section. Inference is severely compromised by a blatantly poor choice of the covariance model.

We illustrate two types of tools you can use with PROC MIXED to help you select a covariance model. First are **graphical tools** to help visualize patterns of correlation between observations at different times. Second are **information criteria** that measure the relative fit of competing covariance models. As noted at the end of the last section, these methods work best when you first rule out covariance structures that are obviously inconsistent with the characteristics of the data you are analyzing.

5.3.2.1 Graphical Methods

You can visualize the correlation structure by plotting changes in covariance and correlation among residuals on the same subject over lag between times of observation. Estimates of correlation and covariance among residuals are easily obtained from the following PROC MIXED statements:

```
proc mixed data=fev1uni;
  class drug patient hour;
  model fev1 = drug|hour;
  repeated / type=un subject=patient(drug) sscp rcorr;
  ods output covparms = cov
        rcorr      = corr;
run;
```

The two ODS statements create new SAS data sets containing the covariances and correlations, respectively. Use these SAS statements to create a plot for visualizing an appropriate covariance model:

```
data times;
  do time1=1 to 8;
    do time2=1 to time1;
      dist=time1-time2;
      output;
    end;
  end;
run;

data covplot; merge times cov;
run;

axis1 order = (0.34 to 0.58 by 0.04)
  minor = none
  offset= (0.2in, 0.2in)
  value = (font=swiss h=2
            '0.34' '0.38' '0.42' '0.46' '0.50' '0.54' '0.58')
  label = (angle=90 f=swiss h=2
            'Covariance of Between Subj Effects');
axis2 order = (0 to 7 by 1)
  minor = none
  offset= (0.2in, 0.2in)
  value = (font=swiss h=2 )
  label = (f=swiss h=2 'Lag');
legend1 value=(font=swiss h=2 )
  label=(f=swiss h=2 'From Time')
  across=2
  mode =protect
  position=(top right inside);
symbol1 color=black interpol=join line=1  value=square;
symbol2 color=black interpol=join line=2  value=circle;
symbol3 color=black interpol=join line=20 value=triangle;
symbol4 color=black interpol=join line=3  value=plus;
symbol5 color=black interpol=join line=4  value=star;
symbol6 color=black interpol=join line=5  value=dot;
symbol7 color=black interpol=join line=6  value=_;
symbol8 color=black interpol=join line=10 value==;
```

```

proc gplot data=covplot;
  plot estimate*dist=time2 / noframe
    vaxis = axis1
    haxis = axis2
    legend = legend1;
run;

```

First, you create a data set TIMES containing the pairs of observation times and the lag between them. This data set is then merged with the covariance data set COV. A partial printout of the resulting data set COVPLOT appears in Output 5.13. The GPLOT procedure with its associated AXIS, LEGEND, and SYMBOL definitions creates a plot of covariance between pairs of repeated measures by lag, shown in Figure 5.3.

Output 5.13 Partial Data Set Containing Lags and Estimated Covariance

time1	time2	dist	CovParm	Estimate
1	1	0	UN(1,1)	0.4541
2	1	1	UN(2,1)	0.4587
2	2	0	UN(2,2)	0.5163
3	1	2	UN(3,1)	0.4441
3	2	1	UN(3,2)	0.4808
3	3	0	UN(3,3)	0.4923
.
.
.
8	6.	2	UN(8,6)	0.4632
8	7	1	UN(8,7)	0.4496
8	8	0	UN(8,8)	0.5031

The data set COVPLOT contains the time pairs and lags and the covariance information created in the ODS step of PROC MIXED. Notice that because of the way the ODS statement constructs the data set COV, the variable TIME2 is actually the observation in the pair that is taken first. In the plot shown in Figure 5.3, notice that there are seven profiles, each corresponding to the time of the first observation in the pair. Since the variable TIME2 is the first observation in the pair, this explains why you use TIME2 in the PLOT statement of PROC GPLOT.

The values plotted at lag=0 in Figure 5.3 are the variances among the observations at each of the eight times. These range from roughly 0.45 to just less than 0.60. This, and the fact that there is no trend of increasing or decreasing variance with time of observation, suggests that a covariance model with constant variance over time is probably adequate. When we discuss model-fitting criteria in Section 5.3.2.2 this will be confirmed more formally.

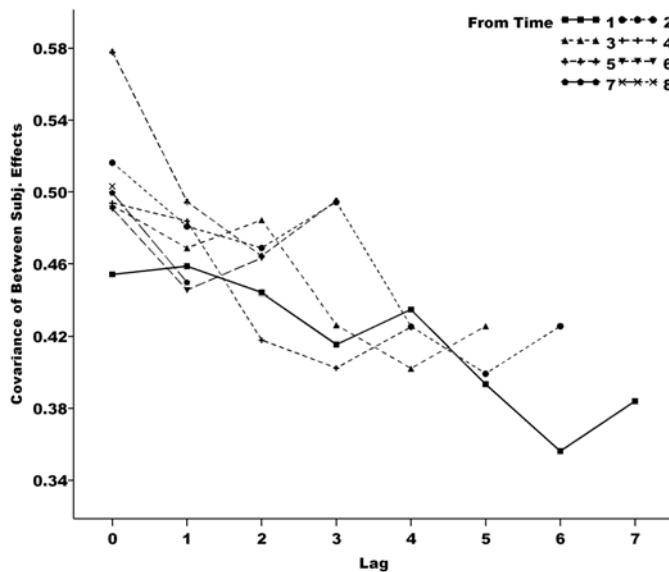
Figure 5.3 Plot of Covariance as a Function of Lag in Time between Pairs of Observations

Figure 5.3 shows that for the FEV1 data, as the lag between pairs of observations increases, covariance tends to decrease. Also, the pattern of decreasing covariance with lag is roughly the same for all reference times. Start with the profile labeled “From Time 1,” which gives the HOUR of the first observation of a given pair of repeated measures. The square symbol on the plot tracks the covariance between pairs of repeated measurements whose first observation occurs at hour 1. The position of the point at lag 0 plots the sample variance of observations taken at hour 1. The plot at lag 1 gives the covariance between hours 1 and 2, lag 2 plots covariance for hours 1 and 3, and so forth. Following the profile across the lags shows how the covariance decreases as lag increases for pairs of observations whose first time of observation is hour 1. You can see that there is a general pattern of decrease from roughly 0.50 at lag 1 to a little less than 0.40. If you follow the plots labeled “From Time” 2, 3, etc., the overall pattern is similar. The covariance among adjacent observations is consistently between 0.45 and 0.50, with the HOUR of the first element of the pair making little difference. The covariance at lag 2 is a bit lower, averaging around 0.45, and does not appear to depend on the hour of the first element of the pair.

You can draw two important conclusions from this plot:

1. An appropriate model allows covariances to decrease with increasing lag, which rules out compound symmetry. Also, an equal variance assumption (across time) is reasonable. In addition, it appears that covariance is strictly a function of lag and does not depend on the time of the first observation in the pair. Therefore, unstructured, ante-dependence, or heterogeneous variance models are probably more general than necessary; there is no visual evidence of changes in variance or covariance-lag relationships over time.

2. The between subject variance component, σ_B^2 , is nonzero. In fact, it is roughly between 0.35 and 0.40. Note that all of the plots of covariance decline for lags 1, 2, and 3, and then seem to flatten out. This is what should happen if the between-subjects variance component is approximately equal to the plotted covariance at the larger lags and there is AR(1) correlation among the observations within each subject. At the larger lags the AR(1) correlation, ρ^{lag} , should approach zero. The nonzero covariance plotted for larger lag results from the intra-class correlation $\rho = \sigma_B^2 / (\sigma_S^2 + \sigma_B^2)$.

You have seen results of using the UN and CS covariance structures in Sections 5.2.1 and 5.2.2. We now fit three other covariance structures to the FEV1 data: AR(1), AR(1)+RE, and TOEP. For completeness, we show SAS statements for all five:

```

proc mixed data=fevluni;
  class drug patient hour;
  model fev1 = drug hour drug*hour;
  repeated hour / sub=patient(drug) type=ar(1);
run;

proc mixed data=fevluni;
  class drug patient hour;
  model fev1 = drug hour drug*hour;
  repeated hour / sub=patient(drug) type=ar(1);
  random patient(drug);
run;

proc mixed data=fevluni;
  class drug patient hour;
  model fev1 = drug hour drug*hour;
  repeated hour / sub=patient(drug) type=toep;
run;

```

The covariance parameter estimates for these three covariance structures are displayed in Outputs 5.14–5.16.

Output 5.14 Covariance Parameter Estimates for AR(1) Structure

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
AR(1)	PATIENT(DRUG)	0.9219
Residual		0.4889

Interpretation

The estimates of parameters in equation (5.10) are $\hat{\sigma}^2 = .4889$ and $\hat{\rho} = .9210$. Thus, the estimated AR(1) covariance function is $(.4889)(.9219)^{lag}$. The variance estimate of .4889 (=covariance at lag=0) seems appropriate for the plot in Figure 5.3. But for greater lags, such as lag = 7, the covariance is $(.4889)(.9219)^7 = .277$, which is too small according to the plot in Figure 5.3.

Output 5.15 Covariance Parameter Estimates for AR(1)+RE Structure

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
PATIENT(DRUG)		0.4145
AR(1)	PATIENT(DRUG)	0.5420
Residual		0.08337

Interpretation

The estimates of parameters in equation (5.15) are $\hat{\sigma}_B^2 = .4145$, $\hat{\sigma}_S^2 = .0833$, and $\hat{\rho} = .542$. The estimated AR(1) covariance function is $.4145 + (.0833).542^{lag}$. Inserting $lag = 0$, you get a variance estimate of $.4145 + .0833 = .4978$. Inserting higher values of the lag, you get covariance estimates consistent with Figure 5.3.

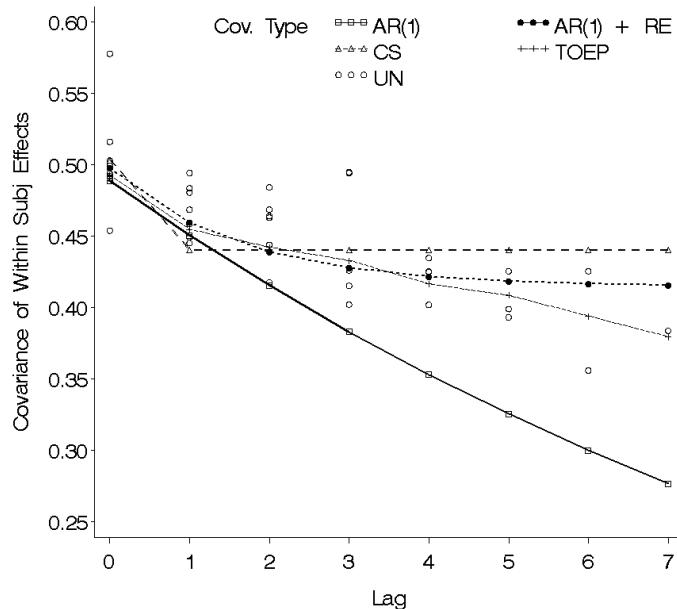
Output 5.16 Covariance Parameter Estimates for Toeplitz Structure

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
TOEP(2)	PATIENT(DRUG)	0.4549
TOEP(3)	PATIENT(DRUG)	0.4425
TOEP(4)	PATIENT(DRUG)	0.4331
TOEP(5)	PATIENT(DRUG)	0.4167
TOEP(6)	PATIENT(DRUG)	0.4087
TOEP(7)	PATIENT(DRUG)	0.3942
TOEP(8)	PATIENT(DRUG)	0.3797
Residual		0.4928

Interpretation

In the notation of equation (5.13), the estimates of variance and covariance parameters are $\hat{\sigma}_0^2 = .4549$, $\hat{\sigma}_{1,2} = .4425$, $\hat{\sigma}_{1,3} = .4331$, ..., $\hat{\sigma}_{1,8} = .3797$. The correlation estimates are, for example, $\hat{\rho}_1 = 0.4549 / 0.4928 = 0.923$. The TOEP($d+1$) parameters estimate covariance at lag d . For example, TOEP(2) estimates correlation between repeated measures $d = 1$ unit apart in time.

Figure 5.4 shows covariance plotted versus lag for five models; AR(1), AR(1)+RE, CS, TOEP, and UN. Except for UN, these covariance relationships can be expressed as a function of lag, meaning that there is only one covariance value for a given lag. Points on the graphs for the functional covariance structures are joined as curves (line segments in the case of TOEP). The unconnected points represent the covariance values from the UN model.

Figure 5.4 Plots of Covariance versus Lag for Five Covariance Models

The compound symmetry covariance is represented by the point (0, .503) followed by the flat line across all lags with covariance .440. The plot with a covariance decreasing from approximately 0.50 to 0.27 as lag increases is what one would expect to see with AR(1) covariance. The plot that decreases from approximately 0.50 to 0.42 is the expected plot from the AR(1)+RE model. You can see that the plots of observed covariance functions in Figure 5.4 correspond most closely to the AR(1)+RE and TOEP models.

5.3.2.2 Information Criteria for Comparing Covariance Structures

Output from PROC MIXED includes values under the heading “Fit Statistics.” These include -2 times the Residual (or REML) Log Likelihood (labeled “ -2 Res Log Likelihood”), and three **information criteria**. In theory, the greater the residual log likelihood, the better the fit of the model. However, somewhat analogously to R^2 in multiple regression, you can always improve the log likelihood by adding parameters to the point of absurdity.

Information criteria printed by PROC MIXED attach penalties to the log likelihood for adding parameters to the model. The penalty is a function of the number of parameters. Each of the information criteria equals -2 Res Log Likelihood *plus* -2 times a function involving the number of covariance model parameters. For example, the penalty for the compound symmetry model is a function of 2, because there are two covariance parameters, σ^2 and ρ . The penalty for an unstructured model is a function of $K(K + 1)/2$ (there are K variances and $K(K - 1)/2$ covariances). Hence, the residual log likelihood for an unstructured model is *always* greater than the residual log likelihood for compound symmetry, but the penalty is always greater as well. Unless the improvement in the residual log likelihood exceeds the size of the penalty, the simpler model yields the higher information criterion and is thus the preferred covariance model.

Two commonly used information criteria are those of Akaike (1974) and Schwarz (1978). A more recent information criterion is a finite-population corrected Akaike criterion developed by Burnham and Anderson (1998). The Akaike information criterion is often referred to by the abbreviation AIC. The finite-population corrected AIC has the abbreviation AICC. Schwarz's Bayesian information criterion is usually referenced as BIC or SBC. Hereafter, we use BIC, in concert with PROC MIXED output. Refer to SAS Help and Documentation for a complete explanation of how these terms are computed.

The basic idea for repeated measures analysis is that among the models for within-subject covariance that are considered plausible in the context of a particular study—e.g., biologically or physically reasonable—the model that *minimizes* AIC, AICC, or BIC is preferred. When AIC, AICC, or BIC is close, the simpler model is generally considered preferable in the interest of using a parsimonious model.

Keselman et al. (1998) compared AIC and BIC for their ability to select “the right” covariance model. Their study used SAS 6.12, which was determined to compute BIC incorrectly. This was corrected in SAS 8.0. Guerin and Stroup (2000) compared the information criteria using SAS 8.0 for their ability to select “the right” model *and* for the impact of choosing “the wrong” model based on the Type I error rate. They found that AIC tends to choose more complex models than BIC. They found that choosing a model that is too simple affects Type I error control more adversely than choosing a model that is too complex. When Type I error control is the highest priority, AIC is the model-fitting criterion of choice. However, if loss of power is relatively more serious, BIC may be preferable. AICC was not available at the time of the Guerin and Stroup study; a reasonable inference from their study is that its performance is similar to AIC, but somewhat less likely to choose a more complex model. Thus, loss of power is less than with AIC, but still greater than with BIC.

You can compare candidate covariance models by running PROC MIXED with the same fixed effects model, varying the RANDOM and REPEATED statements to obtain the AIC, AICC, and BIC for all candidate models. For the FEV1 data, the Toeplitz and AR(1)+RE models appear to warrant consideration based on the covariance plots.

Fit statistics for these models are produced from the statements preceding Output 5.14, and the results appear in Outputs 5.17 and 5.18.

Output 5.17 Fit Statistics for FEV1 Data Using AR(1)+RE Model

Fit Statistics	
-2 Res Log Likelihood	296.0
AIC (smaller is better)	302.0
AICC (smaller is better)	302.1
BIC (smaller is better)	308.9

Interpretation

The value of –2 Residual Log Likelihood is 296.0, and the AR(1)+RE model has three parameters, so the AIC is equal to $296 + 2(3) = 302.0$. The AICC has a small correction for estimating parameters, yielding a value of 302.1. The BIC value is 308.9, larger than AIC, but not greatly so, because of the relatively small number of parameters.

Output 5.18 Fit Statistics for FEV1 Data Using Toeplitz Model

Fit Statistics	
-2 Res Log Likelihood	277.0
AIC (smaller is better)	293.0
AICC (smaller is better)	293.3
BIC (smaller is better)	311.2

Interpretation

The value of –2 Res Log Likelihood for TOEP is 277.0, which is less than the value of –2 Res Log Likelihood for AR(1)+RE (296.0) because TOEP is a more general structure than AR(1)+RE. The TOEP model has eight parameters, so the AIC is equal to $277.0 + 2(8) = 293.0$. The AICC is slightly larger at 293.3. The BIC value is 311.2, which is larger than AIC, but again not greatly so because there are still only a modest number of parameters.

The AIC and AICC for the Toeplitz model are slightly better than the AIC and AICC for AR(1)+RE. But the BIC is slightly worse for TOEP than for AR(1)+RE. You could use either model, but the AR(1) would generally be considered preferable because it is a simpler model.

For completeness you can also fit the compound symmetry and unstructured models.

Important note about compound symmetry, Toeplitz, and unstructured models: In the AR(1)+RE model above, you use both a RANDOM statement for the between-subjects effect, PATIENT(DRUG), and a REPEATED statement for the AR(1) covariance among repeated measures within subjects. The AR(1) component accommodates covariance over and above that induced by between-subjects variation. These two sources of variation are distinct and clearly identifiable AR(1) models. However, this is not true for compound symmetry, Toeplitz, and unstructured covariance.

In Section 5.3.1 you saw that compound symmetry and the model with random between-subjects effect and independent errors are equivalent. Thus, the between-subjects variance component, σ_B^2 , and compound symmetry covariance are not identifiable. This situation also holds, in more complex form, for Toeplitz and unstructured covariance. Therefore, you should **not** use a RANDOM statement for the effect used as SUBJECT= effect in the REPEATED statement. For example, the following SAS statements are inappropriate for the compound symmetry model:

```
proc mixed;
  class drug hour patient;
  model fev1 = drug|hour basefev1;
  random patient(drug);
  repeated / type=cs subject=patient(drug);
run;
```

You should delete the RANDOM statement.

Usually, you should attempt to use separate RANDOM and REPEATED statements for structures such as AR(1), particularly when covariance-by-lag plots as in Figure 5.3 show evidence of nonzero between-subjects variation. Guerin and Stroup (2000) showed that failure

to model a separate between-subjects random effect can adversely affect inference on time and treatment \times time effects.

The AIC, AICC, and BIC results are given in the following table. Also computed are values for the AR(1) model without the RANDOM PATIENT(DRUG) statement.

Model	AIC	AICC	BIC
Compound symmetry	400.6	400.6	405.1
AR(1) with random patient(drug)	302.0	302.1	308.9
AR(1) without random patient(drug)	279.0	279.1	283.6
Toeplitz	293.0	293.3	311.2
Unstructured	269.5	274.6	351.4

You can see that the unstructured model actually has the best AIC and AICC. However, it is a distinctly poor choice according to the BIC. This illustrates the tendency of AIC, and to a lesser extent of AICC, to “choose” more complex models. In this case, given the covariance plots, choosing the unstructured model on the basis of the AIC is probably “modeling overkill,” unless there is some compelling medical or biological process that supports the unstructured model.

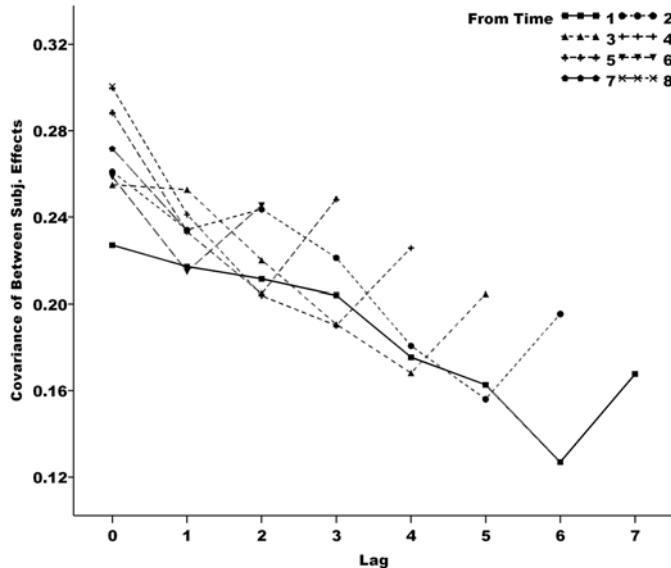
5.3.3 Reassessing the Covariance Structure with a Means Model Accounting for Baseline Measurement

In studies with repeated measures, it is common to have a pre-treatment measure, called a *baseline*. This permits using each subjects as its “own control” to assess the effect of treatment over time. Such a measure is included in the FEV1 data set, with the variable name BASEFEV1. Baseline variables are usually used as *covariates* (Milliken and Johnson 2002). You can refit the means model to include BASEFEV1 as a covariate, output the resulting covariance and correlation matrices, and plot them using the same approach used for Figure 5.3. The revised PROC MIXED statements for the baseline covariate model are as follows:

```
proc mixed data=fev1uni;
  class drug hour patient;
  model fev1 = drug|hour basefev1;
  repeated / type=un sscp subject=patient(drug) rcorr;
  ods output covparms = cov
        rcorr = corr;
run;
```

Figure 5.5 shows the plot of the covariance by lag.

Figure 5.5 Plot of Covariance as a Function of Lag in Time between Pairs of Observations



The pattern of Figure 5.5 is similar to that of Figure 5.3 except that using BASEFEV1 as a covariate substantially reduces the variance—from roughly 0.50 to less than 0.30. With increasing lag the covariance appears to reach an asymptote between 0.15 and 0.20. Again, this suggests a between-subjects variance between 0.15 and 0.20, and an additional within-subject covariance model with covariance a decreasing function of lag but independent of time.

5.3.4 PROC MIXED Analysis of FEV1 Data

Once you have selected the covariance model, you can proceed with the analysis of baseline covariate and the treatment effects just as you would in any other analysis of variance or, in this case, analysis of covariance. Start with the “Type 3 Tests of Fixed Effects” from the output you obtained when you fit the model with AR(1)+RE covariance. Output 5.19 shows the Type 3 tests.

Output 5.19 Type 3 Analysis of Covariance Tests for FEV1 Data

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BASEFEV1	1	483	75.93	<.0001
DRUG	2	68	7.28	0.0014
hour	7	483	17.10	<.0001
DRUG*hour	14	483	3.94	<.0001

You can see that the two main results are as follows:

- There is very strong evidence of a relationship between the baseline covariate BASEFEV1 and the subsequent responses FEV1. The p -value is <0.0001.
- There is strong evidence of a DRUG \times HOUR interaction ($p < 0.0001$). That is, changes in the response variable FEV1 over time are not the same for all DRUG treatments.

As with any factorial structure, inference on the DRUG and HOUR main effects should not proceed until the DRUG \times HOUR interaction is understood.

Important note about degrees of freedom, standard errors, and test statistics: Output 5.19 shows the default denominator degrees of freedom and F -values computed by PROC MIXED. The degree of freedom default is based on traditional analysis of variance assumptions, specifically an independent errors model. Denominator degrees of freedom are often substantially affected by more complex covariance structures, including those typical of repeated measures analysis. Also, PROC MIXED computes so-called naive standard errors and test statistics: it uses estimated covariance parameters in formulas that assume these quantities are known. Kackar and Harville (1984) showed that using estimated covariance parameters in this way results in test statistics that are biased upward and standard errors that are biased downward, for all cases except independent errors models with balanced data. Kenward and Roger (1997) obtained a correction for standard errors and F -statistics and a generalized procedure to obtain degrees of freedom. The Kenward-Roger (KR) correction is applicable to most covariance structures available in PROC MIXED, including all of those used in repeated measures analysis. The KR correction was added as an option with the SAS 8.0 version of PROC MIXED and is strongly recommended whenever MIXED is used for repeated measures. Guerin and Stroup (2000) compared Type I error rates for default versus KR-adjusted test statistics. Their results supported Kenward and Roger's early work: unless you use the adjustment, Type I error rates tend to be highly inflated, especially for more complex covariance structures.

The amended Type 3 statistics for fixed effects appear in Output 5.20. You obtain this output by using the option DDFM=KR in the MODEL statement. The full set of SAS statements is as follows:

```
proc mixed data=fev1uni;
  class drug hour patient;
  model fev1 = basefev1 drug|hour / ddfm=kr;
  random patient(drug);
  repeated / type=ar(1) subject=patient(drug);
run;
```

Output 5.20 Kenward-Roger Adjusted Analysis of Covariance Statistics for FEV1 Data

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BASEFEV1	1	68	76.10	<.0001
DRUG	2	68.8	7.30	0.0013
hour	7	395	16.99	<.0001
DRUG*hour	14	424	3.90	<.0001

You can see that the test statistics for HOUR and DRUG \times HOUR, the statistics that should be most affected, are somewhat lower. The degrees of freedom, especially for BASEFEV1, are also affected. The default showed BASEFEV1 with the within-subjects error denominator degrees of freedom, which makes no sense given that there is only one BASEFEV1 measurement per subject. The adjustment gives BASEFEV1 the between-subjects error degrees of freedom, which is more reasonable. The denominator degrees of freedom for HOUR and DRUG \times HOUR also receive a downward adjustment to account for the fact that an estimated covariance matrix is used. In this case, the basic conclusions of a significant BASEFEV1 covariance effect and a significant DRUG \times HOUR interaction are not changed. In many cases, however, conclusions with the KR adjustment will differ from those obtained using the default.

The next step is to explain the DRUG \times HOUR interaction. To help visualize the interaction, you can plot the DRUG \times HOUR least-squares means over time for each treatment. This is easily done with the MEANPLOT= option in the LSMEANS statement of the GLIMMIX procedure. The corresponding PROC GLIMMIX statements are as follows:

```

ods html;
ods graphics on;

proc glimmix data=fev1uni;
  class drug hour patient;
  model fev1 = drug|hour basefev1 / ddfm=kr;
  random patient(drug);
  random _residual_ / type=ar(1) subject=patient(drug);
  lsmeans drug*hour / plot=meanplot(sliceby=drug join);
  nloptions tech=nrridg;
run;

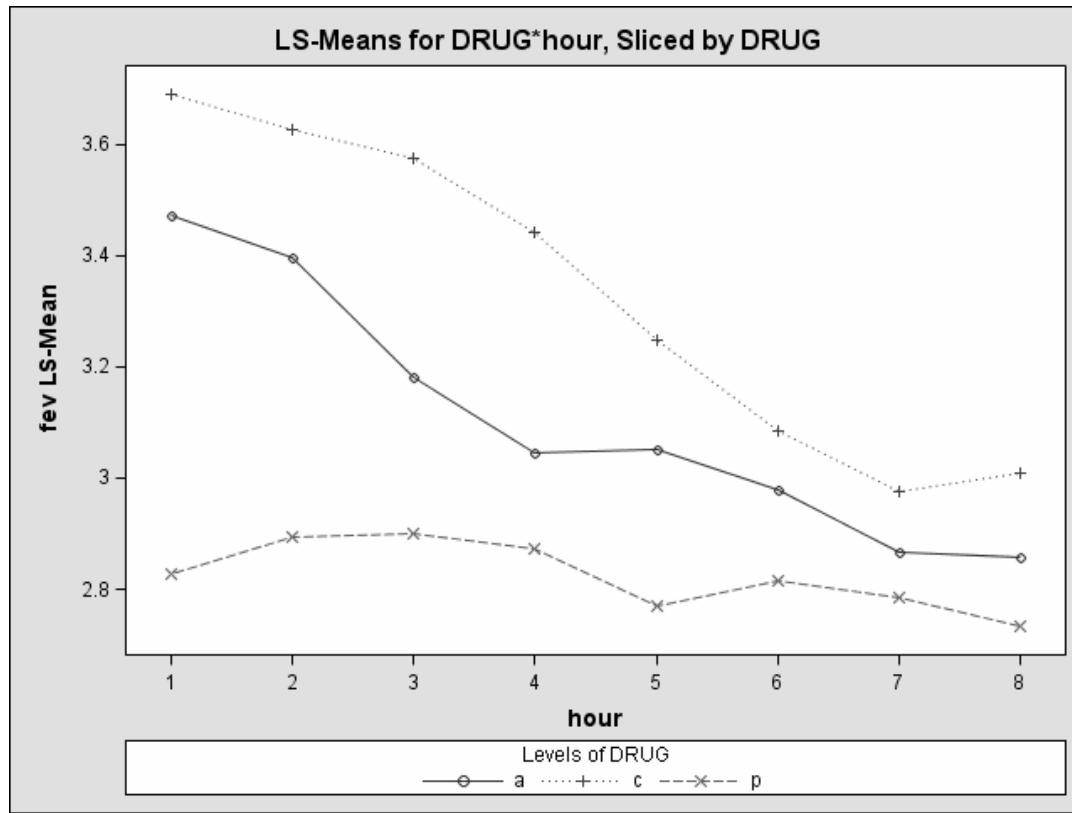
ods graphics off;
ods html close;

```

The ODS HTML statements wrapped around the PROC GLIMMIX code open and close the destination for the output. The ODS GRAPHICS statements similarly turn on and off ODS Statistical Graphics. The PROC GLIMMIX statements are very similar to the PROC MIXED statements. The REPEATED statement in PROC MIXED is replaced with the RANDOM _RESIDUAL_ statement. The NLOPTIONS statement is chosen here to set the optimization technique of PROC GLIMMIX to the same technique used by the MIXED procedure, a Newton-Raphson algorithm with ridging. The LSMEANS statement requests the least-squares means for the DRUG \times HOUR effect. The PLOT= option of the LSMEANS statement requests a graphical display of the interaction least-squares means with separate lines for each DRUG.

Figure 5.6 shows the graphical result produced by PROC GLIMMIX.

Figure 5.6 Plot of LS Means Adjusted for Baseline Covariate by Hour for Each Drug



Inspecting Figure 5.6, you can see that after adjustment for the baseline covariate, the mean responses for the two drugs, A and C, are much greater than those for the placebo, P, for the first hours of measurement. This suggests that the two drugs do improve respiratory performance relative to the placebo initially after the patient uses them. Also, for the two drugs the responses generally decrease over time, whereas for the placebo, P, there is little change. The change is approximately linear over time with a negative slope whose magnitude appears to be greater for drug C than for drug A. This suggests fitting a linear regression over HOUR, possibly with a quadratic term to account for decreasing slope as HOUR increases, and testing its interaction with HOUR. Alternatively, depending on the objectives, you might want to test DRUG differences at specific hours during the experiment. Inference along these lines continues in the next section.

5.3.6 Inference on Treatment and Time Effects of FEV1 Data Using PROC MIXED

The main task of inference on the treatment (DRUG) and time (HOUR) effects is to explain the DRUG \times HOUR interaction in a manner consistent with the objectives of the research. The section presents two approaches, one based on comparisons among the DRUG \times HOUR treatment combination means and the other based on comparing regression of the response

variable, FEV1, on HOUR for the three drug treatments. The mean comparison strategy is very similar to methods for factorial experiments presented in Chapter 3. The regression analysis is similar to procedures for comparing slopes presented in Chapter 7. Each approach is appropriate for some data sets and not for others. Which approach you use for a given data set depends on the objective. The purpose of this section is simply to show how to implement and interpret each approach.

5.3.6.1 Comparisons of DRUG × HOUR Means

Interest usually focuses on three main *types* of tests.

- Estimates or tests of simple effects, either among treatments holding time points constant or vice versa.
- SLICEs to test the effects of DRUG at a given HOUR or HOUR for a given DRUG. The former is more common because the latter is usually addressed more directly by regression methods present later in this section.
- Simple effect tests (contrasts) defined on specific aspects of the DRUG × HOUR interaction.

The following SAS statements demonstrate examples of each of these types of tests. This is not an exhaustive analysis. You can think of different tests that might be of interest. Certainly, different data sets will call for other tests. However, the methods used to construct these examples can be adapted for most tests of potential interest.

```
lsmeans drug*hour/ diff slice=hour;
contrast 'hr=1 vs hr=8 x P vs TRT'
          drug*hour 1 0 0 0 0 0 0 -1
                      1 0 0 0 0 0 0 -1
                      -2 0 0 0 0 0 0 2;
contrast 'hr=1 vs hr=8 x A vs C'
          drug*hour 1 0 0 0 0 0 0 -1
                      -1 0 0 0 0 0 0 1 0;
```

You add these statements to the PROC MIXED statements given above in this section. The SLICE=HOUR options produces tests of the DRUG effect at each time point. Often, researchers who do repeated measures experiments want to measure the change between the first and last time of measurement and want to know if this change is the same for all treatments. The CONTRAST statements perform such a comparison. The contrast “hr=1 vs hr=8 x A vs C” compares the change from the first to last time for the two drugs, A and C. The contrast “hr=1 vs hr=8 x P vs TRT” compares the first to last HOUR change in the placebo to the average of the two drugs. The SLICE and CONTRAST results are shown in Output 5.21.

Output 5.21 PROC MIXED SLICE and CONTRAST Results for FEV1 Data

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
hr=1 vs hr=8 x P vs TRT	1	160	29.78	<.0001
hr=1 vs hr=8 x A vs C	1	160	0.31	0.5773

Tests of Effect Slices					
Effect	hour	Num DF	Den DF	F Value	Pr > F
DRUG*hour	1	2	107	18.03	<.0001
DRUG*hour	2	2	107	12.57	<.0001
DRUG*hour	3	2	107	10.38	<.0001
DRUG*hour	4	2	107	7.67	0.0008
DRUG*hour	5	2	107	5.21	0.0070
DRUG*hour	6	2	107	1.64	0.1998
DRUG*hour	7	2	107	0.82	0.4427
DRUG*hour	8	2	107	1.72	0.1839

You can see that the difference between hour 1 and hour 8 is significantly different for the placebo than it is for the two drug treatments ($p < 0.0001$), but there is no evidence that the change from beginning to end is different for the two drug treatments ($p = 0.5773$). The SLICE results suggest significant differences among DRUG treatments for HOUR 1 through 5, but no statistically significant differences among drugs for HOUR 6, 7 and 8. Note that these are two degree-of-freedom comparisons. You could partition each SLICE into two single degree-of-freedom comparisons, such as A versus C and P versus treated, using contrast statements. For HOUR=1, the contrasts would be as follows:

```
contrast 'A vs C at HOUR=1' drug 1 -1 0
          drug*hour 1 0 0 0 0 0 0 0 -1 0 ;
contrast 'Placebo vs trt at HOUR=1' drug 1 1 -2
          drug*hour 1 0 0 0 0 0 0 0
          1 0 0 0 0 0 0 0 -2 0;
```

These contrast results are not shown.

The DIFF option shown in the LSMEANS statement above produces $24 \times 23/2 = 276$ treatment differences, an unmanageably large output. You can reduce the size of the output to only the simple effects of interest using ODS statements. The following is an example. Add the following statements to the PROC MIXED program:

```
ods exclude lsmeans diffss;
ods output diffss=dhdiff;
```

The ODS EXCLUDE statement prevent the LSMEANS and DIFFS from being printed when you run PROC MIXED. The SLICE results will be printed unless you list the SLICES table in the ODS EXCLUDE statement. Add the following statements to select only the simple effects and eliminate all other DIFFS:

```
data smpleff;
  set dhdiff;
  if drug=_drug or hour=_hour;
proc print data=smpleff;
run;
```

More conveniently, you can use the SLICEDIFF= option in the LSMEANS statement of the GLIMMIX procedure. It filters the comparisons involved in slices so that the levels of a given factor are held fixed. For example, in the following GLIMMIX run, the LSMEANS statement

requests all pairwise comparisons based on the coefficient vectors that define the DRUG \times HOUR least-squares means for a given level of the HOUR effect. The results are shown in Output 5.22.

```
ods select SliceDiffs;
proc glimmix data=fevluni;
  class drug hour patient;
  model fev1 = drug|hour basefev1 / ddfm=kr;
  random patient(drug);
  random _residual_ / type=ar(1) subject=patient(drug);
  lsmeans drug*hour / slicediff=(hour);
  nloptions tech=nrridg;
run;
```

Output 5.22 Simple Effect Differences from PROC GLIMMIX

Simple Effect Comparisons of DRUG*hour Least Squares Means By hour							
Simple Effect Level	DRUG	_DRUG	Estimate	Standard Error	DF	t Value	Pr > t
hour 1	a	c	-0.2182	0.1494	107.3	-1.46	0.1471
hour 1	a	p	0.6447	0.1495	107.3	4.31	<.0001
hour 1	c	p	0.8629	0.1494	107.3	5.77	<.0001
hour 2	a	c	-0.2303	0.1494	107.3	-1.54	0.1262
hour 2	a	p	0.5022	0.1495	107.3	3.36	0.0011
hour 2	c	p	0.7325	0.1494	107.3	4.90	<.0001
hour 3	a	c	-0.3941	0.1494	107.3	-2.64	0.0096
hour 3	a	p	0.2838	0.1495	107.3	1.90	0.0602
hour 3	c	p	0.6779	0.1494	107.3	4.54	<.0001
hour 4	a	c	-0.3978	0.1494	107.3	-2.66	0.0090
hour 4	a	p	0.1730	0.1495	107.3	1.16	0.2496
hour 4	c	p	0.5708	0.1494	107.3	3.82	0.0002
hour 5	a	c	-0.1966	0.1494	107.3	-1.32	0.1912
hour 5	a	p	0.2830	0.1495	107.3	1.89	0.0610
hour 5	c	p	0.4796	0.1494	107.3	3.21	0.0018
hour 6	a	c	-0.1066	0.1494	107.3	-0.71	0.4773
hour 6	a	p	0.1617	0.1495	107.3	1.08	0.2816
hour 6	c	p	0.2683	0.1494	107.3	1.80	0.0753
hour 7	a	c	-0.1091	0.1494	107.3	-0.73	0.4670
hour 7	a	p	0.08175	0.1495	107.3	0.55	0.5855
hour 7	c	p	0.1908	0.1494	107.3	1.28	0.2043
hour 8	a	c	-0.1528	0.1494	107.3	-1.02	0.3088
hour 8	a	p	0.1238	0.1495	107.3	0.83	0.4092
hour 8	c	p	0.2767	0.1494	107.3	1.85	0.0668

For example, at hours 1 and 2, there are no significant differences between drugs A and C, but there are significant differences from the placebo treatment. With increasing value of HOUR, all treatment comparisons tend to become nonsignificant. Note that this is the very meaning of “DRUG and HOUR interact”: a comparison of one factor (DRUG in Output 5.22) depends on the level of the other factor (HOUR in Output 5.22) that the comparison is performed at.

To better understand the SLICEDIFF= option, consider the coefficient matrix \mathbf{L}_1 that must be applied to compute the least-squares means for the DRUG \times HOUR effect at hour 1. This coefficient matrix has three rows that correspond to drugs A, C, and P, say, \mathbf{l}_{1a} , \mathbf{l}_{1c} , \mathbf{l}_{1p} . The F -test of a DRUG \times HOUR slice at hour 1 tests the simultaneous hypothesis $H_0: \mathbf{l}_{1a}\beta - \mathbf{l}_{1c}\beta = 0$, $\mathbf{l}_{1a}\beta - \mathbf{l}_{1p}\beta = 0$. The SLICEDIFF option in the previous GLIMMIX run tests the separate hypotheses $H: \mathbf{l}_{1a}\beta - \mathbf{l}_{1c}\beta = 0$, $H: \mathbf{l}_{1a}\beta - \mathbf{l}_{1p}\beta = 0$, $H: \mathbf{l}_{1c}\beta - \mathbf{l}_{1p}\beta = 0$.

You can compare least-squares means with even greater control by using the LSMESTIMATE statement in PROC GLIMMIX that enables you to specify general linear combinations of least-squares means. Adding the following statement to the previous run produces Output 5.22:

```
lsmestimate Drug*Hour
    1  -1,
    1   0  -1,
    1   0   0  -1,
    1   0   0   0  -1,
    1   0   0   0   0  -1,
    1   0   0   0   0   0  -1,
    1   0   0   0   0   0   0  -1 / Ftest;
```

The LSMESTIMATE statement tests seven linear combinations of the DRUG \times HOUR least-squares means. These correspond to differences from the hour 1 mean for drug A. The FTEST option in the LSMESTIMATE statement also produces a joint test for the seven estimates. This joint test equals the F -test for drug A.

Output 5.23 shows the differences between the FEV1 mean at HOUR 1 and each of the subsequent time points for DRUG A. You can see that with the AR(1) covariance structure, the standard error of a difference increases, and the degrees of freedom decrease, as the time lag increases. For DRUG A, the change from HOUR 1 to 2 is not significant, but the change from HOUR 1 to subsequent time points is.

5.3.6.2 Comparisons Using Regression

From the plot in Figure 5.7, there appears to be a linear regression of FEV1 on HOUR with a negative slope. The slope appears to be different for each drug treatment. The slopes appear to be less negative as HOUR increases, indicating a possible quadratic trend. You can test these impressions. One approach is to write orthogonal polynomial contrasts for linear, quadratic, etc. effects and, more importantly, their interactions with treatment contrasts of interest. As you will see in Chapter 7, you can do this far less tediously using analysis of covariance methods. Create a new data set with variables H and H2 defined as follows:

```
data fev1regr;
  set fev1uni;
  h=hour;
  h2=hour*hour;
run;
```

Output 5.23 Linear Combination of Least-Squares Means with PROC GLIMMIX

Least Squares Means Estimates						
Effect	Label	Estimate	Standard Error	DF	t Value	Pr > t
DRUG*hour	1	0.0767	0.0568	436.4	1.35	0.1777
DRUG*hour	2	0.2896	0.0705	456.3	4.11	<.0001
DRUG*hour	3	0.4271	0.0768	334	5.56	<.0001
DRUG*hour	4	0.4200	0.0800	247.8	5.25	<.0001
DRUG*hour	5	0.4942	0.0816	200.5	6.06	<.0001
DRUG*hour	6	0.6050	0.0824	174.4	7.34	<.0001
DRUG*hour	7	0.6154	0.0828	159.6	7.43	<.0001

Least Squares Means Ftest				
Effect	Num DF	Den DF	F Value	Pr > F
DRUG*hour	7	394.7	10.02	<.0001

The following SAS statements allow you to do these three things:

- see if there is a significant quadratic component to the regression
- see if the slopes for the linear and, if applicable, quadratic regressions are the same for drug treatments
- see if there is any lack of fit resulting from regression of higher order than quadratic

```
proc mixed data=fev1regr;
  class drug hour patient;
  model fev1=basefev1 drug h h2 hour
    drug*h drug*h2 drug*hour/htype=1;
  random patient(drug);
  repeated / type=ar(1) subject=patient(drug);
run;
```

You can see that this program is similar to analysis of covariance for qualitative \times quantitative factorial treatment structures presented in Chapter 4. Here, there is an additional covariate, BASEFEV1. When you attempt to run this program with the REPEATED covariance error structure, you often get the following warning:

```
The Mixed Procedure
WARNING: Stopped because of infinite likelihood.
```

The REML procedure cannot simultaneously compute the analysis of covariance MODEL and the correlated error structure in the REPEATED statement. You can get around this problem by using the covariance parameters already estimated from the mean comparison analysis and preventing PROC MIXED from attempting to obtain new REML estimates. Use the following statements:

```

proc mixed data=fev1regr noprofile;
  class drug hour patient;
  model fev1 = basefev1 drug h h2 hour
    drug*h drug*h2 drug*hour/htype=1;
  random patient(drug);
  repeated / type=ar(1) subject=patient(drug);
  parms (0.1848) (0.08309) (0.5401) / noiter;
run;

```

The PARMS statement gives the estimates of σ_B^2 , σ^2 , and ρ , obtained previously. You can determine the order the variance and covariance components appear in the PARMS statement by looking at the order in which they are printed in the standard MIXED analysis, e.g., Output 5.14. The NOPROFILE and NOITER commands, *used together*, cause MIXED to use these estimates literally as is. If you use NOITER without the NOPROFILE option in the PROC MIXED statement, this procedure will not work, because MIXED continues to update the residual variance that is profiled from the optimization. The NOPROFILE option includes the residual variance in the optimization, and you can fix its value. Note also that you cannot use the DDFM=KR option with this particular analysis. Output 5.24 shows the results.

Output 5.24 Type 1 Tests of Fixed Effects to Assess Lack of Fit, Order of Polynomial Regression, and Equality of Slopes for Mixed Model Analysis of FEV1 Data with AR(1) Correlated Error Structure

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BASEFEV1	1	482	76.51	<.0001
DRUG	2	68	7.28	0.0014
h	1	482	117.96	<.0001
h2	1	482	0.43	0.5143
hour	5	482	1.27	0.2756
h*DRUG	2	482	17.82	<.0001
h2*DRUG	2	482	2.20	0.1114
DRUG*hour	10	482	1.54	0.1231

Notice that the order of the “Covariance Parameter Estimates” on this analysis is different from the previous results in Output 5.15. The NOPROFILE option causes the reordering. The BASEFEV1 denominator degrees of freedom are clearly wrong. You could correct this by using the DDF option with the MODEL statement to manually set degrees of freedom. This is not done here because inference with this analysis is not concerned with BASEFEV1. The important elements are as follows:

- Neither DRUG \times HOUR nor HOUR is significant, indicating no lack of fit from the quadratic model—i.e., no third- or higher-order regression effects.
- Neither DRUG \times H2 nor H2 is statistically significant. The DRUG \times H2 term tests the equality of quadratic regressions, if any, for the treatments. Since there is no evidence of unequal slope, we test the quadratic main effect, H2, which is also not significant. Taken together, these terms indicate that there is no evidence of quadratic regression effects.

- DRUG \times H is statistically significant. There is evidence of unequal linear regressions of FEV1 on HOUR for the three drug treatments.

You can then fit separate regressions of FEV1 on HOUR for each drug using the following statements:

```
proc mixed data=fev1regr;
  class drug patient;
  model fev1=basefev1 drug h(drug) / noint solution
                                ddfm=kr htype=1;
  random patient(drug);
  repeated / type=ar(1) subject=patient(drug);
run;
```

This program is similar to examples shown in Chapter 7. Output 5.25 shows relevant results.

Output 5.25 Estimates of Parameters for Linear Regression of FEV1 on HOUR for Each Treatment Using Mixed Model Analysis with AR(1) Correlated Error Structure

Solution for Fixed Effects						
Effect	DRUG	Estimate	Standard Error	DF	t Value	Pr > t
BASEFEV1		0.8947	0.1026	68	8.72	<.0001
DRUG	a	1.1487	0.2940	72.5	3.91	0.0002
DRUG	c	1.4407	0.2917	72.5	4.94	<.0001
DRUG	p	0.5146	0.2909	72.6	1.77	0.0811
h(DRUG)	a	-0.08887	0.01164	113	-7.63	<.0001
h(DRUG)	c	-0.1057	0.01164	113	-9.08	<.0001
h(DRUG)	p	-0.01583	0.01164	113	-1.36	0.1766

You could add CONTRAST statements to the program above to see if there is evidence that regressions for drug A and C are different. You could also test the placebo versus the treated drugs, but in context this difference has already been established. The parameter estimates shown above give you the linear regression equations adjusted for BASEFEV1. For example, for drug A, the regression is $1.15 - 0.089 \times \text{HOUR}$. Consistent with Figure 5.7, the slopes for drug A and C are significantly different from 0 ($p < 0.0001$), whereas the slope for the placebo is not ($p = 0.1766$). The contrasts for comparing the drug A versus C regressions are as follows:

```
contrast 'drug a vs c intercept' drug 1 -1 0;
contrast 'drug a vs c slope'      h(drug) 1 -1 0;
```

The results are shown in Output 5.26.

Output 5.26 Test of Equality of Regression for Drug A versus Drug C

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
drug a vs c intercept	1	111	3.69	0.0572
drug a vs c slope	1	113	1.04	0.3094

There is some evidence that the intercepts are different ($p = 0.0572$), indicating a difference between drugs A and C immediately after the treatment is applied. There is no evidence of a difference in slopes ($p = 0.3094$).

5.4 Example: Unequally Spaced Repeated Measures

In the longitudinal data setting, in which repeated measures on subjects occur over time, it is often the case that the measurements are made on time intervals that are not equal. For example, consider the heart rate profiles in Figure 3.3, which are generated by the following SAS program. See Data Set 5.4, “Heart Rates,” in Appendix 2, “Data Sets,” for the complete data set.

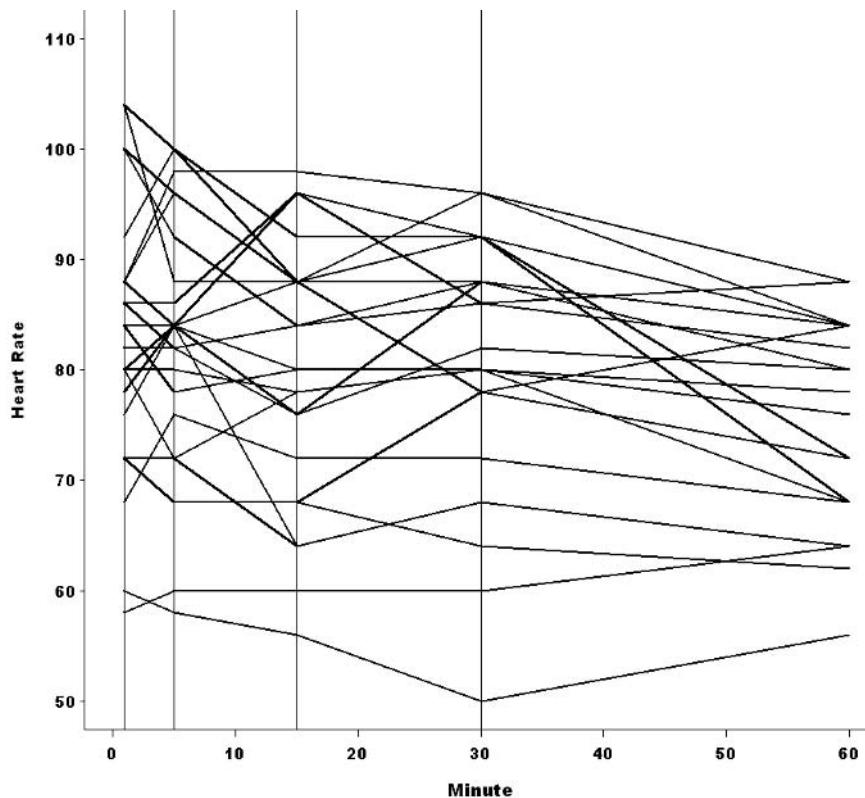
```

data hr;
  input patient drug$ basehr hr1 hr5 hr15 hr30 hr1h;
  array hra{5}  hr1 hr5 hr15 hr30 hr1h;
  do i = 1 to 5;
    if (i = 1) then minute = 1;
    else if (i = 2) then minute = 5;
    else if (i = 3) then minute = 15;
    else if (i = 4) then minute = 30;
    else minute = 60;
    hours = minute / 60;
    hours1 = hours;
    hr = hra{i} ;
    output;
  end;
  drop i hr1 hr5 hr15 hr30 hr1h;
  datalines;
...datalines...
run;

symbol i=join r=24 c=black;
proc gplot data=hr;
  plot hr*minute=patient / nolegend vminor=1 hminor=0;
run;

```

Figure 5.7 Unequally Spaced Repeated Heart Rate Measurements for 24 Patients;
Vertical Lines Drawn at Measurement Intervals



These data, from the pharmaceutical industry, consist of repeated measurements on the heart rates of 24 patients at 5 unequally spaced time intervals: 1 minute, 5 minutes, 15 minutes, 30 minutes, and 1 hour. Each patient is subjected to one of three possible drug treatment levels: a, b, and p, the last being a placebo.

Model

For data of this sort, it is sensible to consider some kind of time-series covariance structure, where the correlations of the repeated measurements are assumed to be smaller for observations that are further apart in time. However, many of the time-series covariance structures available in PROC MIXED are inappropriate because they assume equal spacing. The structures that are inappropriate include AR(1), TOEP, and ARMA(1,1). The CS and UN structures are still appropriate; however, CS assumes that the correlations remain constant, and UN is often too general. Another model that may be appropriate is the random coefficient model discussed in Chapter 7. This section focuses on appropriate time-series structures.

To fit a time-series-type covariance structure in which the correlations decline as a function of time, you can use any one of the spatial structures available in PROC MIXED. The most common of these are SP(POW) (spatial power law), SP(GAU) (Gaussian), and SP(SPH) (spherical). Chapter 11 discusses the use of these structures for spatial data; however, they are also useful for unequally spaced longitudinal measurements. The connection is that the unequally spaced data can be viewed as a spatial process in one dimension.

The SP(POW) structure for unequally spaced data provides a direct generalization of the AR(1) structure for equally spaced data. SP(POW) models the covariance between two measurements at times T1 and T2 as

$$\text{Cov}[Y_{t_1}, Y_{t_2}] = \sigma^2 \rho^{|t_1 - t_2|}$$

where ρ is an autoregressive parameter assumed to satisfy $|\rho| < 1$ and σ^2 is an overall variance.

Program

You can fit this structure to the heart rate data with the following PROC MIXED program:

```
proc mixed data=hr order=data;
  class drug hours patient;
  model hr = drug|hours basehr;
  repeated hours / type=sp(pow) (hours1) sub=patient r rcorr;
  random int / subject=patient v;
run;
```

The ORDER=DATA option is used in the PROC MIXED statement to preserve the ordering of the levels of the class variable HOUR (time is coded here in hours). The fixed effects model consists of different cell means for each drug-hour combination, and the baseline heart rate is included as a covariate.

The REPEATED statement sets up the SP(POW) structure for each patient. The REPEATED effect HOURS informs PROC MIXED of the time level of the current observation. The TYPE= option specifies the structure using the continuous variable HOURS1 to indicate the time levels. HOURS1 is an exact copy of HOURS in the HR data set; however, HOURS1 is not included in the CLASS statement, and so PROC MIXED considers it to be continuous. The SUB= option defines patients to be a blocking factor where data from different blocks are assumed to be independent. The R option requests the printout of the conditional covariance matrix corresponding to the first level of PATIENT, and RCORR prints this matrix in correlation form. The V option on the RANDOM statement request that the marginal covariance matrix be displayed. The difference between the result of the R option in the REPEATED and the result of the V option in the RANDOM statement is simply that the latter displays $\text{Var}[\mathbf{Y}] = \mathbf{ZGZ}' + \mathbf{R}$, whereas the former displays $\text{Var}[\mathbf{Y}|\mathbf{u}] = \mathbf{R}$.

Selected output from this run is shown in Output 5.27.

Results

Output 5.27 Results for Unequally Spaced Model

Class Level Information		
Class	Levels	Values
drug	3	p b a
time	5	1 5 15 30 60
patient	24	201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 232

Estimated R Matrix for patient 201					
Row	Col1	Col2	Col3	Col4	Col5
1	35.0362	20.3185	5.2039	0.6745	0.01133
2	20.3185	35.0362	8.9734	1.1631	0.01954
3	5.2039	8.9734	35.0362	4.5412	0.07629
4	0.6745	1.1631	4.5412	35.0362	0.5886
5	0.01133	0.01954	0.07629	0.5886	35.0362

Estimated R Correlation Matrix for patient 201					
Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.5799	0.1485	0.01925	0.000323
2	0.5799	1.0000	0.2561	0.03320	0.000558
3	0.1485	0.2561	1.0000	0.1296	0.002178
4	0.01925	0.03320	0.1296	1.0000	0.01680
5	0.000323	0.000558	0.002178	0.01680	1.0000

Estimated V Matrix for patient 201					
Row	Col1	Col2	Col3	Col4	Col5
1	75.1070	60.3894	45.2748	40.7454	40.0822
2	60.3894	75.1070	49.0442	41.2340	40.0904
3	45.2748	49.0442	75.1070	44.6121	40.1472
4	40.7454	41.2340	44.6121	75.1070	40.6595
5	40.0822	40.0904	40.1472	40.6595	75.1070

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	patient	40.0709
SP(POW)	patient	0.8727
Residual		35.0362

Fit Statistics	
-2 Res Log Likelihood	727.9
AIC (smaller is better)	733.9
AICC (smaller is better)	734.1
BIC (smaller is better)	737.4

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	84	1.46	0.2382
time	4	84	5.30	0.0007
drug*time	8	84	1.10	0.3699
basehr	1	84	23.22	<.0001

Interpretation

The “Class Level Information” table shows the 3 levels of DRUG, the 5 levels of HOURS (in the order that they appeared in the data), and the 24 levels of PATIENT.

The “Estimated R Matrix for patient 201” table reveals the estimate of σ^2 to be 35.03. This is not the variance of an observation however, which is $35.03 + 40.07 = 75.1$. σ^2 represents the variance among the repeated measures for a given patient. Note how the covariances are much smaller for time points that are further apart. The correlations in the “Estimated R Correlation Matrix for patient 201” table reveal the same decline. The estimated variance matrix of \mathbf{Y} is displayed in the “Estimated V Matrix for patient 201” table.

The estimate of ρ for this example is 0.8727, as shown in the “Covariance Parameter Estimates” table. The scale of this estimate depends upon the scale selected for the HOURS1 variable; however, the fixed effects estimates and their standard errors are scale invariant.

The “Fit Statistics” table displays several restricted likelihood-based statistics that can be useful in comparing covariance structures, for example, to compare TYPE=SP(POW) against TYPE=CS or TYPE=UN.

Finally, the “Type 3 Tests of Fixed Effects” table reveals that the drug had little effect, that there is evidence of change over time, and that the baseline heart rate is highly associated with subsequent heart rate measurements. All of the tests in this table are based on the SP(POW) covariance model and account for the unequally spaced measurements.

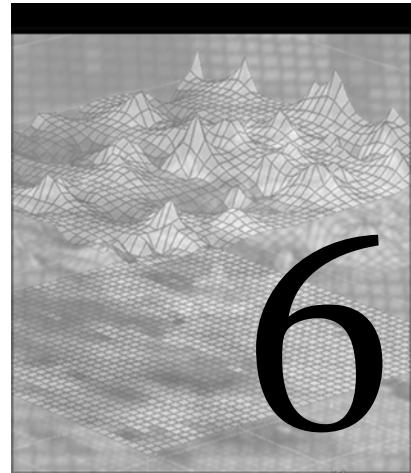
5.5 Summary

Repeated measures data need mixed models because of correlations between measurements on the same subject. Modeling the covariance structure is a preliminary step in the analysis of repeated measures data using mixed model methodology. The MIXED procedure has several covariance structures available for valid estimation of covariance in the data. An example from exercise therapy used the first-order autoregressive structure.

Missing data can have a devastating effect on multivariate methods of repeated measures analysis because subjects with incomplete data are discarded. Mixed model methodology using PROC MIXED does not require complete data. The exercise therapy example data with some values randomly deleted were used for illustration.

Autoregressive covariance estimation with the AR(1) option requires equally spaced observations. But unequally spaced data with autoregressive correlation can be analyzed using spatial covariance options.

The final two sections of this chapter considered two special types of repeated measures: those that are unequally spaced and those that are repeated in more than one dimension. The models discussed for handling the former type involve generalizations of common time-series structures such as AR(1) to handle the unequal spacing. The model for the latter type involves separate variance components for each dimension.



Best Linear Unbiased Prediction

6.1	Introduction	206
6.2	Examples of BLUP.....	206
6.2.1	Random Effects Model	206
6.2.2	Two-Way Mixed Model.....	207
6.2.3	A Random Coefficient Regression Model.....	208
6.2.4	A Mixed Model with Multiple Error Terms.....	209
6.3	Basic Concepts of BLUP.....	210
6.4	Example: Obtaining BLUPs in a Random Effects Model	212
6.4.1	Program Using the MIXED Procedure.....	213
6.4.2	Comparison of PROC MIXED and PROC GLM Results	215
6.4.3	Relationship between Sire Means and BLUPs	218
6.5	Example: Two-Factor Mixed Model	219
6.5.1	Model	219
6.5.2	Program to Obtain Estimates and Predictors.....	220
6.5.3	Intermediate Inference	222
6.5.4	Broad Space BLUP	222
6.5.5	Comparison of PROC MIXED with PROC GLM	223
6.6	A Multilocation Example.....	226
6.6.1	Model Issues for Multicenter Trial	226
6.6.2	Analysis with Fixed Location Effects.....	227
6.6.3	Analysis with Random Location Effects.....	231
6.7	Location-Specific Inference in Multicenter Example	234
6.8	Summary	241

6.1 Introduction

The first five chapters have focused primarily on fixed effects inference with mixed models. The main role of random effects in these models is in determining the standard error of an estimated fixed effect or the proper form of a test statistic. In some cases, notably in Chapter 3, estimates of the variance components are of intrinsic interest. One of the main features distinguishing mixed model methodology from conventional linear model methods is the ability to estimate “realized values of random variables”—that is, specific random effects or linear functions of random effects. For example, Henderson (1963) developed a procedure for predicting breeding values of randomly selected sires in animal genetics experiments. To do this, he used estimates of the random sire effects in a mixed model. He called his procedure **best linear unbiased prediction**, or **BLUP**. Harville (1976) showed that Henderson’s procedure had a valid theoretical basis: it can be justified as an extension of the Gauss-Markov theorem. More recently, the use of predictors based on estimated random effects has gained widespread acceptance and has been applied to statistical problems in many disciplines.

The purpose of this chapter is to introduce the fundamental concepts of best linear unbiased prediction, hereafter referred to as BLUP, and to demonstrate their use with PROC MIXED. Section 6.2 presents some introductory motivating examples. Section 6.3 discusses the basic ideas and terminology of BLUP. Sections 6.4 through 6.7 use three examples of increasing complexity to show how you can use PROC MIXED for best linear unbiased prediction.

6.2 Examples of BLUP

McLean et al. (1991) discuss various possible **inference spaces** available in working with mixed models. Robinson (1991) wrote an excellent general discussion of BLUP. This section presents a number of examples to illustrate in nontechnical terms the distinction between **estimation**—as defined in classical linear model theory—and BLUP. Section 6.3 presents a more technical discussion of BLUP theory and methods.

6.2.1 Random Effects Model

The random effects model was discussed in Chapter 3. A common example occurs in animal breeding. A group of bulls are randomly selected. Each bull is mated to a number of cows. A trait of interest (e.g., weight gain) is measured in the offspring. The model for such an experiment is

$$Y_{ijk} = \mu + s_i + d(s)_{ij} + e_{ijk} \quad (6.1)$$

where

Y_{ijk} is the measurement on the k^{th} offspring of the j^{th} dam (cow) mated to the i^{th} sire (bull)

μ is the intercept

s_i is the effect of the i^{th} sire

$d(s)_{ij}$ is the effect of the j^{th} dam mated to the i^{th} sire

e_{ijk} is the residual

The effects s_i , $d(s)_{ij}$, and e_{ijk} are each *iid* normal with mean 0 and variances σ_s^2 , σ_d^2 , and σ^2 , respectively. Additionally, all random effects are assumed to be independent of one another.

Typically, the initial purpose of such experiments is to estimate the variance associated with the random effects (e.g., sire and dam). In the animal breeding example, for instance, researchers typically want to know whether a trait can be inherited and if so, how much of the inheritance is paternal or maternal. Variance component estimates are used to estimate “heritability.”

A secondary objective of many such trials is to identify superior animals for breeding purposes. For example, if the trait of interest does turn out to be strongly inherited, say, through the sire, then the animal breeder would want to identify the best sires for future breeding. In order to do this, the sire breeding value, $\mu + s_i$, must be estimated. You can see that this is a linear combination of a fixed effect, μ , and a random effect, s_i .

The difference between fixed and random effects can be further illustrated using the nested model. Suppose each of the sires represents a particular breed—e.g., sire 1 is an Angus, sire 2 is a Hereford, etc. If there is one sire per breed, then it is reasonable to regard sire (now a.k.a. “breed”) as a fixed effect. The inclusion of specific breeds in the experiment is a reproducible decision, whereas the random sampling of sires from a population of bulls from the same breed is not.

6.2.2 Two-Way Mixed Model

The two-way mixed model was discussed in Chapter 2. McLean et al. (1991) used the following example, typical of two-way models in a manufacturing application. Different machines used in a production process are to be compared. Machine operators are randomly selected from a population of possible operators. Each operator runs each machine; two observations are taken for each operator-machine combination.

The resulting model is

$$Y_{ijk} = \mu + \tau_i + O_j + (\tau O)_{ij} + e_{ijk} \quad (6.2)$$

where

Y_{ijk} is the k^{th} observation on the i^{th} machine run by the j^{th} operator

μ is the intercept

τ_i is the effect of the i^{th} machine (fixed)

O_j is the effect of the j^{th} operator (random)

$(\tau O)_{ij}$ is the ij^{th} machine-by-operator interaction

e_{ijk} is random error

The operator, machine-by-operator, and error effects are assumed *iid* normal with mean 0 and variance components σ_O^2 , σ_{MO}^2 , and σ^2 , respectively.

The analyses discussed in Chapter 1 focus on inference about fixed effects, in this case, the machine effects. For example, you can estimate the mean performance of the i^{th} machine as $\mu + \tau_i$, or estimate or test the difference between machines, e.g., the first and second machines,

$\tau_1 - \tau_2$. These estimates and tests address the objective of assessing machine performance, say, for the purpose of choosing the best machine.

In this study, a manager may also be interested in assessing the performance of various operators under his or her supervision. The manager might want to know the performance of the j^{th} operator averaged over all machines,

$$\mu + (1/m) \sum_i \tau_i + O_j$$

where m is the number of machines in the study. The manager might also want to assess the performance of a given operator on a specific machine, $\mu + \tau_i + O_j + (\tau O)_{ij}$. These objectives must be addressed by estimating the required linear combinations of fixed and random effects, i.e., the required BLUPs.

6.2.3 A Random Coefficient Regression Model

Random coefficient regression models are discussed in detail in Chapter 8. To further illustrate the kinds of applications for which BLUP is useful, this section presents a nontechnical preview. A common objective of clinical trials is to assess the relationship between drug dosage and a physiological response. Suppose that subjects are randomly assigned to receive a given dosage of a drug. The subjects' response is then measured. A model for these data is

$$Y_{ij} = \beta_0 + s_i + (\beta_1 + d_i)X_{ij} + e_{ij} \quad (6.3)$$

where

Y_{ij} is the response of the i^{th} subject at the j^{th} dose level

β_0 is the fixed intercept

β_1 is the fixed slope

s_i is the random deviation of the i^{th} subject's intercept from β_0

d_i is the random deviation of the i^{th} subject's slope from β_1

X_{ij} is the j^{th} dose level assigned to subject i

e_{ij} is random error

The random effects, s_i , d_i , and e_{ij} are assumed iid normal with mean 0 and variance components σ_s^2 , σ_d^2 , and σ_e^2 , respectively.

In conventional regression theory, the **subject-specific** terms, s_i and d_i , do not appear in the model. Inference focuses on estimating the intercept and slope and using them to obtain predicted response to a given dose level, i.e., $\hat{\beta}_0 + \hat{\beta}_1 X_{ij}$. The estimates of intercept, slope, and predicted response are implicitly averages over the entire population of subjects.

In the mixed model above including the random regression coefficients, s_i and d_i , the same **population-average** estimates used in conventional regression can also be obtained. However, in a clinical trial where the subjects happen to be patients under treatment for an illness, the specific response of a given patient to drug dosage may also be of interest. The attending physician, for example, may want to know how an individual patient responds to a given dose.

In the random coefficient mixed model, you assume that the population-average relationship between dose and response is given by the conventional regression equation, $\beta_0 + \beta_1 X_{ij}$, but each subject's specific response to dosage can be estimated by

$$\beta_0 + s_i + (\beta_1 + d_i)X_{ij}$$

The terminology “population average” versus “subject-specific” is due to Zeger et al. (1988). The subject-specific estimate is a form of BLUP.

6.2.4 A Mixed Model with Multiple Error Terms

Mixed models with multiple error terms were introduced in Chapter 4. A common type of multior error model in which BLUP may be of interest occurs in a multilocation trial. Examples of multilocation trials include medical and pharmaceutical research, where similar experiments are often conducted at several clinics or hospitals, or technology-transfer research in agriculture, where similar experiments are often conducted at several farms. The purpose of such trials is to broaden the scope of inference, for example, to a wide variety of situations that might be encountered when an experimental treatment is adopted for practical use.

A typical multilocation trial is conducted as follows. Each of t treatments is replicated r times at each of s sites or locations. Ideally, the sites are a random sample from a target population. How well this ideal is met in practice can be a controversial matter; assume that sites are random for this discussion. The model implied by such a trial is

$$Y_{ijk} = \mu + s_i + r(s)_{ij} + \tau_k + (s\tau)_{ik} + e_{ijk} \quad (6.4)$$

where

Y_{ijk} is the observation on the j^{th} replication on the k^{th} treatment at the i^{th} site or location

s_i is the site effect

$r(s)_{ij}$ is the replication within-site effect

τ_k is the treatment effect

$(s\tau)_{ik}$ is the site-by-treatment interaction

The effects s_i , $r(s)_{ij}$, $(s\tau)_{ik}$, and e_{ijk} are each assumed *iid* normal with mean 0 and variances σ_s^2 , σ_{rs}^2 , σ_{rt}^2 , and σ^2 , respectively. The random effects are assumed independent of one another.

In conventional fixed effects analysis of multilocation trials, inference focuses on average treatment performance throughout the target population. Thus, estimates of treatment means, expressed as $\mu + \tau_k$, or differences between two treatments, $\tau_k - \tau_{k'}$, are the main objectives.

In many practical situations, you may be interested in the specific performance of treatments at a given site. For example, you may want to know the site-specific treatment mean, expressed as $\mu + s_i + \tau_k + (s\tau)_{ik}$, or a site-specific treatment difference, $\tau_k - \tau_{k'} + (s\tau)_{ik} - (s\tau)_{ik'}$. These estimates are of particular interest if you suspect that different treatments perform better under different environmental conditions, represented by the various locations in the trial.

Traditionally, site-specific inference has been approached first by determining if a significant site-by-treatment interaction exists and then by analyzing each site separately if the interaction

is present. However, this approach limits the power and precision of inference at each site. The mixed model approach using BLUP permits site-specific inference using information from the entire trial for all sites simultaneously.

6.3 Basic Concepts of BLUP

This section presents the basic concepts and terminology of best linear unbiased prediction required to follow the examples discussed in the remainder of the chapter. This section focuses on application and interpretation. Appendix 1 contains additional theoretical detail.

The basic form of a linear mixed model is

$$Y_j = \sum_i \beta_i X_{ji} + \sum_k u_k Z_{jk} + e_j$$

where

Y_j is the j^{th} observation

β_i are fixed effect parameters

X_{ji} are constants associated with the fixed effects

u_k are random effects

Z_{jk} are constants associated with the random effects

e_j is the j^{th} residual error

Specific forms of the linear models were discussed in Section 6.2 and in the previous chapters. Alternatively, you can write the mixed model in matrix form as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$.

The expected value of an observation is

$$E[\mathbf{Y}] = E[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}] = \mathbf{X}\boldsymbol{\beta}$$

since the expected values of the random effect vector \mathbf{u} and the error vector \mathbf{e} are 0. This is called the **unconditional expectation**, or the mean of \mathbf{Y} averaged over all possible \mathbf{u} . The subtlety of this quantity is important: in practical terms, the observed levels of the random effects are a random sample of a larger population. The unconditional expectation is the mean of \mathbf{Y} over the *entire population*.

The conditional expectation of \mathbf{Y} given \mathbf{u} , denoted $E[\mathbf{Y}|\mathbf{u}]$, is

$$E[\mathbf{Y}|\mathbf{u}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

In practical terms, this is the mean of \mathbf{Y} for the specific set of levels of the random effect *actually observed*.

The unconditional mean is thus a population-wide average, whereas the conditional mean is an average specific to an observed set of random effects. Because the set of observed levels of the random factors is not an *exact* duplicate of the entire population, the conditional and unconditional means are not equal, in general.

In the previous chapters, statistical inference is based on linear combinations of the fixed effects. Linear combinations of **fixed effects**, denoted $\sum_i \mathbf{K}_i \boldsymbol{\beta}_i$, are called **estimable functions** if they can be constructed from a linear combination of unconditional means of the observations. That is, if $\mathbf{K}'\boldsymbol{\beta} = \mathbf{T}'\mathbf{E}[\mathbf{Y}] = \mathbf{T}'\mathbf{X}\boldsymbol{\beta}$ for some \mathbf{T} , then it is estimable. Quantities such as regression coefficients, treatment means, treatment differences, contrasts, and simple effects in factorial experiments are all common examples of estimable functions.

Estimable functions do not depend on the random effects. The examples discussed in Section 6.2 introduced linear combinations of both the $\boldsymbol{\beta}_i$ and \mathbf{u}_j of interest in many practical situations. A generalization of the estimable function is required for such cases. Linear combinations of the fixed and random effects, $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$, can be formed from linear combinations of the conditional means. Such linear combinations are called **predictable functions**. A function $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$ is predictable if its $\mathbf{K}'\boldsymbol{\beta}$ component is estimable.

The mixed model equations, discussed in Appendix 1, provide solutions for both estimable and predictable functions. Using the mixed model equation solution for $\boldsymbol{\beta}$ in an estimable function results in the **best linear unbiased estimate** (BLUE) of $\mathbf{K}'\boldsymbol{\beta}$. For predictable functions, the solutions for $\boldsymbol{\beta}$ and \mathbf{u} provide the best linear unbiased *predictor* (BLUP) of $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$.

To summarize, linear combinations of fixed effects only are called estimable functions. The solution of the mixed model equations results in estimates, or BLUES, of $\mathbf{K}'\boldsymbol{\beta}$. Linear combinations of fixed *and* random effects are called predictable functions. Solving the mixed model equations yields *predictors*, or BLUPs, of $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$.

Estimates, or BLUES, and predictors, or BLUPs, imply different targets of statistical inference. Various terminology has been developed to describe the differences. McLean et al. (1991) discussed the *inference space* and defined **broad** versus **narrow** inference. Zeger et al. (1988) discussed *population-wide* versus *subject-specific* inference. To oversimplify, *broad* and population-wide generally refer to inference based exclusively on fixed effects and estimable functions, whereas *narrow* and subject-specific generally refer to inference based on predictable functions.

Technically, the broad/narrow and population-wide/subject-specific terminology is not interchangeable, and the distinctions are somewhat more subtle. You can better understand the distinctions with the following example. Section 6.2.2 considers a study involving machines (fixed) and operators (random). The model is

$$Y_{ijk} = \mu + \tau_i + O_j + (\tau O)_{ij} + e_{ijk}$$

where τ_i , O_j , and $(\tau O)_{ij}$ are the machine, operator, and machine-by-operator effects, respectively. The unconditional mean of an observation is $E[Y_{ijkl}] = \mu + \tau_i$. If you wanted an estimate of the average performance of the i^{th} machine over the entire population of operators—those actually observed in the study *and* those in the population but not observed—you would estimate $\mu + \tau_i$. Similarly, if you wanted to estimate the average difference between two machines over the entire population, you would estimate $\tau_i - \tau_j$. Both of these terms involve only fixed effects, and thus they are *estimates*.

Alternatively, you may want to assess the performance of a machine for the specific set of operators observed. Perhaps the machine's designer wants an estimate over the entire population, but as the manager of your company, you want to know how *your* operators, *specifically*, are doing. If so, you want to determine the conditional mean, $\mu + \tau_i + (1/J)\sum O_j +$

$(1/J)\Sigma(\tau O)_{ij}$, where J is the number of operators. In terms of a difference, you want to determine $\tau_i - \tau_{i'}$ + $(1/J)\Sigma[(\tau O)_{ij} - (\tau O)_{i'j}]$. These expressions represent what McLean et al. (1991) called the *narrow* inference space, as contrasted with *broad* inference, which uses the unconditional mean, $\mu + \tau_i$, and difference, $\tau_i - \tau_{i'}$. Narrow inference restricts attention to only the operators observed, whereas broad inference expands attention to the entire population. Both narrow inference terms, the conditional mean and difference, are *predictors* because they involve both fixed and random effects. You can see that the narrow inference terms share the same estimable functions as their broad inference analogs, but the narrow space terms have additional random effects.

Finally, suppose, as a manager, you want to use the study to evaluate an individual employee or to determine which machine is best to assign to that particular operator. In such cases, you do not want averages. You want a quantity that is specific to a given operator. To assess a given operator, averaged over all machines, you want to determine one of two possible predictable functions:

$$(1) \quad \mu + (1/I) \sum \tau_i + O_j + (1/I) \sum \sum (\tau O)_{ij}$$

$$(2) \quad \mu + (1/I) \sum \tau_i + O_j$$

where I is the number of machines. The narrow space BLUP is obtained from equation (1), whereas equation (2) yields the broad space BLUP. To assess a specific operator's performance on a given machine, you want to determine the conditional mean, $\mu + \tau_i + O_j + (\tau O)_{ij}$. To compare an individual operator's performance on two specific machines, you want to determine $\tau_i - \tau_{i'} + (\tau O)_{ij} - (\tau O)_{i'j}$. Each of these are what Zeger et al. (1988) called *subject-specific* terms. If you drop the random effects components of each subject-specific predictable function, you have its population-wide analog.

The remaining sections present three examples using PROC MIXED. They show how to set up, compute, and interpret various BLUPs.

6.4 Example: Obtaining BLUPs in a Random Effects Model

This section considers a data set based on the animal breeding example introduced in Section 6.2.1. Five sires were randomly sampled from a population. Each sire was mated to two dams. Two offspring per sire-dam combination were observed. The average daily gain (ADG) of each offspring was recorded. The data are given as Data Set 6.4, "Genetic Evaluation," in Appendix 2, "Data Sets." The model for this data set is model (6.1) in Section 6.2.1.

As noted in Section 6.2.1, animal breeders are often interested in the breeding value of the i^{th} sire, that is, the BLUP of the predictable function: $\mu + s_i$. Actually, three BLUPs for the i^{th} sire might be defined. The first is $\mu + s_i$, the population-wide or broad inference space BLUP, which predicts the performance of the i^{th} sire across the entire population of dams to which he might potentially be mated. The second is $\mu + s_i + (1/2)\sum d(s)_{ij}$, the narrow inference space BLUP assessing the performance of the sire on those dams to which he was actually mated. The third is the conditional mean, $\mu + s_i + d(s)_{ij}$, a "dam-specific" BLUP assessing the performance of a specific sire-by-dam combination.

6.4.1 Program Using the MIXED Procedure

You can obtain solutions for the model effects and the three BLUPs defined above using the following PROC MIXED statements.

Program

```
proc mixed data=genetic_evaluation;
  class sire dam;
  model adg= / ddfm=kr;
  random sire dam(sire);
  estimate 'sire 1 BLUP "broad" '
    intercept 1 | sire 1 0;
  estimate 'sire 1 BLUP "narrow" '
    intercept 2 | sire 2 0
    dam(sire) 1 1 0 0 0 0 0 0 / divisor=2;
  estimate 'sire 1 BLUP with dam 1'
    intercept 1 | sire 1 0
    dam(sire) 1 0;
run;
```

Some things to note about this program follow:

- This model has no fixed effects (other than intercept, which is included by default). Therefore, the MODEL statement serves only to identify the dependent variable (ADG). No independent variables appear in the MODEL statement.
- The MODEL statement uses the DDFM=KR option. This is important for two reasons. First, similar to discussions (e.g., of Satterthwaite's approximation) in Chapters 4 and 5, the prediction error variance of the BLUP may involve a linear combination of variance components. The degrees of freedom used to construct interval estimates or evaluate tests must take this into account. Second, the prediction error variance computed with estimated variance components tends to be biased downward. This is similar to the underestimation of standard errors with unbalanced and repeated measures data (see Chapter 5), but the effect can be even more severe. For BLUPs, the DDFM=KR option invokes a bias correction due to Prasad and Rao (1990) and Harville and Jeske (1992).
- The model effects, SIRE and DAM(SIRE), are both random and thus appear in the RANDOM statement.
- You use the ESTIMATE statement to define predictable functions. All fixed effect coefficients must appear first and then all random effect coefficients. Fixed and random effect coefficients must be separated by a vertical bar (|).
- For the BLUPs defined in this program, the only fixed effect is the intercept, μ . All other coefficients appear to the right of the vertical bar.
- As with ESTIMATE statements shown in previous chapters for fixed effects only, all coefficients for an effect following the last one specified are set to 0. For example, in the first ESTIMATE statement, "sire 1 BLUP broad," SIRE 1 0 causes the coefficients for the remaining sire effects (s_3 , s_4 , and s_5) to default to zero. DAM(SIRE) 0 sets all coefficients for dam to zero. Alternatively, you can explicitly give all the coefficients for every effect in the model for the predictable function: $\mu + s_1$. The ESTIMATE statement then appears as follows:

```
estimate 'sire 1 BLUP broad'
  intercept 1 |
  sire 1 0 0 0 0
  dam(sire) 0 0 0 0 0 0 0 0 0;
```

The second ESTIMATE statement gives all coefficients explicitly for the predictable function defining the narrow inference space BLUP, $\mu + s_1 + (\frac{1}{2})\sum d(s)_{1j}$. The third ESTIMATE statement uses defaults to define the function specific to the sire 1, dam 1 BLUP, $\mu + s_1 + d(s)_{11}$.

Results

The results are given in Output 6.1.

Output 6.1 PROC MIXED Output for Sire-Dam Random Effects Model

Covariance Parameter Estimates	
Cov Parm	Estimate
sire	0.05130
dam(sire)	0.03701
Residual	0.03870

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
sire 1 BLUP "broad"	2.2057	0.1859	8.91	11.86	<.0001
sire 1 BLUP "narrow"	2.1609	0.1003	12.1	21.55	<.0001
sire 1 BLUP with dam 1	2.1002	0.1428	16.1	14.71	<.0001

Interpretation

In Output 6.1, the REML estimates of the variance components are given under “Covariance Parameter Estimates.” The estimate of the sire variance, σ_s^2 , is 0.0513. The estimated dam variance, σ_D^2 , is 0.0370, and the estimated residual variance is 0.0387.

The “Estimates” output gives the three BLUPs. The column labeled “Estimate” gives the estimated BLUP. Technically, it is not an estimate, but a predictor. The “Std Error” is technically not a standard error, but is the square root of the estimated prediction error variance (the prediction standard error). These are fine points of mixed model jargon. For practical purposes, estimates and predictors have similar roles in inference, as do standard errors and square root prediction error variances. Unless otherwise noted, the rest of this chapter uses **standard error** to refer to the square root prediction error variance. Note that the standard errors shown here are corrected for bias using the Prasad-Rao procedure—a consequence of the DDFM=KR option.

For the three predictable functions, the results are shown in Table 6.1.

Table 6.1 ESTIMATE Statement Results

Label	Predictable Function	BLUP	Std. Error
sire 1 broad	$\mu + s_1$	2.206	0.1859
sire 1 narrow	$\mu + s_1 + (\frac{1}{2})\sum d(s)_{1j}$	2.161	0.1003
sire 1 dam1	$\mu + s_1 + d(s)_{11}$	2.100	0.1428

The BLUPs are not the same. This is typical of changes in the inference space among BLUPs. In this case, the prediction for sire 1 for the entire population of dams is higher than for the average of the dams to which he was actually mated. This relationship is specific to these data and does not necessarily hold in general. The prediction error is largest for broad inference, smallest for narrow inference, and between the two for dam-specific inference. This relationship *does* hold in general. In narrow inference, you have data on the specific sire-dam combinations you are predicting. In broad inference, your predictions apply to sire-dam combinations that are theoretically possible but have not been observed. Broad inference, therefore, is made with less precision than narrow inference: with broad inference there is more uncertainty and hence greater variability.

PROC MIXED computes the standard error by substituting the REML variance component estimates into the formula for the prediction error variance assuming known variance components. These formulae are given in Appendix 1. This is often called the **naive** method of estimating prediction error variance. Kackar and Harville (1984) showed that the naive estimate is biased downward. They suggested a correction for the bias. When the DDFM=KR option is in effect, the MIXED procedure applies the Prasad-Rao-Harville-Jeske standard error adjustment.

6.4.2 Comparison of PROC MIXED and PROC GLM Results

In the past, researchers interested in quantities such as sire BLUPs in random effects models used procedures such as PROC VARCOMP to obtain variance component estimates and then obtained sire means using PROC GLM (or PROC ANOVA if the data were balanced). The conceptual inconsistency of this practice was obvious to all: sires were treated as random to obtain variance component estimates, but they were treated as fixed to obtain means. However, the state of software development left little alternative for researchers who were not expert programmers or who lacked access to such experts. The BLUPs computed by PROC MIXED are conceptually consistent. How different are the BLUPs and the sire means PROC GLM computes?

Analysis with PROC GLM

You can obtain the “traditional” analysis using the following program:

```
proc glm data=genetic_evaluation;
  class sire dam;
  model adg=sire dam(sire);
  random sire dam(sire);
  lsmeans sire / stderr;
  lsmeans sire / e=dam(sire) stderr;
  estimate 'sire 1 BLUP "broad"' 
    intercept 1 sire 1 0;
  estimate 'sire 1 BLUP "narrow"' 
    intercept 2 sire 2 0
    dam(sire) 1 1 0 0 0 0 0 0 / divisor=2;
  estimate 'sire 1 BLUP with dam 1'
    intercept 1 sire 1 0
    dam(sire) 1 0;
run;
```

The first LSMEANS statement obtains the default standard errors, which are based on the residual error variance, σ^2 , only. The second LSMEANS statement reflects an attempt to account for the fact that dams are random.

Results

The results are given in Output 6.2.

Output 6.2 PROC GLM Output for Sire-Dam Random Effects Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1.83520000	0.20391111	5.27	0.0079
Error	10	0.38700000	0.03870000		
Corrected Total	19	2.22220000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
sire	4	1.27165000	0.31791250	8.21	0.0033
dam(sire)	5	0.56355000	0.11271000	2.91	0.0707

Parameter	Estimate	Standard Error	t Value	Pr > t
sire 1 BLUP "broad"	2.13750000	0.09836158	21.73	<.0001
sire 1 BLUP "narrow"	2.13750000	0.09836158	21.73	<.0001
sire 1 BLUP with dam 1	2.04500000	0.13910428	14.70	<.0001

Least Squares Means

		Standard Error	Pr > t
1	2.13750000	0.09836158	<.0001
2	2.24000000	0.09836158	<.0001
3	2.60250000	0.09836158	<.0001
4	2.02000000	0.09836158	<.0001
5	2.65000000	0.09836158	<.0001

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III MS for dam(sire) as an Error Term

sire	adg LSMEAN	Standard Error	Pr > t
1	2.13750000	0.16786155	<.0001
2	2.24000000	0.16786155	<.0001
3	2.60250000	0.16786155	<.0001
4	2.02000000	0.16786155	<.0001
5	2.65000000	0.16786155	<.0001

The mean for sire 1 is 2.138. The default standard error is 0.0984; the standard error accounting for random dam effects is 0.168. The estimate of the sire 1-dam 1 combination is 2.045 with a standard error of 0.139. None of these values agree with any of the BLUPs or their square root prediction error variances. What is PROC GLM actually doing? You can find out by running a PROC MIXED program to produce the same results as PROC GLM, as follows.

PROC MIXED Program to Duplicate PROC GLM Results

You duplicate the PROC GLM results with PROC MIXED by defining SIRE to be a fixed effect. The default standard error results from defining the narrow inference space BLUP for sire 1; it has the same coefficients as before, but now SIRE is fixed, whereas in the previous PROC MIXED program SIRE was random. The standard error accounting for random dam effects results from defining the broad space estimable function for sire 1. The PROC MIXED program follows.

Program

```
proc mixed data=genetic_evaluation;
  class sire dam;
  model adg=sire / ddfm=kr;
  random dam(sire);
  lsmeans sire;
  estimate 'sire 1 BLUP "broad"' 
            intercept 1 sire 1 0;
  estimate 'sire 1 BLUP "narrow"' 
            intercept 2 sire 2 0 |
            dam(sire) 1 1 0 0 0 0 0 0 0 0 / divisor=2;
  estimate 'sire 1 BLUP with dam 1'
            intercept 1 sire 1 0 |
            dam(sire) 1 0;
run;
```

Results

The results are given in Output 6.3.

Output 6.3 PROC MIXED Output for Sire-Dam Random Effects Model

Least Squares Means						
Effect	sire	Estimate	Standard Error	DF	t Value	Pr > t
sire	1	2.1375	0.1679	5	12.73	<.0001
sire	2	2.2400	0.1679	5	13.34	<.0001
sire	3	2.6025	0.1679	5	15.50	<.0001
sire	4	2.0200	0.1679	5	12.03	<.0001
sire	5	2.6500	0.1679	5	15.79	<.0001

Estimates						
Label	Estimate	Standard Error	DF	t Value	Pr > t	
sire 1 BLUP "broad"	2.1375	0.1679	5	12.73	<.0001	
sire 1 BLUP "narrow"	2.1375	0.09836	10	21.73	<.0001	
sire 1 BLUP with dam 1	2.0768	0.1415	14	14.68	<.0001	

Interpretation

The LSMEANS computed by PROC MIXED are identical to the broad space sire means. They produce identical results to the PROC GLM LSMEANS statement using the E=DAM(SIRE) option to account for random dam effects. The narrow space sire 1 mean produces the same result as the PROC GLM LSMEANS using the default standard error. The LSMEANS statement in PROC GLM, by default, assumes that all model effects are fixed. The RANDOM statement in PROC GLM has no impact on this default, since GLM does not recognize the distinction between the **X** (fixed) and **Z** (random) matrices in its computations. The ‘sire 1 BLUP with dam 1’ produces a slightly different output because it is based on random effect estimates not averaged over all levels of the fixed effect. This results in “shrinkage,” discussed in the following section.

6.4.3 Relationship between Sire Means and BLUPs

These data illustrate the relationship between the estimate of a mean and its BLUP analog. Whereas a fixed effects mean is a simple average—reflecting the assumption that the entire population of levels has been observed in the data set—the BLUP is a regression toward the overall mean based on the variance components of the model effects. Regression toward the mean is sometimes called **shrinkage estimation**.

For sire 1, the sample mean is $\bar{y}_{1\bullet} = 2.1375$. Its “broad” BLUP is $\mu + s_1$. Since μ is a fixed effect, you estimate it from the overall sample mean $\bar{y}_{\bullet\bullet} = 2.33$. The random effect, s_1 , is estimated from its conditional mean, $E[s_1 | \{y_{ijk}\}]$, where $\{y_{ijk}\}$ denotes the set of all observations. In this case, this reduces to $\hat{s}_{1,BLUP} = E[s_1 | \bar{y}_{1\bullet}]$, which is equal to

$$E[s_1] + \text{Cov}[s_1, \bar{y}_{1\bullet}] \left(\text{Var}[\bar{y}_{1\bullet}] \right)^{-1} (\bar{y}_{1\bullet} - \bar{y}_{\bullet\bullet})$$

Since $E[s_1] = 0$, this is equal to

$$0 + \sigma_s^2 \times \left[\frac{\sigma^2 + 6\sigma_D^2 + 12\sigma_s^2}{12} \right]^{-1} \times (\bar{y}_{1\bullet} - \bar{y}_{\bullet\bullet})$$

Substituting the estimates of the variance components and means yields

$$\hat{s}_1 = 0.0513 \times \frac{12}{(0.0387 + 6 \times 0.03701 + 12 \times 0.0513)} \times (2.1375 - 2.33) = -0.1243$$

Thus, the estimated sire 1 BLUP is $\hat{\mu} + \hat{s}_1 = 2.33 - 0.1243 = 2.2057$, as seen in Output 6.1.

Strictly speaking, BLUP assumes that the variance components are known. In practice, you use estimated variance components, so the resulting predictor is not a true BLUP. Instead, in mixed model vernacular, the predictor calculated using the estimated variance components is called an **EBLUP**. Henceforth in this chapter, “BLUP” refers to EBLUP.

The basic idea of a shrinkage estimator is that it *moves* the sire mean toward the overall mean. The degree of shrinkage depends on the magnitude of the variance. A large sire variance results in very little shrinkage, whereas a smaller variance results in more shrinkage toward μ .

The advantage of the shrinkage estimate is that estimated means well above or below μ are regressed toward μ consistent with the magnitude of σ_s^2 relative to the total variance of the observations. Thus, extreme means are attenuated by knowledge of the underlying variability, and the risk of misinterpreting the data is reduced.

6.5 Example: Two-Factor Mixed Model

The second example is based on the machine-operator study described in Section 6.2.2 and used to illustrate various BLUP concepts in Section 6.3. This example focuses on using PROC MIXED to obtain the BLUPs discussed in Section 6.3 and on relationships between these BLUPs and estimable functions frequently computed using PROC GLM.

Two different types of machine (variable MACHINE) were compared. Three operators (variable OPERATOR) were randomly sampled from a population. Two observations were taken on each operator for each machine. The response variable (Y) was a performance criterion of interest in the study. The data are given in Data Set 6.5, “Machine Operators,” in Appendix 2, “Data Sets.”

6.5.1 Model

The mixed model for this study is model (6.2) given in Section 6.2.2.

Section 6.3 considered three basic types of functions of the model effects of potential interest.

- **Estimable functions.** These are estimable linear combinations of fixed effects only. These correspond to broad inference for the machines in the McLean et al. (1991) terminology and population-average inference using the Zeger et al. (1988) terminology. Consider two examples:

e1 Machine 1 mean, $\mu + \tau_1$

e2 Machine difference, $\tau_1 - \tau_2$

- **Narrow inference predictable functions.** These are predictable functions that limit inference to the operators actually observed. Inference is based on conditional expectations given the observed operators. This is narrow inference as described by McLean et al. (1991). Consider two examples:

e3 Machine 1 BLUP given observed operators

$$\mu + \tau_1 + (1/3) \sum_j O_j + (1/3) \sum_j (\tau O)_{1j}$$

e4 Machine difference BLUP given observed operators

$$\tau_1 - \tau_2 + (1/3) \left[\sum_j (\tau O)_{1j} - \sum_j (\tau O)_{2j} \right]$$

- **Subject-specific predictable functions.** These are BLUPs applicable to individual operators. Consider three examples:

e5 Operator 1 BLUP, averaged over all machines

$$\mu + (1/2)\tau_1 + O_1 + (1/2)\sum_j (\tau O)_{i1}$$

e6 BLUP for operator 1 using machine 1

$$\mu + \tau_1 + O_1 + (\tau O)_{11}$$

e7 BLUP for the difference between machines specific to operator 1

$$\tau_1 - \tau_2 + (\tau O)_{11} - (\tau O)_{21}$$

6.5.2 Program to Obtain Estimates and Predictors

Program

You can obtain estimates and predictors for these functions using the following PROC MIXED statements:

```
proc mixed data=machine;
  class machine operator;
  model y=machine/ddfm=kr;
  random operator machine*operator;
  lsmeans machine;
  estimate 'BLUE - mach 1'
    intercept 1
    machine 1 0;
  estimate 'BLUE - diff'
    machine 1 -1;
  estimate 'BLUP - m 1 narrow'
    intercept 3
    machine 3 0 |
    operator 1 1 1
    machine*operator 1 1 1 0 0 0 / divisor=3;
  estimate 'BLUP - diff narrow'
    machine 3 -3 |
    machine*operator 1 1 1 -1 -1 -1 / divisor=3;
  estimate 'BLUP - oper 1'
    intercept 2
    machine 1 1 |
    operator 2 0 0
    machine*operator 1 0 0 1 0 0 / divisor=2;
  estimate 'BLUP - m 1 op 1'
    intercept 1
    machine 1 0 |
    operator 1 0 0
    machine*operator 1 0 0 0 0 0;
  estimate 'BLUP - diff op 1'
    machine 1 -1 |
    machine*operator 1 0 0 -1 0 0;
run;
```

The ESTIMATE statements labeled “BLUE” are the two broad or population-averaged estimable functions. The ESTIMATE statements labeled “BLUP...narrow” correspond to the two narrow inference space BLUPs. The final set of three ESTIMATE statements refer to the subject-specific BLUPs.

Results

The results of this program are given in Output 6.4.

Output 6.4 PROC MIXED Output for Machine-Operator Two-Way Mixed Model

Covariance Parameter Estimates	
Cov Parm	Estimate
operator	0.1073
machine*operator	0.05100
Residual	0.04852

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
BLUE – mach 1	50.9483	0.2467	2.97	206.50	<.0001
BLUE – diff	-1.0083	0.2240	2	-4.50	0.0460
BLUP - m1 narrow	50.9483	0.08993	6	566.53	<.0001
BLUP - diff nrw	-1.0083	0.1272	6	-7.93	0.0002
BLUP - oper 1	51.7366	0.1151	6.7	449.30	<.0001
BLUP - m1 op1	51.2979	0.1724	7.89	297.48	<.0001
BLUP - diff op1	-0.8773	0.2567	7.98	-3.42	0.0092
BLUP - m1 interm	50.9483	0.1584	2	321.65	<.0001
BLUP -op 1 intrm	51.6820	0.2391	2.99	216.15	<.0001

Least Squares Means						
Effect	machine	Estimate	Standard Error	DF	t Value	Pr > t
machine	1	50.9483	0.2467	2.97	206.50	<.0001
machine	2	51.9567	0.2467	2.97	210.59	<.0001

Interpretation

Major points about the output follow:

- For the broad and narrow inference space, the estimates of the machine means are the same, 50.9483. However, the standard errors for the two inference spaces are different. The standard error for the broad space is 0.2467. For the narrow BLUP the standard error is 0.0899.

The standard error in the broad inference space results from determining the variance of an estimated machine mean or difference, as derived in Chapters 2 and 4. The standard error for the narrow inference space is what you would get if you defined OPERATOR (and hence MACHINE \times OPERATOR) as fixed, rather than random.

- The same result holds for the machine difference as well. The two inference spaces yield identical estimates, -1.008, but the standard errors, 0.224 for broad inference and 0.1272 for narrow inference, are different.

Again, the broad standard error results from the same derivations of the variance of treatment differences presented in previous chapters. The narrow standard error corresponds to the standard error you would obtain if operator effects were defined as fixed.

- The subject-specific BLUPs are computed using methods discussed in Appendix 1. Unlike the broad and narrow space estimates and predictors, they do not correspond to any straightforward, easy-to-derive formula.

6.5.3 Intermediate Inference

McLean et al. (1991) discussed a third inference space, which they called **intermediate** inference. For example, an alternative to the two predictable functions for the machine 1 mean presented above is

$$\mu + \tau_1 + (1/3) \sum_j O_j$$

This predictable function is conditioned on the operator main effects but not on the operator-machine interactions. For a manager assessing a machine's performance with a particular group of operators, this makes sense if operator performance averaged over machines is predictable, but specific interactions are not.

Aside from its potential application, this function has some theoretical features that help shed light on traditional ways of doing things in PROC GLM. You can compute the intermediate BLUP in PROC MIXED using the following statement:

```
estimate 'BLUP - m 1 interm'
    intercept 3
    machine 3 0 |
    operator 1 1 1
    machine*operator 0 0 0 0 0 0 / divisor=3;
```

The result is included in Output 6.4 above.

As was the case for the broad and narrow machine 1 means, the estimate is the same, 50.9483. But the standard error is different. It is 0.158—less than the broad but greater than the narrow space standard error. You can derive this standard error by defining the machine and operator main effects as fixed, but the machine-by-operator effects as random.

6.5.4 Broad Space BLUP

You can also vary the inference space for subject-specific BLUPs. For example, a broad space predictable function for operator 1 averaged across machines is

$$\mu + (1/2) \tau_1 + O_1$$

This function assumes that you can estimate the machine effects and predict the operator 1 effect but that you do not want to restrict inference to the particular way in which operator 1 interacted with the machines in this study. Thus, this is broader inference than results from including the $(\tau O)_{i1}$ effects.

You can compute the broad space BLUP for operator 1 in PROC MIXED using the following statement:

```
estimate 'BLUP - op 1 broad'
    intercept 2
    machine 1 1 |
    operator 2 0 0
    machine*operator 0 0 0 0 0 0/divisor=2;
```

The results appear in the last row of the “Estimates” table in Output 6.4. The predictor is 51.682 with a standard error of 0.2391.

6.5.5 Comparison of PROC MIXED with PROC GLM

You can also analyze these data using PROC GLM. The estimable or predictable functions for the various inference spaces help interpret various aspects of the PROC GLM output, particular standard errors.

Program

To compute the analysis using PROC GLM, use the following statements:

```
proc glm data=machine;
  class machine operator;
  model y = machine|operator;
  random operator machine*operator/test;
  lsmeans machine operator machine*operator/stderr;
  lsmeans machine/stderr e=machine*operator;
  estimate 'diff' machine 1 -1/e;
run;
```

Results

The results appear in Output 6.5 and Output 6.6.

Output 6.5 PROC GLM Results for Machine-Operator Two-Way Model

Source	Type III Expected Mean Square
machine	Var(Error) + 2 Var(machine*operator) + Q(machine)
operator	Var(Error) + 2 Var(machine*operator) + 4 Var(operator)
machine*operator	Var(Error) + 2 Var(machine*operator)

machine	y LSMEAN	Standard Error	Pr > t
1	50.9483333	0.0899305	<.0001
2	51.9566667	0.0899305	<.0001

operator	y LSMEAN	Standard Error	Pr > t
1	51.7625000	0.1101420	<.0001
2	51.5675000	0.1101420	<.0001
3	51.0275000	0.1101420	<.0001

machine	operator	y LSMEAN	Standard Error	Pr > t
1	1	51.3550000	0.1557642	<.0001
1	2	50.8400000	0.1557642	<.0001
1	3	50.6500000	0.1557642	<.0001
2	1	52.1700000	0.1557642	<.0001
2	2	52.2950000	0.1557642	<.0001
2	3	51.4050000	0.1557642	<.0001

Interpretation

The following are pertinent to this discussion:

- The LSMEANS for machine are computed using estimable functions with the same coefficients as the narrow space BLUP for machine,

$$\mu + \tau_1 + (1/3) \sum_j O_j + (1/3) \sum_j (\tau O)_{1j}$$

The estimate of the LSMEANS for machine is the same (e.g., 50.9483 for machine 1) as that obtained for the broad, narrow, and intermediate estimates in PROC MIXED.

Output 6.6 PROC GLM Machine-Operator Output

*Least Squares Means
Standard Errors and Probabilities Calculated Using the Type III MS for machine*operator as an Error Term*

machine	y LSMEAN	Standard Error	Pr > t
1	50.9483333	0.1583947	<.0001
2	51.9566667	0.1583947	<.0001

Coefficients for Estimate diff		
		Row 1
Intercept		0
machine	1	1
machine	2	-1
operator	1	0
operator	2	0
operator	3	0

Coefficients for Estimate diff	
	Row 1
machine*operator 1 1	0.3333333333
machine*operator 1 2	0.3333333333
machine*operator 1 3	0.3333333333
machine*operator 2 1	-0.3333333333
machine*operator 2 2	-0.3333333333
machine*operator 2 3	-0.3333333333

Parameter	Estimate	Standard Error	t Value	Pr > t
diff	-1.00833333	0.12718097	-7.93	0.0002

Interpretation

- The default standard error for the LSMEANS for machine is the standard error for the **narrow space BLUP**—0.0899.
- Traditionally, you override the default error term using the optional error term E=MACHINE*OPERATOR to account for the fact that OPERATOR effects are random. The resulting standard error is 0.159. Using this option results in the **intermediate space BLUP** estimate and standard error.
- The ESTIMATE statement to assess the difference between machines in PROC GLM uses the same coefficients and yields the same results as the narrow space BLUP for machine difference. The ESTIMATE statement in PROC GLM has no option to override the default standard error.
- PROC GLM cannot compute the broad space estimates and standard errors for machine means and differences. They can be computed in PROC MIXED. This is important because in the vast majority of practical applications, the broad inference space is of primary—if not exclusive—interest.
- The least-squares means for OPERATOR and MACHINE \times OPERATOR are computed by standard linear model methods for fixed effects. They do not, in general, yield the same results as the corresponding BLUPs computed in PROC MIXED. PROC GLM cannot compute subject-specific BLUPs. For the model as defined in this example, the OPERATOR and MACHINE \times OPERATOR estimates computed by PROC GLM are inappropriate.

The ANOVA table and the F -values computed using the RANDOM statement with the TEST option in PROC GLM are correct. This is one of the few aspects of the PROC GLM analysis that *are* correct for this model.

6.6 A Multilocation Example

The summary section of Chapter 4 mentioned multilocation experiments as an important class of factorial experiments. The typical features of multilocation data are as follows:

- Treatments form one factor (often, the treatments themselves may have a factorial structure).
- Locations form another factor.
- Location \times treatment interactions and location-specific treatment simple effects are often of interest.
- Locations are typically a sample of a larger population and hence most appropriately regarded as random effects.
- Thus, inference on location \times treatment and location-specific treatment simple effects requires BLUP methodology.

This section illustrates multilocation mixed model analysis using the following example. The example is a multicenter trial to compare 4 treatments. Treatments were observed at each of 9 centers or locations. At each location, a randomized complete block design with 3 blocks was used.

The data appear as Data Set 6.6, “Multicenter Trial,” in Appendix 2, “Data Sets.” The model is equation (6.4), given in Section 6.2.4.

6.6.1 Model Issues for Multicenter Trial

Data Set 6.6 can be described by the model

$$Y_{ijk} = \mu + \tau_i + c_j + (\tau c)_{ij} + b(c)_{jk} + e_{ijk}$$

where

Y_{ijk} denotes the observation on the i^{th} treatment on the k^{th} block at the j^{th} center

μ denotes the intercept

τ_i is the effect of the i^{th} treatment

c_j is the effect of the j^{th} center

$(\tau c)_{ij}$ is the ij^{th} center \times treatment interaction effect

$b(c)_{jk}$ is the effect of the k^{th} block at the j^{th} center

e_{ijk} is random error associated with the ijk^{th} observation

The block effects are random, mutually independent, and distributed $N(0, \sigma_B^2)$, and random error effects are independent and distributed $N(0, \sigma^2)$. The remaining assumptions depend on the

specifics of the study and are considered controversial by many statisticians. There are two primary issues:

1. Should center effects be considered fixed or random?
2. Should the center \times treatment effect be included in the model?

It is not the purpose of this section to resolve this controversy. There are situations for which one model approach may clearly be appropriate and other situations for which the same model approach is clearly not appropriate. There is no “one size fits all” set of assumptions. In this section we first review the circumstances that should accompany each set of assumptions and then illustrate the analysis. Because this chapter concerns best linear unbiased prediction, most of the focus is on the random locations case, especially when location-specific treatment effects, a form of BLUP, are of interest.

First, a brief review of the circumstances that warrant the various assumptions.

Should center effects be considered fixed or random? Following the discussion in Chapters 1 and 2, the relevant question is, Do the centers observed constitute the entire population of interest, or are they a sample of a larger target population? Put another way, do center effects represent a sample whose characteristics can be plausibly described by a probability distribution? If centers do not plausibly represent a probability distribution and have been deliberately selected for specific reasons (similar to reasons you would apply to select treatment levels), then locations are fixed. Otherwise, they are random. Centers are often considered random simply because otherwise there would not be sufficient denominator degrees of freedom to test treatment effects. This “reasoning by convenience” is fallacious. It is easy to show by simulation that it is an open invitation to drastically excessive Type I error rates.

Should the center \times treatment effect be included in the model? The rationale for excluding center \times treatment from the model follows the experimental unit criterion for identifying terms that belong in the model discussed in Chapter 4. For the multicenter trial in this example, the experimental unit is block \times center \times treatment. Center \times treatment does not correspond to any physical unit in the design. However, if you exclude it from the model, you implicitly assume that treatment effects must be alike at all locations. In many, if not most, multicenter trials, this is a questionable leap of faith. The relevant question to ask is thus, Can center-specific differences in treatment effect be ruled out? If the answer to this question is “no” or “not sure,” then center \times treatment should be retained in the model.

6.6.2 Analysis with Fixed Location Effects

You use the following PROC MIXED statements to implement the basic elements of the analysis assuming the center effects are fixed. To reiterate from the discussion above, you assume center (or location) effects are fixed if they *are* the entire population being studied or if they have been deliberately selected for specific reasons. For example, locations may represent particular soil types or climatic conditions in an agronomic trial or well-defined socioeconomic or risk groups in a clinical trial. Historically, especially in pre-PROC MIXED days, this was considered the “standard” analysis. Output 6.7 shows relevant results from this analysis.

Program

```
proc mixed data=MultiCenter;
  class location block treatment;
  model response=location treatment location*treatment;
  random block(location);
  lsmeans treatment;
  lsmeans location*treatment/slice=location;
run;
```

Output 6.7 Multicenter Data Analysis with Fixed Location Effects

Covariance Parameter Estimates	
Cov Parm	Estimate
block(location)	0.005719
Residual	0.03475

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
location	8	18	22.80	<.0001
treatment	3	54	8.23	0.0001
location*treatment	24	54	1.45	0.1306

Least Squares Means							
Effect	location	treatment	Estimate	Standard Error	DF	t Value	Pr > t
treatment		1	2.9196	0.03871	54	75.41	<.0001
treatment		2	2.7304	0.03871	54	70.53	<.0001
treatment		3	2.9200	0.03871	54	75.42	<.0001
treatment		4	2.9589	0.03871	54	76.43	<.0001

Tests of Effect Slices					
Effect	location	Num DF	Den DF	F Value	Pr > F
location*treatment	A	3	54	0.54	0.6544
location*treatment	B	3	54	2.22	0.0965
location*treatment	C	3	54	0.66	0.5777
location*treatment	D	3	54	2.89	0.0439
location*treatment	E	3	54	2.85	0.0456
location*treatment	F	3	54	2.45	0.0735
location*treatment	G	3	54	3.72	0.0167

Tests of Effect Slices					
Effect	location	Num DF	Den DF	F Value	Pr > F
location*treatment	H	3	54	3.49	0.0217
location*treatment	I	3	54	0.97	0.4120

Interpretation

From the “Type 3 Tests of Fixed Effects” table you start with the test of no location \times treatment interaction. The F -value, 1.45, and p -value, 0.1306, appear to indicate that there is no evidence of an interaction—that is, no evidence of location-specific treatment effects (however, more about this later). This appears to justify assessing treatment main effects. The treatment F -value is 8.23 with $p < 0.0001$. The treatment least-squares means indicate that the mean response to treatment 2 (2.73) is less than the other treatments, which range from 2.91 to 2.96. You could add a DIFF option to the first LSMEANS statement to obtain interval estimates or tests of specific pairwise differences, or you could add CONTRAST or ESTIMATE statements to address specific objectives. These are not shown here, but are similar to examples in previous chapters.

Note that as with any factorial experiment, inference is a two-step process:

1. Location \times treatment interaction is evaluated first. If it is non-negligible, you focus on simple effects (for example, starting with the “Tests of Effect Slices” output).
2. Only if interaction is deemed negligible does one proceed to inference on the treatment main effects.

In this experiment, both treatment and location \times treatment effects are tested using the residual error term. Hence, there are 54 denominator degrees of freedom for each test.

The “Tests of Effect Slices” table suggests that one should be cautious about this approach. Three of the nine locations (A, C, and I) show no evidence of treatment effect. Two locations, B and F, show at best marginal evidence. Only four of the nine locations show statistically significant differences among the treatment means. Output 6.8 shows the location \times treatment least-squares means for these four locations (D, E, G, and H, where treatment effects are clearly significant; in the interest of space, the other locations are not shown).

Output 6.8 Location \times Treatment Least-Squares Means for Locations with Significant Treatment Effects

Least Squares Means							
Effect	location	treatment	Estimate	Standard Error	DF	t Value	Pr > t
location*treatment	D	1	2.4067	0.1161	54	20.72	<.0001
location*treatment	D	2	2.4867	0.1161	54	21.41	<.0001
location*treatment	D	3	2.6300	0.1161	54	22.64	<.0001
location*treatment	D	4	2.8233	0.1161	54	24.31	<.0001
location*treatment	E	1	2.8467	0.1161	54	24.51	<.0001
location*treatment	E	2	2.6000	0.1161	54	22.39	<.0001

Least Squares Means							
Effect	location	treatment	Estimate	Standard Error	DF	t Value	Pr > t
location*treatment	E	3	2.9733	0.1161	54	25.60	<.0001
location*treatment	E	4	2.9967	0.1161	54	25.80	<.0001
location*treatment	G	1	3.1500	0.1161	54	27.12	<.0001
location*treatment	G	2	2.6700	0.1161	54	22.99	<.0001
location*treatment	G	3	2.9200	0.1161	54	25.14	<.0001
location*treatment	G	4	2.7767	0.1161	54	23.91	<.0001
location*treatment	H	1	3.4533	0.1161	54	29.73	<.0001
location*treatment	H	2	3.0400	0.1161	54	26.17	<.0001
location*treatment	H	3	3.3967	0.1161	54	29.25	<.0001
location*treatment	H	4	3.4633	0.1161	54	29.82	<.0001

Comments on Location × Treatment LS Means

Of the four locations that show unambiguously significant treatment effects, two of them, D and G, show a pattern of treatment differences unlike the pattern among main effect means. At location D, the mean of treatment 1 is less than the mean of treatment 2 and the mean of treatment 3, while it is noticeably greater than that of treatment 2, is substantially less than the mean of treatment 4. For location G, treatments 2 and 4 have means that are nearly equal and relatively low, whereas treatments 1 and 3 have relatively high means.

There are two ways to look at the implications of the slices by location and location × treatment least-squares means. First, one can make them the focus of the analysis. Second, one can dismiss them as relatively unimportant and proceed with inference based on treatment main effects only.

The rationale for the former approach, focusing on simple effects, is as follows. The *F* test for location × treatment effect has 24 *numerator* degrees of freedom. Many statistical methods texts warn about the lack of power of tests with many numerator degrees of freedom. Snedecor and Cochran (1989), for example, suggest that when degrees of freedom for interaction are large, the interaction should not be disregarded unless the *p*-value is much greater than α -levels one would usually use—e.g., greater than 0.20. By this argument, the location × treatment *p*-value, 0.1306, should be considered evidence of non-ignorable interaction and further inference should be on location-specific simple effects.

The rationale for the latter approach, focusing on main effects, is as follows. You could argue that the location-specific effects observed in these data are just an expression of minor localized differences in the population. By this argument, simple effect inference would be regarded as “ghost-chasing” or overinterpreting the data, and thus a distraction from the “big picture”—i.e., differences among main effect means. The *p*-values shown here are unadjusted. For these data, if they were adjusted for multiplicity, for example, using Bonferroni’s procedure, the apparent significance might disappear. However, even with such adjustments, dilemmas of the sort illustrated by these data are possible. Clearly, this is a judgment call. There are data sets where this approach can be justified. Such a judgment requires experience and knowledge of the subject matter of the study. This argument could easily be rephrased, “the location-specific effects observed in these data are just an expression of localized variation in the population,” which essentially defines locations as random effects.

6.6.3 Analysis with Random Location Effects

With the random (or location) effects model, the center effects, c_j , are assumed to be mutually independent and distributed $N(0, \sigma_c^2)$ and the center \times treatment effects, (τ_{ij}) , are assumed to be mutually independent and distributed $N(0, \sigma_{TC}^2)$. The most common forms of inference are population-wide (broad)—that is, inference on treatment means, $\mu + \tau_i$ or location- (center-) specific. The center-specific case is equivalent to subject-specific inference discussed in previous sections of this chapter. Interest would typically focus on treatment effects at specific centers or sets of centers.

A special case of BLUP, called “narrow inference” in this discussion, produces results similar to the fixed center-effects model. In multilocation trials, you ordinarily would not be interested in narrow inference unless you consider locations fixed. However, it is of interest for the sake of comparison; this section presents an example.

6.6.3.1 SAS Program: Basic Analysis and Population-Averaged Inference

You can use the following SAS statements to compute the basis analysis—that is, estimates of variance components, overall tests of treatment effects, and estimates of and comparisons among treatment LS-means.

Program

```
proc mixed data=MultiCenter;
  class location block treatment;
  model response=treatment / ddfm=KR;
  random location block(location) location*treatment;
  lsmeans treatment / diff;
run;
```

Results

Output 6.9 shows relevant results.

Output 6.9 Random Location Effect Analysis of Multicenter Data: Broad Inference

Covariance Parameter Estimates	
Cov Parm	Estimate
location	0.1034
block(location)	0.005719
location*treatment	0.005170
Residual	0.03475

Least Squares Means						
Effect	treatment	Estimate	Standard Error	DF	t Value	Pr > t
treatment	1	2.9196	0.1164	9.9	25.07	<.0001
treatment	2	2.7304	0.1164	9.9	23.45	<.0001
treatment	3	2.9200	0.1164	9.9	25.08	<.0001
treatment	4	2.9589	0.1164	9.9	25.41	<.0001

Differences of Least Squares Means							
Effect	treatment	_treatment	Estimate	Standard Error	DF	t Value	Pr > t
treatment	1	2	0.1893	0.06102	24	3.10	0.0049
treatment	1	3	-0.00037	0.06102	24	-0.01	0.9952
treatment	1	4	-0.03926	0.06102	24	-0.64	0.5260
treatment	2	3	-0.1896	0.06102	24	-3.11	0.0048
treatment	2	4	-0.2285	0.06102	24	-3.75	0.0010
treatment	3	4	-0.03889	0.06102	24	-0.64	0.5299

Interpretation

The block(location) and error variance component estimates are identical to those obtained with fixed location effect analysis ($\hat{\sigma}_B^2 = 0.0057$, $\hat{\sigma}^2 = 0.03475$). The location and location \times treatment variance components are $\hat{\sigma}_C^2 = 0.1034$ and $\hat{\sigma}_{TC}^2 = 0.0517$, respectively. The treatment means are identical to those in the fixed location analysis, but the standard errors are not. For fixed location, the standard error of a mean is

$$\sqrt{\frac{\hat{\sigma}^2 + \hat{\sigma}_B^2}{36}} = 0.03871$$

whereas for random locations the standard error is

$$\sqrt{\frac{\hat{\sigma}^2 + \hat{\sigma}_B^2 + 4\hat{\sigma}_{TC}^2 + 4\hat{\sigma}_C^2}{36}} = 0.1164$$

Similarly, with random location effects, the standard error of a treatment difference is

$$\sqrt{\frac{2(\hat{\sigma}^2 + 4\hat{\sigma}_{TC}^2)}{36}} = 0.061$$

whereas with the fixed location analysis the standard error would be based on $\hat{\sigma}^2$ only. The standard errors reflect differences in the sources of random variations assumed to affect treatment effects in the two models. The interval estimates you obtain from the random location analysis are assumed to apply to the entire population of centers, of which the locations in the data set are only a sample. On the other hand, the interval estimates for treatment means obtained from fixed location analysis assume that the locations in the data set *are* the population. From a random location perspective, they are similar to “narrow” inference space treatment mean BLUPs.

6.6.3.2 Narrow Inference Space Treatment Means

The narrow inference space treatment mean BLUPs are defined using the same estimable function as fixed location analysis uses to compute treatment least-squares means:

$\mu + \tau_i + \bar{c}_i + (\bar{\tau} \bar{c})_{ii}$ In the random location effect analysis, you add ESTIMATE statements to the

PROC MIXED program shown in the last section (6.3.3.1) to compute these BLUPs. For example, for treatment 1, consider the following ESTIMATE statements:

```

estimate 'trt 1 LS Mean'
    intercept 1
    treatment 1 0 0 0;
estimate 'trt 1 BLUP "Narrow"'
    intercept 9
    treatment 9 |
    location 1 1 1 1 1 1 1 1 1
    location*treatment 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0
    1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0
    1 0 0 0 /divisor=9;
estimate 'trt 1 BLUP "alt"'
    intercept 9
    treatment 9 |
    location 1 1 1 1 1 1 1 1 1 / divisor=9;
estimate 'trt1 vs trt2'
    treatment 1 -1 0;
estimate 'trt 1 v 2 BLUP'
    treatment 9 -9 0 0 |
    location*treatment 1 -1 0 0 1 -1 0 0 1 -1 0 0
    1 -1 0 0 1 -1 0 0 1 -1 0 0
    1 -1 0 0 1 -1 0 0 1 -1 0 0 /
    divisor=9;

```

The first statement reproduces the random location least-squares means. The second statement applies the definition of the least-square mean for the fixed location analysis. Note that location and location \times treatment effects appear to the right of the vertical bar (|), as they are random effects in this analysis. The third statement is a modified BLUP, similar to the “intermediate” inference space BLUP introduced in Section 6.5. The final two statements estimate the difference between treatments 1 and 2: the first between least-squares means as defined in the random location model; the second between treatment 1 and 2 BLUPs, whose coefficients are identical to the difference between least-squares means in the fixed location effect model. Output 6.10 shows the results.

Output 6.10 Estimates Comparing LS-Means versus BLUPs for Treatment Means

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
trt 1 LS Mean	2.9196	0.1164	9.9	25.07	<.0001
trt 1 BLUP "Narrow"	2.9196	0.03871	67.9	75.41	<.0001
trt 1 BLUP "alt"	2.9196	0.04553	26.5	64.12	<.0001
trt1 vs trt2	0.1893	0.06102	24	3.10	0.0049
trt 1 v 2 BLUP	0.1893	0.05073	54	3.73	0.0005

Interpretation

“Trt 1 LS Mean” and “trt 1 BLUP ‘narrow’” yield estimates and standard errors identical to random location and fixed locations least-squares means, respectively. You can see that the estimates and standard errors are identical to the random location LS-mean and the fixed

location LS-mean for treatment 1, respectively. Likewise, the “trt1 vs trt2” estimate is identical to the difference between treatment 1 and 2 LS-means for the random location output, whereas the “trt 1 v 2 BLUP” produces output identical to the fixed location estimates of treatment difference. These estimates help clarify what the random and fixed location analyses are doing. The random location effect analysis produces interval estimates based on assuming that uncertainty results from the locations being a sample of the population, whereas the fixed location analysis assumes that uncertainty results only from experimental error within locations. If locations represent a larger population, the latter interval estimate is unrealistic. The alternate “trt 1 BLUP” allows for uncertainty to result from variation in treatment effects among locations, but not from variation in the locations themselves.

6.7 Location-Specific Inference in Multicenter Example

Location-specific inference means that you assess location main effect BLUPs, location \times treatment BLUPs, and simple effect differences between treatments at specific locations. As in previous examples, you use BLUPs rather than means because the BLUPs, unlike the means, use information about the variance among random location effects. Conventional means are not appropriate because they imply estimable functions based on both treatment and locations effects; this implication is misleading, and produces inappropriate interval estimates, since location is not a fixed effect.

Obtaining BLUPs

To illustrate, suppose you want the BLUP for treatment 1 at location A. The predictable function for this BLUP is $\mu + \tau_1 + c_A + (\tau c)_{1A}$. Similarly, for treatment 2, the BLUP is $\mu + \tau_2 + c_A + (\tau c)_{2A}$. The location-specific difference between treatment 1 and 2 is thus $\tau_1 - \tau_2 + (\tau c)_{1A} - (\tau c)_{2A}$. Similar BLUPs can be constructed for any location \times treatment combination or treatment difference at any given location.

You can also compute BLUPs for locations (averaged over all treatments), or differences between locations, if these are of interest in a particular study. For example, the predictable function for the BLUP of location A is

$$\mu + (1/4) \sum_i \tau_i + c_A + (1/4) \sum_i (\tau c)_{iA}$$

The BLUP of the difference between two locations would be, for example,

$$c_A - c_B + (1/4) \sum_i (\tau c)_{iA} - (1/4) \sum_i (\tau c)_{iB}$$

for location A versus location B. An alternative location A BLUP would be $\mu + \frac{1}{4} \sum \tau_i + c_A$. This would broaden the inference space to allow variation among treatment effects at each location (the location \times treatments effects) to be considered a contributor to uncertainty in this predictor. The BLUP for location difference with the equivalent inference space would be $c_A - c_B$.

You can expand the application of location-specific treatment differences to obtain the BLUP equivalent of the location \times treatment SLICE in the fixed location analysis. For example, the simultaneous contrast

$$\begin{aligned}\tau_1 - \tau_2 + (\tau c)_{1A} - (\tau c)_{2A} \\ \tau_1 - \tau_3 + (\tau c)_{1A} - (\tau c)_{3A} \\ \tau_1 - \tau_4 + (\tau c)_{1A} - (\tau c)_{4A}\end{aligned}$$

defines the BLUP equivalent for the slice among treatments for location A. This set is based on the result that the set of contrasts comparing treatment 1 to 2, 1 to 3, and 1 to 4 jointly imply that all treatment effects are equal. You could use any three independent contrast such that if they are all zero, then the treatment effects are equal, and obtain the same result.

Implementation Using PROC MIXED

You can compute these BLUPs using the following PROC MIXED statements:

```
proc mixed data=MultiCenter;
  class location block treatment;
  model response=treatment/ddfm=KR;
  random location block(location) location*treatment;
  estimate 'trt1 at loc A blup'
    intercept 1
    treatment 1 0 0 0 |
    location 1 0
    location*treatment 1 0;
  estimate 'trt2 at loc A blup'
    intercept 1
    treatment 0 1 0 0 |
    location 1 0
    location*treatment 0 1 0;
  estimate 'trt3 at loc A blup'
    intercept 1
    treatment 0 0 1 0 |
    location 1 0
    location*treatment 0 0 1 0;
  estimate 'trt4 at loc A blup'
    intercept 1
    treatment 0 0 0 1 |
    location 1 0
    location*treatment 0 0 0 1;
  estimate 'trt1 at loc B blup'
    intercept 1
    treatment 1 0 0 0 |
    location 0 1 0
    location*treatment 0 0 0 0 1 0;
  estimate 'trt2 at loc B blup'
    intercept 1
    treatment 0 1 0 0 |
    location 0 1 0
    location*treatment 0 0 0 0 0 1 0;
  estimate 'trt3 at loc B blup'
    intercept 1
    treatment 0 0 1 0 |
    location 0 1 0
    location*treatment 0 0 0 0 0 0 1 0;
  estimate 'trt4 at loc B blup'
    intercept 1
    treatment 0 0 0 1 |
    location 0 1 0
    location*treatment 0 0 0 0 0 0 0 1;
  estimate 'trt 1 v 2 at loc A'
    treatment 1 -1 0 |
```

```

      location*treatment 1 -1 0;
estimate 'trt 1 v 3 at loc A'
      treatment 1 -1 0 |
      location*treatment 0 1 -1 0;
estimate 'trt 1 v 4 at loc A'
      treatment 1 -1 0 |
      location*treatment 0 0 1 -1;
contrast 'slice at loc A'
      treatment 1 -1 0   |
      location*treatment 1 -1 0,
      treatment 0 1 -1 0 |
      location*treatment 0 1 -1 0,
      treatment 0 1 0 -1 |
      location*treatment 0 1 0 -1;
contrast 'slice at loc A'
      treatment 1-1 0   |
      location*treatment 1 -1 0,
      treatment 1 0 -1 0 |
      location*treatment 1 0 -1 0,
      treatment 1 0 0 -1 |
      location*treatment 1 0 0 -1;
contrast 'slice at loc B'
      treatment 1 -1 0   |
      location*treatment 0 0 0 0 1 -1 0,
      treatment 1 0 -1 0 |
      location*treatment 0 0 0 0 1 0 -1 0,
      treatment 1 0 0 -1 |
      location*treatment 0 0 0 0 1 0 0 -1;
run;

```

The KR (KENWARDROGER) option in the MODEL statement computes the approximate degrees of freedom and the bias-corrected standard errors and test statistics for the BLUPs using the Prasad-Rao-Jeske-Harville procedure discussed earlier in this chapter. Note that the contrast entitled ‘slice at loc A’ shows two different sets of independent comparisons among the treatments that both define the overall equality of treatment effects. You can see from the results shown in Output 6.11 that they produce identical results.

Output 6.11 PROC MIXED Location-Specific Results for Multicenter Data

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
trt1 at loc A blup	3.0443	0.09687	26.9	31.43	<.0001
trt2 at loc A blup	2.8723	0.09687	26.9	29.65	<.0001
trt3 at loc A blup	3.0291	0.09687	26.9	31.27	<.0001
trt4 at loc A blup	3.0879	0.09687	26.9	31.88	<.0001
trt1 at loc B blup	2.4172	0.09687	26.9	24.95	<.0001
trt2 at loc B blup	2.2668	0.09687	26.9	23.40	<.0001
trt3 at loc B blup	2.3608	0.09687	26.9	24.37	<.0001
trt4 at loc B blup	2.5060	0.09687	26.9	25.87	<.0001
trt1 at loc H blup	3.3797	0.09687	26.9	34.89	<.0001
trt2 at loc H blup	3.1213	0.09687	26.9	32.22	<.0001
trt3 at loc H blup	3.3624	0.09687	26.9	34.71	<.0001

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
trt4 at loc H blup	3.4099	0.09687	26.9	35.20	<.0001
trt 1 v 2 at loc A	0.1720	0.1112	8.89	1.55	0.1566
trt 1 v 3 at loc A	0.2221	0.1257	5.1	1.77	0.1364
trt 1 v 4 at loc A	0.1694	0.1211	5.83	1.40	0.2127

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
slice at loc A	3	8.89	1.43	0.2971
slice at loc A	3	8.89	1.43	0.2971
slice at loc B	3	8.89	1.63	0.2510
slice at loc C	3	8.89	1.19	0.3682
slice at loc D	3	8.89	1.98	0.1877
slice at loc E	3	8.89	2.57	0.1203
slice at loc F	3	8.89	2.42	0.1339
slice at loc G	3	8.89	2.30	0.1465
slice at loc H	3	8.89	2.86	0.0978
slice at loc I	3	8.89	1.11	0.3965

Interpretation

The “Estimates” table shows BLUPs for the four treatments at locations A, B, and H. The SAS code for location H was not shown in the program, but it is a straightforward adaptation of the code for locations A and B. The ESTIMATE statements can easily be adapted to the other locations as well.

The BLUPs are similar, but not equal, to the location \times treatment least-squares means for the fixed location analysis shown in Output 6.8. The differences are as follows:

3. The BLUPs are shrinkage estimators as described in Section 6.4.3 and are regressed toward the mean.
4. The standard errors reflect differences in the sources of variation considered random.

The SLICE BLUPs test the same differences, conceptually, as the fixed location slices, except these, too, are based on shrinkage estimators and on a broader inference space. Output 6.11 shows the results for all slices. In the interest of space, the SAS statements for locations C through I are omitted, but are defined similarly to those for locations A and B.

In many multilocation and multicenter trials, there are questions regarding the **stability** of treatments. In other words, do the same treatments perform well or poorly in all locations, or are there as-yet-unknown subpopulations such that some treatments perform optimally at certain types of locations whereas other treatments are better suited for other locations? By assessing the simple effects of treatments at each location using the BLUPs demonstrated in this example, the stability issue can be addressed using mixed model methods.

Note that for this example, the BLUP-based slices do not show nearly the extent of discrepancies seen in the fixed location slices. When fixed location analysis is used inappropriately for locations that are a sample of the population, the slices based in location \times treatment least-squares means (Output 6.7c) tend to be overly sensitive to variation in treatment effects among locations. In other words, use of fixed location analysis when center effects are more appropriately regarded as random increases the Type I error rate.

Output 6.12 shows results for BLUP-based slices defined using the same contrasts, but omitting the DDFM=KR option. This shows what happens when the standard errors are not corrected for bias and degrees of freedom are not approximated. You can see that the Type I error rate tends to be increased without the DDFM=KR correction. The DDFM=SATTERTH option corrects the degrees of freedom, but not the standard errors. The Type I error rate is lower without the DDFM option, but it is still excessive.

Output 6.12 MIXED Location-Specific Results for Multicenter Data, No DDFM=KR Option

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
slice at loc A	3	24	1.98	0.1434
slice at loc A	3	24	1.98	0.1434
slice at loc B	3	24	2.26	0.1077
slice at loc C	3	24	1.65	0.2052
slice at loc D	3	24	2.75	0.0651
slice at loc E	3	24	3.55	0.0294
slice at loc F	3	24	3.35	0.0357
slice at loc G	3	24	3.19	0.0419
slice at loc H	3	24	3.95	0.0201
slice at loc I	3	24	1.53	0.2319

Finally, you can define BLUPs on subsets of locations. For example, the following statements allow you to look at the four treatment BLUPs at their corresponding slices for three subsets of the locations:

1. A and C, two locations with no significant treatment effects
2. B, D, and I, three locations with marginally significant or nonsignificant differences among the means and somewhat different ranks among the treatment BLUPs than at other locations
3. E though H, the four locations with significant slices in the fixed location analysis

These subsets are mainly for demonstration purposes. Obviously, you could define other subsets that would be of greater interest, depending on additional specifics about the locations or the objectives of the study. The SAS statements are as follows. Output 6.13 shows the results, using the DDFM=KR option.

```

estimate 'trt1 @ Locs A,C'
  intercept 2
  treatment 2 0 0 0 |
  location 1 0 1 0 0 0 0 0 0
  location*treatment 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=2;
estimate 'trt2 @ Locs A,C'
  intercept 2
  treatment 0 2 0 0 |
  location 1 0 1 0 0 0 0 0 0
  location*treatment 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=2;
estimate 'trt3 @ Locs A,C'
  intercept 2
  treatment 0 0 2 0 |
  location 1 0 1 0 0 0 0 0 0
  location*treatment 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=2;
estimate 'trt4 @ Locs A,C'
  intercept 2
  treatment 0 0 0 2 |
  location 1 0 1 0 0 0 0 0 0
  location*treatment 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=2;
estimate 'trt1 @ Locs B,D,I'
  intercept 3
  treatment 3 0 0 0 |
  location 0 1 0 1 0 0 0 0 1
  location*treatment 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=3;
estimate 'trt2 @ Locs B,D,I'
  intercept 3
  treatment 0 3 0 0 |
  location 0 1 0 1 0 0 0 0 1
  location*treatment 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=3;
estimate 'trt3 @ Locs B,D,I'
  intercept 3
  treatment 0 0 3 0 |
  location 0 1 0 1 0 0 0 0 1
  location*treatment 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=3;
estimate 'trt4 @ Locs B,D,I'
  intercept 3
  treatment 0 0 0 3 |
  location 0 1 0 1 0 0 0 0 1
  location*treatment 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / divisor=3;
estimate 'trt1 @ Locs E,F,G,H'
  intercept 4
  treatment 4 0 0 0 |
  location 0 0 0 0 1 1 1 1 0
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 / divisor=4;

```

```

estimate 'trt2 @ Locs E,F,G,H'
  intercept 4
  treatment 0 4 0 0 |
  location 0 0 0 0 1 1 1 1 0
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0
    0 0 0 0 / divisor=4;
estimate 'trt3 @ Locs E,F,G,H'
  intercept 4
  treatment 0 0 4 0 |
  location 0 0 0 0 1 1 1 1 0
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0
    0 0 0 0 / divisor=4;
estimate 'trt4 @ Locs E,F,G,H'
  intercept 4
  treatment 0 0 0 4 |
  location 0 0 0 0 1 1 1 1 0
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1
    0 0 0 0 / divisor=4;
contrast 'slice at locs A,C'
  treatment 2 -2 0 |
  location*treatment 1 -1 0 0 0 0 0 0 0 1 -1 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0,
  treatment 2 0 -2 0 |
  location*treatment 1 0 -1 0 0 0 0 0 0 1 0 -1 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0,
  treatment 2 0 0 -2 |
  location*treatment 1 0 0 -1 0 0 0 0 0 1 0 0 -1 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0;
contrast 'slice at locs B,D,I'
  treatment 3 -3 0 |
  location*treatment 0 0 0 0 1 -1 0 0 0 0 0 0 0 1 -1 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 -1 0 0,
  treatment 3 0 -3 0 |
  location*treatment 0 0 0 0 1 0 -1 0 0 0 0 0 0 1 0 -1 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 0 -1 0,
  treatment 3 0 0 -3 |
  location*treatment 0 0 0 0 1 0 0 -1 0 0 0 0 0 1 0 0 -1 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 0 0 -1;
contrast 'slice at locs E,F,G,H'
  treatment 4 -4 0 |
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 -1 0 0 1 -1 0 0 1 -1 0 0 1 -1 0 0
    0 0 0 0,
  treatment 4 0 -4 0 |
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 0 -1 0 1 0 -1 0 1 0 -1 0 1 0 -1 0
    0 0 0 0,
  treatment 4 0 0 -4 |
  location*treatment 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    1 0 0 -1 1 0 0 -1 1 0 0 -1 1 0 0 -1
    0 0 0 0;

```

Output 6.13 MIXED Multicenter Results for Subsets of Locations

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
trt1 @ Locs A,C	3.0679	0.07035	36.8	43.61	<.0001
trt2 @ Locs A,C	2.9011	0.07035	36.8	41.24	<.0001
trt3 @ Locs A,C	3.0733	0.07035	36.8	43.69	<.0001
trt4 @ Locs A,C	3.0884	0.07035	36.8	43.90	<.0001
trt1 @ Locs B,D,I	2.5407	0.06447	26.3	39.41	<.0001
trt2 @ Locs B,D,I	2.3946	0.05891	48.1	40.65	<.0001
trt3 @ Locs B,D,I	2.5469	0.05891	48.1	43.24	<.0001
trt4 @ Locs B,D,I	2.6078	0.05891	48.1	44.27	<.0001
trt1 @ Locs E,F,G,H	3.1365	0.05226	59.2	60.02	<.0001
trt2 @ Locs E,F,G,H	2.8969	0.05226	59.2	55.43	<.0001
trt3 @ Locs E,F,G,H	3.1231	0.05226	59.2	59.76	<.0001
trt4 @ Locs E,F,G,H	3.1575	0.05226	59.2	60.42	<.0001

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
slice at locs A,C	3	17.6	2.27	0.1162
slice at locs B,D,I	3	32.4	3.23	0.0352
slice at locs E,F,G,H	3	53.1	7.24	0.0004

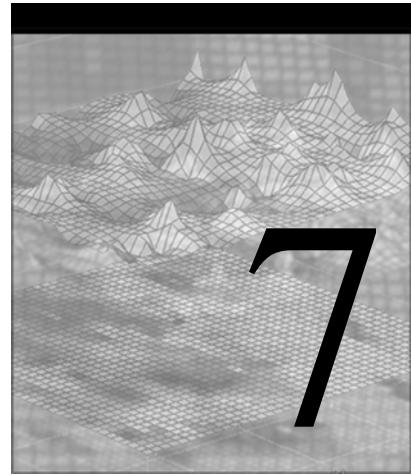
Interpretation

The results in the “Estimates” table tend to suggest that, despite difference in individual locations, the patterns of differences among the means for the three subsets of locations are similar. The “Contrasts” table shows treatment effect *p*-values ranging from 0.0004 to 0.1162. These discrepancies are considerably less than among individual locations. Depending on the context and specific objectives, this output may be taken as evidence that there are differences among subpopulations represented by the three subsets or as evidence that treatment effects are similar enough among all locations to justify population-wide or broad inference on the treatment means.

6.8 Summary

The chapter presented a nontechnical introduction to best linear unbiased prediction (BLUP) and the distinction between estimation and prediction in mixed models. Section 6.2 presented several examples to illustrate estimates and predictors that are of interest in practical situations.

Section 6.3 developed the main ideas, illustrating them through a two-way mixed model. The notion of inference space and how estimates and predictors can be manipulated to widen or narrow the scope of inference was introduced. Sections 6.4 through 6.7 showed how to obtain BLUPs using PROC MIXED and how to work with the various inference spaces implied by the estimates and BLUPs presented in these examples. Three data sets were used to show how to construct the programs and how to interpret the output.



Analysis of Covariance

7.1	Introduction	244
7.2	One-Way Fixed Effects Treatment Structure with Simple Linear Regression Models.....	245
7.2.1	Comparing the Regression Models.....	247
7.2.2	Summary of an Analysis of Covariance Strategy	250
7.2.3	Extensions to More Complex Regression Models	250
7.3	Example: One-Way Treatment Structure in a Randomized Complete Block Design Structure—Equal Slopes Model.....	251
7.3.1	Step 1: Fit the Model to Test the Slopes-Equal-to-Zero Hypothesis	251
7.3.2	Step 2: Determine If a Common Slope Model Is Adequate to Describe the Data	253
7.3.3	Step 3: Fit a Common Slope Model	255
7.3.4	Mixed Model Estimator: A Combined Estimator	258
7.4	Example: One-Way Treatment Structure in an Incomplete Block Design Structure—Time to Boil Water.....	263
7.4.1	Step 1: Fit the Model to Test the Equal Slopes Hypothesis.....	264
7.4.2	Step 2: Fit the Unequal Slopes Model.....	265
7.4.3	Step 3: Test for Lack of Fit of the Simple Linear Regression Model	267
7.4.4	Step 4: Testing the Equality of the Models at a Preselected Value of the Covariate	269
7.5	Example: One-Way Treatment Structure in a Balanced Incomplete Block Design Structure	272
7.5.1	Step 1: Fit the Unequal Slopes Model.....	272
7.5.2	Step 2: Test the Equal Slopes Hypothesis.....	274
7.5.3	Step 3: Fit Model with Unequal Slopes for Each Level of GRP	276
7.6	Example: One-Way Treatment Structure in an Unbalanced Incomplete Block Design Structure	281
7.6.1	Step 1: Fit the Unequal Slopes Model and Test the Slopes-Equal-to-Zero Hypothesis.....	282

7.6.2 Step 2: Test the Equal Slopes Hypothesis.....	283
7.6.3 Step 3: Fit a Common Slope Model	284
7.7 Example: Split-Plot Design with the Covariate Measured on the Large-Size Experimental Unit or Whole Plot.....	286
7.7.1 Step 1: Fit Unequal Slopes Model and Test Slopes-Equal-to-Zero Hypothesis.....	286
7.7.2 Step 2: Fit Factorial Effects Model for Intercepts and Slopes	288
7.7.3 Step 3: Fit Model with Slopes Expressed as Main Effects of the Two Factors.....	289
7.7.4 Step 4: Fit Model with Unequal Slopes for Each Gender.....	290
7.8 Example: Split-Plot Design with the Covariate Measured on the Small-Size Experimental Unit or Subplot.....	297
7.8.1 Step 1: Fit Model with Factorial Effects for Both Intercepts and Slopes	298
7.8.2 Step 2: Fit the Factorial Effects for Intercepts and a Main Effects Representation for the Slopes	299
7.8.3 Step 3: Fit the Factorial Effects Model for Intercepts and the Slopes as a Function of Teaching Method.....	300
7.8.4 Step 4: Fit a Common Slope Model	301
7.8.5 Comparison with PROC GLM	305
7.9 Example: Complex Strip-Plot Design with the Covariate Measured on an Intermediate-Size Experimental Unit.....	308
7.10 Summary.....	315

7.1 Introduction

Analysis of covariance is a strategy for analyzing data from a designed experiment where, in addition to the response variable, one or more continuous variables are measured on each experimental unit. Ideally, the additional continuous variables, called **covariates**, should be determined before the treatments have been applied to the experimental units. At a minimum, the values of the covariates should not be affected by the applied treatments. The analysis of covariance is often described as a method to remove variability in the experiment by accounting for variability in the experimental units that could not be controlled by the design structure. A more global view of analysis of covariance describes it as a **methodology to compare a series of regression models** (Milliken and Johnson 2002). The analysis of covariance model is a model that consists of both **classification**, or qualitative, variables and **continuous**, or quantitative, variables. The discussion in this chapter uses treatment or treatment combination to denote the levels of the classification variables in the model and covariates to denote the continuous variables. Categorical variables can be used as covariates in the modeling process, but those issues are not discussed here.

A major objective of an analysis is to make comparisons among the levels of the treatments or treatment combinations as if the experimental units all have the same value of the covariate or covariates. This process is accomplished by fitting regression models to the data from each of

the treatment combinations and then making comparisons among the regression lines at a common value of the covariate or common values of the covariates.

The basic model consists of the same functional relationship between response and the covariates with possibly different parameters for each of the treatments or treatment combinations. In essence, the basic model consists of a different **regression model** for each treatment or treatment combination in the treatment structure. Analysis of covariance is a strategy for making decisions about the form of the model and then comparing the model's parameters for each combination of classification variables in the treatment structure. The analysis of a one-way treatment structure in a completely randomized design structure, where the levels of the treatment are fixed effects, is used to establish the basic analysis of covariance strategy, which is the topic of Section 7.2. Most applications of analysis of covariance are to experimental designs with fixed treatment effects and some type of blocking. A random coefficient regression model is used to describe the data when the regression models occur for each of the levels of a random effect. Random coefficient models are discussed in Chapter 8. When blocking is used in the design structure, more than one size of experimental unit is generated. Models with fixed effects and more than one size experimental unit are **mixed models**. Section 7.5 presents an example of the analysis of such a design. These examples involve blocking and demonstrate the aspects of the mixed models equations concerning the combining of information from the intra-block and inter-block analyses.

Next the discussion expands to fixed effect treatment structure models with more complex design structures including nested and split-plot types of designs. The mixed model is very important in the analysis of complex design structures, as information about the coefficients of the covariates in the model can occur at several levels of the model and the solution of the mixed models equations combines the information from all sources into a single estimate.

The following sections present models for complete block, incomplete block, and balanced incomplete block designs, as well as three split-plot type designs where

- the covariate is measured on the large size of experimental unit
- the covariate is measured on the small size of experimental unit
- the covariate is measured on an intermediate size of experimental unit

7.2 One-Way Fixed Effects Treatment Structure with Simple Linear Regression Models

Consider a situation where data are collected from an experiment involving a one-way treatment structure in a completely randomized design, and the experimental units exhibit considerable variability that cannot be controlled by using some form of blocking. But, the researcher believes, say, from past experiences, that one or more characteristics of the experimental units can help describe some of the variability between experimental units. In this section, only one covariate is considered for the development of the strategy of analysis of covariance. Before the treatments were applied to the experimental units, the researcher measured the value of a covariate. The covariate should be considered a priori to be related to the variability observed in the experimental units. By measuring a covariate, the researcher attempts to account for variability in the experimental units that cannot conveniently be removed by blocking the experimental units into homogeneous groups. For example, in a feeding study the response to different diets may be affected by the size of the animals or the age of the animals. The animal scientist may measure the initial weight of each animal before the start of the experiment—i.e.,

before they are randomly assigned to the diets. The response variable can be the average daily gain of each animal computed for the amount of time the animals are in the study. At this point we have a fixed effects treatment structure (a set of s diets), a random sample of animals from a population of animals with initial weights denoted by x , and the response, average daily gain, denoted by Y . The average daily gain values are calculated by dividing the total amount gained during the feeding trial by the number of days the animals were on trial. A **statistical model** that can be used to describe the relationships among the response variable Y , the classification variable for diets, the covariate x , and the experimental units or animals is

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij} \quad (7.1)$$

where

$$i = 1, 2, \dots, s$$

$$j = 1, 2, \dots, n$$

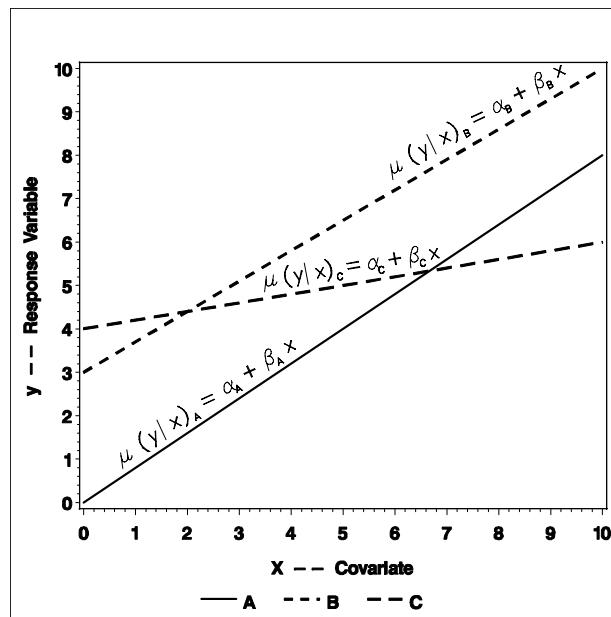
$$e_{ij} \sim iid N(0, \sigma_e^2)$$

α_i is the intercept of the model for treatment i

β_i is the slope of the model for treatment i

Model (7.1) represents a set of **simple linear regression** lines with possibly different slopes and intercepts as exhibited in Figure 7.1, where the mean of the response from the i^{th} treatment at a given value of the covariate is expressed as $\mu(Y|x)_i = \alpha_i + \beta_i x$.

Figure 7.1 Three Simple Regression Lines with Unequal Slopes



Before you use these models to describe data, you must be sure that the simple linear regression models (or the selected regression model) adequately describe the data for each treatment. This process involves identification of outliers, testing for the equality of variances, and checking the adequacy of the models for each treatment. This aspect of the analysis of covariance is often overlooked when put into the context of measuring variables to adjust for variability in the experimental units. When the analysis of covariance is cast in the context of comparing

regression models, you realize that all **regression diagnostics** used for usual regression analysis should be utilized for fitting and comparing these models.

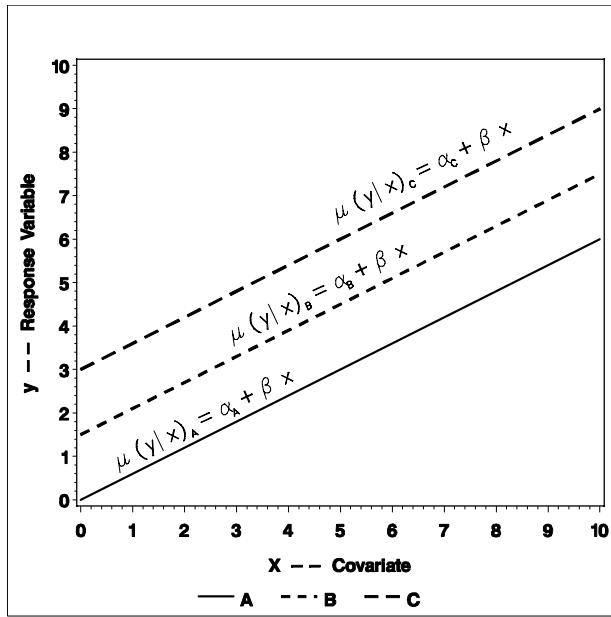
A basic philosophy is to use the simplest possible model for the covariate part of the model; that is, avoid using a more complicated model than is necessary to describe the relationship between the mean of the response variable and the covariate (Milliken and Johnson 2002). Using the simplest model prevents overfitting of the data and provides more degrees of freedom for estimating the error. In this light, the first hypothesis to be tested is to determine if a model without the covariate can be used to adequately describe the data; i.e., test $H_0: \beta_1 = \beta_2 = \dots = \beta_s = 0$ versus H_a : (not H_0), called the **slopes-equal-to-zero** hypothesis. (If you have graphed the data for each treatment, you probably already have an idea if the slopes are different from zero.) If there is evidence that the slopes are not all zero, then determine if a model with a common slope can be used to describe the data; i.e., test $H_0: \beta_1 = \beta_2 = \dots = \beta_s = \beta$ versus H_a : (not H_0), where β is unspecified. If the common slope model can be used to adequately describe the data, the process of comparing the regression models is greatly simplified. The common slope analysis of covariance model is

$$Y_{ij} = \alpha_i + \beta x_{ij} + e_{ij} \quad (7.2)$$

where $\alpha_1, \alpha_2, \dots, \alpha_s$ are the intercepts for treatments 1, 2, ..., s and β is the common slope.

As in model (7.1), we assume $e_{ij} \sim iid N(0, \sigma_e^2)$. Common slope models can be represented as shown in Figure 7.2.

Figure 7.2 Three Simple Linear Regression Models with Equal Slopes



7.2.1 Comparing the Regression Models

After you decide on the form of the covariate part of the model—i.e., determine if the covariate is useful and if so, determine the form of the slopes in the model—the next step is to make comparisons among the regression models, which may include comparisons among the intercepts, among the slopes, or among the models evaluated at specified values of the covariate. Making

comparisons among the models at a specific value of the covariate, say, x_0 , is tantamount to treatment comparisons as if all experimental units had a covariate value equal to x_0 .

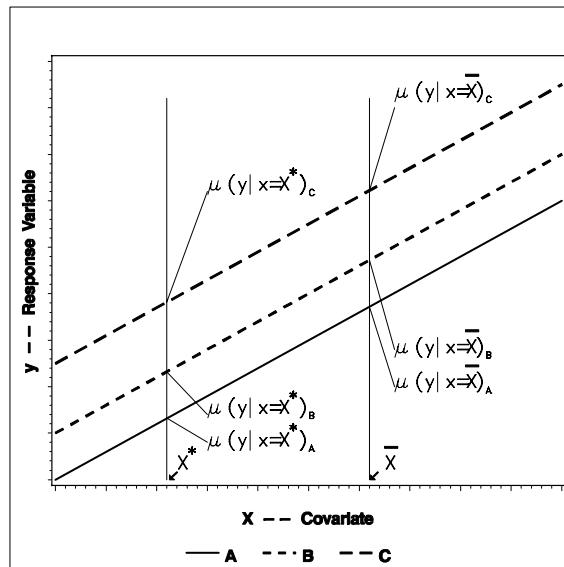
If the slopes are all zero, the covariate is not needed in the model, and you compare the means (intercepts for regression lines with zero slopes) using standard analysis of variance techniques (as in Chapter 3), such as testing equality of the means, investigating contrasts of the means, or carrying out multiple comparisons. If models with common slopes are adequate to describe the data, they form a series of parallel lines and can be compared by estimating the distance between the lines. The estimated regression models with common slope—i.e., the estimated means of Y for a given value of x for treatment i —are $\hat{\mu}(y|x)_i = \hat{\alpha}_i + \hat{\beta}x$, $i = 1, 2, \dots, s$.

A comparison of the regression lines for treatments one and two at $x = x^*$ is

$$\hat{\mu}(y|x^*)_1 - \hat{\mu}(y|x^*)_2 = \hat{\alpha}_1 - \hat{\alpha}_2$$

This is a comparison of the two intercepts and is independent of the value of the covariate. The quantities $\hat{\mu}(y|x^*)_i$, $i = 1, 2, \dots, s$, are predicted values obtained from the estimated regression lines evaluated at $x = x^*$, as shown in Figure 7.3. These predicted values are called **adjusted means** or **least-squares means**. Also displayed in Figure 7.3 are the estimated values of the regression lines at $x = \bar{x}$. The predicted values evaluated at $x = \bar{x}$ are the usual adjusted means computed for analysis of covariance and they are designated as least-squares means by the SAS System (LSMEANS statement). An important point is that adjusted means can be computed at values of x other than $x = \bar{x}$. Predictions from a regression model evaluated at some value $x_0 \neq \bar{x}$ can be useful. For regression models with a common slope, the differences among adjusted means involve only the intercepts; that is, they are independent of the value of the covariate (Figure 7.3).

Figure 7.3 Comparison of Equal Slopes Regression Models at $\mu(y|x=X^*)_k$, Which Denotes the Mean of the k^{th} Treatment Evaluated at x^* , and $\mu(y|x=\bar{X})_k$, Which Denotes the Mean of the k^{th} Treatment Evaluated at \bar{X}



For the unequal slopes models, the estimated mean of y for a given value of x for the i^{th} treatment is $\hat{\mu}(y|x)_i = \hat{\alpha}_i + \hat{\beta}_i x$, $i = 1, 2, \dots, s$. A comparison of the regression lines for treatments one and two evaluated at $x = x^*$ is

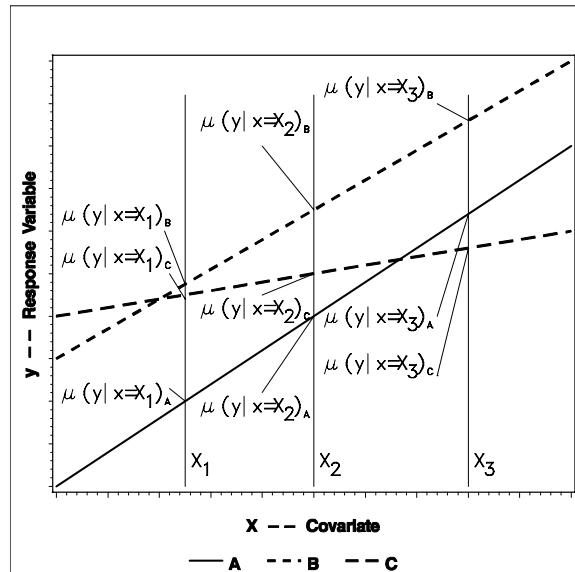
$$\hat{\mu}(y|x^*)_1 - \hat{\mu}(y|x^*)_2 = \hat{\alpha}_1 + \hat{\beta}_1 x^* - \hat{\alpha}_2 - \hat{\beta}_2 x^* = \hat{\alpha}_1 - \hat{\alpha}_2 + (\hat{\beta}_1 - \hat{\beta}_2)x^*$$

This—unlike for the common slope model—does depend on the value of the covariate. Because the comparison of treatments now depends on the value of the covariate, this model is called the **covariate by treatment interaction model**. To provide an appropriate analysis when the slopes are unequal, the models should be compared at a minimum of three values of x , such as at a low, middle, and high value of x . Potential choices are the minimum value of x , the mean value of x , and the maximum value of x , or the lower γ percentile, the median, and the upper γ percentile of the distribution of the x values. Reasonable choices for γ are 5, 10, or 25. At each of the selected values of x , contrasts of interest or multiple comparisons should be used to compare the treatments. Figure 7.4 displays three models being compared at three values of x , where the points of intersection of the models with the vertical lines are the respective adjusted treatment means. Other choices for the analysis include the following:

- using a multiple comparison procedure to make pairwise comparisons between slopes in an attempt to form groups of treatments (treatments *within* a group have common slopes, and treatments *between* groups have unequal slopes)
- constructing confidence bands about differences of models for each pair of treatments in an attempt to determine ranges of the covariate where the models are likely to have similar response levels and where the models are likely to have different response levels

These paths of analysis are used in some of the examples that follow. But first, the discussion to this point is summarized in the next section.

Figure 7.4 Comparisons of Unequal Slope Regression Models for Three Models at Three Values of X ; $\mu(y|x=X^*)_k$ Denotes the Mean of the k^{th} Treatment Evaluated at X^*



7.2.2 Summary of an Analysis of Covariance Strategy

Making sure the selected functional form of the covariate part of the model does in fact describe the data is the first essential step in the analysis of covariance strategy. If the simple linear regression model is not appropriate, then a more complex model must be selected for the analysis. To this point, we have considered only the simple linear regression model as the functional form and the following discussion centers on that model. There are several ways to approach the process of determining the form of the covariate part of the model. One strategy is summarized in Table 7.1. Following the steps in Table 7.1 provides a procedure to determine the form of the covariate part of the model. The resulting model can be used to compare the treatments by comparing the resulting regression models.

Table 7.1 Strategy for Determining the Form of the Covariate Part of the Models for Simple Linear Regression Models

Step	Action
1	Make sure the simple linear regression model describes the data from each treatment and that regression diagnostics are used to validate the assumptions.
2	Test the hypothesis that all the slopes are equal to zero. <ul style="list-style-type: none"> a) If fail to reject, go to step 3. b) If reject, go to step 4.
3	Fit a common slope model to the data and test the hypothesis that the slope is equal to zero. <ul style="list-style-type: none"> a) If fail to reject, compare the treatment means using analysis of variance. b) If reject, use a parallel lines model and compare the treatment regression models by comparing the intercepts or adjusted means (LS-means).
4	Test the hypothesis that the slopes are equal. <ul style="list-style-type: none"> a) If fail to reject, use a common slope model and compare the treatment regression models by comparing the intercepts or adjusted means (LS-means). b) If reject, go to step 5.
5	Use the unequal slopes model. <ul style="list-style-type: none"> a) Compare the slopes of the treatments. b) Compare the models at a minimum of three values of the covariate. c) Construct confidence bands about the differences of selected pairs of models.

7.2.3 Extensions to More Complex Regression Models

To this point, the discussion has focused on the simple linear regression model, but the same methodology can be extended to any regression model required to adequately describe the data. For example, a quadratic model in x may be required to describe the relationship between y and x for each treatment. You still need to include the simplest form of the covariate model by determining if the slopes are equal across treatments for x^2 as well as for x . If there are several

possible covariates, then the equality of slopes must be investigated for each possible covariate. In this case, the steps in Table 7.1 must be followed for each covariate. The strategy does not involve comparing the treatments by analysis of variance unless it is decided that the slopes for all covariates across all treatments are zero. When the slopes are unequal for a covariate, the models need to be compared at a minimum of three values for that covariate. If there are k covariates with unequal slopes and the models are to be compared at combinations of three values of each covariate, then there are 3^k combinations where the models need to be compared. Thus, it is imperative that the covariate part of the model be simplified as much as possible.

7.3 Example: One-Way Treatment Structure in a Randomized Complete Block Design Structure—Equal Slopes Model

The data in Data Set 7.3, “Average Daily Gain,” in Appendix 2, “Data Sets,” are average daily gains (ADG) of steers fed for 160 days. The treatments are four diets consisting of a base ration and three levels of a medicated feed additive added to the base ration. The objective of the experiment is to determine the optimal level of feed additive to maximize the average daily gain. The steers were housed in barns, the blocking factor, where each barn held four steers and the steers were individually fed.

The 32 steers were randomly assigned to the eight barns, and the four diets were randomly assigned to the four steers in each barn. Because the steers were of varying initial weights, the initial weights (iwt) were measured for use as a possible covariate. The model is constructed assuming the diets are a fixed effect, the barns or blocks are a random effect, and the functional form of the covariate part of the model is a simple linear regression model.

7.3.1 Step 1: Fit the Model to Test the Slopes-Equal-to-Zero Hypothesis

Model

A model to describe this data is

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + b_j + e_{ij} \quad (7.3)$$

where

$$i = 1, 2, 3, 4$$

$$j = 1, 2, \dots, 8$$

α_i denotes the intercept of the i^{th} diet model

β_i denotes the slope of the i^{th} diet model

$b_j \sim N(0, \sigma_b^2)$ denotes the effect of the j^{th} block

$e_{ij} \sim iid N(0, \sigma_e^2)$ denotes the experimental unit error

b_j and e_{ij} are independent random variables

Program

The following program fits model (7.3):

```

proc mixed data=ex7_3_rcb;
  class trt blk;
  model adg=trt iwt*trt/noint solution ddfm=kr;
  random blk;
run;

```

Selected results are presented in Output 7.1.

Results

Output 7.1 Fit of the Unequal Slope Model and Test of the Slopes-Equal-to-Zero Hypothesis

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.2593
Residual	0.04943

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	0	0.4391	0.7153	19.7	0.61	0.5463
trt	10	1.4261	0.6435	20.5	2.22	0.0381
trt	20	0.4796	0.5538	21.4	0.87	0.3960
trt	30	0.2001	0.7799	19.3	0.26	0.8002
iwt*trt	0	0.002294	0.001759	17.4	1.30	0.2091
iwt*trt	10	0.001083	0.001499	17.7	0.72	0.4795
iwt*trt	20	0.003366	0.001300	17.7	2.59	0.0187
iwt*trt	30	0.004448	0.002095	17.4	2.12	0.0484

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	4	19.1	1.30	0.3062
iwt*trt	4	17.5	3.43	0.0306

Interpretation

The “Covariance Parameter Estimates” table reports the estimates of the two variance components, one for the block-to-block variation, denoted by BLK, and one for the residual variance. The NOINT option in the MODEL statement is used so that PROC MIXED fits a model with nonsingular design matrix and thus provides estimates of the intercepts and slopes, as shown in the “Solution for Fixed Effects” table. If the NOINT option is not used, the design matrix corresponding to the intercepts is singular and the estimates of the intercepts satisfy the set-to-zero restrictions (see Milliken and Johnson 1992, Chapter 6). The MODEL statement includes the $iwt \times trt$ term in the model without the iwt term. As you see a little later, excluding the iwt term enables the covariate part of the model to be nonsingular, thus providing estimates of the slopes. Including the iwt provides estimates of the slopes satisfying the set-to-zero restriction. The “Solution for Fixed Effects” table contains the estimates of the intercepts

(Effect=trt) and slopes (Effect=iwt*trt) for each of the four treatments. The Type 3 F -statistic corresponding to iwt*trt in the “Type 3 Tests of Fixed Effects” table tests the hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus H_a : (not H_0). The p -value of 0.0306 indicates that the slopes are most likely not all equal to zero. The Type 3 F -statistic corresponding to TRT in the “Type 3 Tests of Fixed Effects tests” table the hypothesis $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus H_a : (not H_0), a comparison of the regression models at iwt=0. This does not provide meaningful comparisons, as the initial weight of the animals cannot be zero.

7.3.2 Step 2: Determine If a Common Slope Model Is Adequate to Describe the Data

Model

The following code fits the model

$$Y_{ij} = \alpha_4 + (\alpha_i - \alpha_4) + \beta_4 x_{ij} + (\beta_i - \beta_4) x_{ij} + b_j + e_{ij} \quad (7.4)$$

where

$$i = 1, 2, 3, 4$$

$$j = 1, 2, \dots, 8$$

α_i denotes the intercept of the i^{th} diet model

β_i denotes the slope of the i^{th} diet model

$b_j \sim N(0, \sigma_b^2)$ denotes the effect of the j^{th} block

$e_{ij} \sim \text{iid } N(0, \sigma_e^2)$ denotes the experimental unit error

b_j and e_{ij} are independent random variables

Model (7.4) is a reparameterization of model (7.3) where the slopes have been re-expressed as $\beta_i = \beta_4 + (\beta_i - \beta_4)$. In this parameterization the Type 3 F -statistic corresponding to IWT*TRT tests the hypothesis of equal slopes.

Program

The following code fits model (7.4) to the data:

```
proc mixed data=ex7_3_rcb;
  class trt blk;
  model adg=trt iwt iwt*trt/solution ddfm=kr;
  random blk;
run;
```

When the MODEL statement includes the iwt term (the covariate) and the iwt \times trt term (the interaction), the model becomes singular and the set-to-zero restrictions are used to handle the singularity. The coefficient of iwt becomes the slope corresponding to the last treatment, and the coefficients of iwt \times trt become the deviations of the given treatment’s slope from that of the last treatment.

Results

Selected results of fitting model (7.4) are displayed in Output 7.2; the estimates of the fixed effects reflect the above parameterization of the model.

Output 7.2 Fit of the Unequal Slope Model and Test of the Equal Slope Hypothesis

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.2593
Residual	0.04943

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		0.2001	0.7799	19.3	0.26	0.8002
trt	0	0.2390	1.0544	17.5	0.23	0.8233
trt	10	1.2260	0.9617	17.5	1.27	0.2191
trt	20	0.2795	0.9382	17.5	0.30	0.7693
trt	30	0
iwt		0.004448	0.002095	17.4	2.12	0.0484
iwt*trt	0	-0.00215	0.002808	17.5	-0.77	0.4532
iwt*trt	10	-0.00337	0.002533	17.4	-1.33	0.2011
iwt*trt	20	-0.00108	0.002508	17.5	-0.43	0.6713
iwt*trt	30	0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	17.4	0.86	0.4780
iwt	1	17.6	10.50	0.0046
iwt*trt	3	17.4	0.93	0.4485

The estimates of the variance components are identical to those in Output 7.1. Comparing information from the “Solution for Fixed Effects” table in Outputs 7.1 and 7.2, you see that the estimate corresponding to the iwt effect in Output 7.2 is the same as the estimate corresponding to iwt*trt 30 in Output 7.1, and the estimate corresponding to iwt*trt 0 in Output 7.2 is equal to the difference in the slopes (iwt*trt 0) – (iwt*trt 30) in Output 7.1. The Type 3 F -statistic for the iwt*trt effect in the “Type 3 Tests of Fixed Effects” table tests the hypothesis $H_0: \beta_1 - \beta_4 = \beta_2 - \beta_4 = \beta_3 - \beta_4 = 0$ versus $H_a:$ (not H_0). Note that this null hypothesis implies $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$, where β is unspecified. The p -value of 0.4485 indicates that there is insufficient evidence to reject the notion of unequal slopes. A common slope model should be adequate to describe the relationship between average daily gain and IWT across the four treatments. The Type 3 F -statistic corresponding to the trt effect in the “Type 3 Tests of Fixed Effects” table tests the hypothesis $H_0: \alpha_1 - \alpha_4 = \alpha_2 - \alpha_4 = \alpha_3 - \alpha_4 = 0$ versus $H_a:$ (not H_0), which compares the regressions at iwt=0, and again, is not meaningful because the initial weight of the animals cannot be zero.

7.3.3 Step 3: Fit a Common Slope Model

The preceding results indicate that the common slope model appears to be adequate to describe the data. (Plot the residuals by treatment to verify there are no patterns.)

Model

A common slope model is

$$Y_{ij} = \alpha_i + \beta x_{ij} + b_j + e_{ij} \quad (7.5)$$

where

$$i = 1, 2, 3, 4$$

$$j = 1, 2, \dots, 8$$

α_i denotes the intercept of the i^{th} diet model

β denotes the common slope of the diet models

$b_j \sim N(0, \sigma_b^2)$ denotes the effect of the j^{th} block

$e_{ij} \sim iid N(0, \sigma_e^2)$ denotes the experimental unit error

b_j and e_{ij} are independent random variables

Program

The following program fits model (7.5) to the data and uses ESTIMATE and LSMEANS statements to provide the final analysis:

```
proc mixed data=ex7_3_rcb;
  class trt blk;
  model adg = trt iwt/ solution ddfm=kr;
  random blk;
  estimate 'linear' trt -3 -1 1 3;
  estimate 'quad' trt -1 1 1 -1;
  estimate 'cubic' trt -1 3 -3 1;
  lsmeans trt / diff;
  lsmeans trt / at iwt=300;
  lsmeans trt / at iwt=500;
run;
```

Results

The results of fitting model (7.5) to the data are shown in Output 7.3.

Output 7.3 Fit of the Equal Slope Model

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.2408
Residual	0.05008

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		0.8011	0.3584	27	2.24	0.0339
trt	0	-0.5521	0.1149	20	-4.81	0.0001
trt	10	-0.06857	0.1191	20.1	-0.58	0.5713
trt	20	-0.08813	0.1164	20	-0.76	0.4577
trt	30	0
iwt		0.002780	0.000842	21.1	3.30	0.0034

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	20	10.16	0.0003
iwt	1	21.1	10.89	0.0034

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
linear	1.6367	0.3643	20	4.49	0.0002
quad	0.3954	0.1650	20	2.40	0.0265
cubic	0.6108	0.3538	19.9	1.73	0.0998

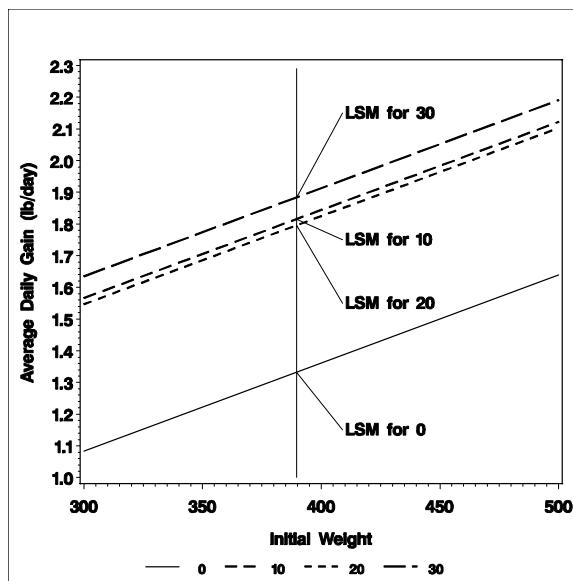
Least Squares Means							
Effect	trt	iwt	Estimate	Standard Error	DF	t Value	Pr > t
trt	0	389.59	1.3320	0.1907	9.02	6.98	<.0001
trt	10	389.59	1.8155	0.1914	9.13	9.49	<.0001
trt	20	389.59	1.7959	0.1908	9.04	9.41	<.0001
trt	30	389.59	1.8841	0.1923	9.29	9.80	<.0001
trt	0	300.00	1.0829	0.2056	11.7	5.27	0.0002
trt	10	300.00	1.5665	0.2116	12.8	7.40	<.0001
trt	20	300.00	1.5469	0.2079	12.1	7.44	<.0001
trt	30	300.00	1.6350	0.1973	10.2	8.29	<.0001
trt	0	500.00	1.6389	0.2116	12.8	7.75	<.0001
trt	10	500.00	2.1224	0.2056	11.7	10.32	<.0001
trt	20	500.00	2.1028	0.2091	12.3	10.06	<.0001
trt	30	500.00	2.1910	0.2241	15.2	9.78	<.0001

Differences of Least Squares Means								
Effect	trt	_trt	iwt	Estimate	Standard Error	DF	t Value	Pr > t
trt	0	10	389.59	-0.4835	0.1129	20	-4.28	0.0004
trt	0	20	389.59	-0.4639	0.1121	19.9	-4.14	0.0005
trt	0	30	389.59	-0.5521	0.1149	20	-4.81	0.0001
trt	10	20	389.59	0.01956	0.1122	19.9	0.17	0.8634
trt	10	30	389.59	-0.06857	0.1191	20.1	-0.58	0.5713
trt	20	30	389.59	-0.08813	0.1164	20	-0.76	0.4577

Interpretation

The estimates of the variance components are shown in the “Covariance Parameter Estimates” table; $\hat{\sigma}_e^2 = 0.05008$ and $\hat{\sigma}_b^2 = 0.2408$. Estimates of the intercepts (using the set-to-zero solution) and of the common slope are found in the “Solution for Fixed Effects” table. The estimate of the intercept for treatment 0 is computed as $0.8011 - 0.5521 = 0.2480$. The “Type 3 Tests of Fixed Effects” table indicates that the common slope is significantly different from zero ($p = 0.0034$) and that the distances between the regression lines are significantly different ($p = 0.0003$). The comparisons of the intercepts (Type 3 test for trt effect) are in fact comparisons of the distances between the parallel models, which is an interesting hypothesis. Figure 7.5 displays the four parallel regression (common slope) models, where ADG is the response variable and iwt is the regressor variable or covariate.

Figure 7.5 Plot of Estimated Regression Lines with LSMEANS

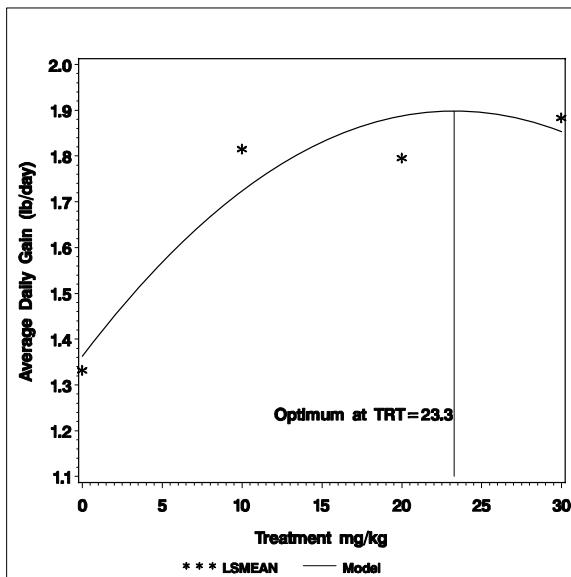


The “Least Squares Means” output provides estimates of the regression models at the average value of iwt, which in this case equals 389.59. The least-squares means were also requested for values of iwt of 300 and 500 to provide the endpoints of the regression lines in Figure 7.5.

Because the treatment levels are quantitative—i.e., the levels of the feed additive are 0 pmm, 10 pmm, 20 pmm, and 30 pmm—the ESTIMATE statements are used to evaluate the shape of the response curve—i.e., to evaluate how the ADG means change as the amount of feed additive

changes. The coefficients in the ESTIMATE statements correspond to the linear, quadratic, and cubic orthogonal polynomials for four equally spaced levels. The ESTIMATE statements were used to evaluate the curvature of the average daily gain least-squares means as a function of dose or treatment. The results are shown in the “Estimates” table of Output 7.3 and indicate that there are significant linear and quadratic effects ($p = 0.002$ and $p = 0.0265$, respectively) and a marginal cubic effect ($p = 0.0998$). It is expected that the true response is of the general quadratic shape, and thus the second degree model was used to describe the relationship among the average daily gain means and the level of dose. The response curve showing the relationship between the mean ADG and the levels of feed additive is displayed in Figure 7.6.

Figure 7.6 Plot of Least-Squares Means as a Function of Treatment Level



The pairwise comparisons results in the “Differences of Least Squares Means” table in Output 7.3 indicate that treatments 10, 20, and 30 are significantly different from the control but are not different among themselves. A quadratic function was fit to the least-squares means evaluated at 389.59, and the line in Figure 7.6 is a plot of the estimated quadratic function. The maximum predicted average daily gain occurs at a dose of 23.3 ppm. The complete analysis involves using some strategy to estimate the level of feed additive that provides the optimal ADG response, a task beyond the scope of this presentation (for a detailed discussion of methods of estimating the optimal dose, see Remmenga et al. 1997).

7.3.4 Mixed Model Estimator: A Combined Estimator

One of the important features of the mixed models analysis is that it combines information from all parts of the model into its estimates of the fixed effects of the model. For this example, there is information about the common slope of the models from within-block as well as between-block comparisons. The mixed model estimate of the slope is constructed by combining the within-block (intra-block) information with the between-block (inter-block) information. This example provides an opportunity to demonstrate this process of information combination.

Intra-block Information

The intra-block information is obtained using PROC MIXED, where the following code fits the common slope model with blk included in the MODEL statement. Thus, the intra-block analysis is obtained by assuming that the blocks are fixed.

Program

```
proc mixed data=ex7_3_rcb;
  class blk trt;
  model adg= trt iwt blk/solution;
run;
```

Results

The results from this program are listed in Output 7.4, which includes the estimates of the variance, common slope (bold), intercepts, and block effects and the fixed effects analysis of variance table.

Output 7.4 Fit of the Equal Slope Model to Obtain Intra-block Information about the Common Slope Using PROC MIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	0.05000

Solution for Fixed Effects							
Effect	blk	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			-0.1834	0.3081	20	-0.60	0.5584
trt		0	-0.5456	0.1148	20	-4.75	0.0001
trt		10	-0.05847	0.1191	20	-0.49	0.6287
trt		20	-0.08022	0.1163	20	-0.69	0.4983
trt		30	0
iwt			0.002572	0.000845	20	3.04	0.0064
blk	1		0.8697	0.1632	20	5.33	<.0001
blk	2		1.2210	0.1673	20	7.30	<.0001
blk	3		1.3292	0.1670	20	7.96	<.0001
blk	4		1.5935	0.1673	20	9.52	<.0001
blk	5		1.5182	0.1621	20	9.36	<.0001
blk	6		1.0935	0.1593	20	6.87	<.0001
blk	7		0.8508	0.1590	20	5.35	<.0001
blk	8		0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	20	10.12	0.0003
iwt	1	20	9.27	0.0064
blk	7	20	18.17	<.0001

Interpretation

The intra-block estimate of the common slope (row iwt in the “Solutions for Fixed Effects” table) and its associated estimated standard error based on the intra-block analysis are

$$\hat{\beta}_a = 0.002572 \text{ and } \hat{\sigma}_{\hat{\beta}_a} = 0.000845$$

Inter-block Information

The inter-block model is constructed from the model for the block means (or totals) as

$$\bar{y}_{ij} = \bar{\mu} + \beta \bar{x}_{ij} + b_j + \bar{e}_{ij}$$

where

$$j = 1, 2, \dots, 8$$

$$Var[\bar{y}_{ij}] = Var[b_j + \bar{e}_{ij}] = \sigma_b^2 + \frac{\sigma_e^2}{4} = \frac{4\sigma_b^2 + \sigma_e^2}{4}$$

The inter-block model is fit to the data using PROC MIXED with the following code. PROC MEANS is used to compute the block means, and the WEIGHT NADP statement scales the weighted sums of squares to make them comparable to sums of squares in Output 7.4. (We are weighting by the block size because the variance of a block mean is a scalar multiple of the reciprocal of the block size.)

Program

```

proc sort data=ex7_3_rcb; by blk;
proc means mean n data=ex7_3_rcb;
  by blk;
  var adg iwt;
  output out=means mean = madg miwt
            n      = nadg niwt;
run;
proc mixed data=means;
  model madg = miwt / solution;
  weight nadg;
run;

```

Results

The results are given in Output 7.5 where madg and miwt denote the block means for the ADG and iwt, respectively.

Output 7.5 Fit of the Equal Slope Model to Obtain Inter-block Information about the Common Slope Using PROC MIXED

Model Information	
Data Set	WORK.MEANS
Dependent Variable	madg
Weight Variable	nadg

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	0.8202

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-2.2053	1.7570	6	-1.26	0.2561
miwt	0.01004	0.004491	6	2.24	0.0667

Interpretation

The inter-block information about the common slope is in the “Solution for Fixed Effects” portion of Output 7.5. The row corresponding to miwt (mean initial weight) yields

$$\hat{\beta}_r = 0.01004 \text{ and } \hat{\sigma}_{\hat{\beta}_r} = 0.004491.$$

Combined Estimate of the Common Slope

The mean squared error estimate from the intra-block analysis in Output 7.4 is 0.05000. This is the estimate of σ_e^2 . The mean squared error estimate from the inter-block analysis in Output 7.5 is 0.8202. This is the estimate of $\sigma_e^2 + 4\sigma_b^2$. The set of equations yielding methods of moments estimates of the variance components is

$$\begin{aligned} E[MSERROR(\text{intra-block})] &= \sigma_e^2 \\ E[MSERROR(\text{inter-block})] &= \sigma_e^2 + 4\sigma_b^2 \end{aligned}$$

The solution to the method of moments equations for the two variance components is (see Chapter 3)

$$\hat{\sigma}_e^2 = 0.05000$$

$$\hat{\sigma}_b^2 = \frac{MSERROR(\text{inter}) - MSERROR(\text{intra})}{4} = 0.1926$$

The combined estimate of the slope is computed as the weighted average of the intra-block estimate of the slope and the inter-block estimate of the slope where the weights are the inverses of the respective variances as

$$\hat{\beta}_c = \frac{1}{\frac{1}{\hat{\sigma}_{\hat{\beta}_a}^2} + \frac{1}{\hat{\sigma}_{\hat{\beta}_r}^2}} \left[\frac{\hat{\beta}_a}{\hat{\sigma}_{\hat{\beta}_a}^2} + \frac{\hat{\beta}_r}{\hat{\sigma}_{\hat{\beta}_r}^2} \right]$$

The combined estimate of the common slope using the method of moments estimates of the variance components is $\hat{\beta}_c = 0.0028267$ and $\hat{\sigma}_{\hat{\beta}_c} = 0.00082997$. The following program uses the preceding two estimates of the variance components (via the HOLD=1,2 option) to provide the combined estimate of the slope and its estimated standard error.

Program

```
proc mixed data=ex7_3_rcb covtest;
  class trt blk;
  model adg=trt iwt / solution;
  random blk;
  parms (.1926) (.0500) / hold=1,2;
run;
```

Results

The results are in Output 7.6, where the estimate of the slope and its estimated standard error are $\hat{\beta}_c = 0.002827$ and $\hat{\sigma}_{\hat{\beta}_c} = 0.000830$, respectively.

Output 7.6 Fit of the Equal Slope Model Using Method of Moments Estimates of the Variance Components to Obtain Combined Intra-block–Inter-block Estimate of the Slope and Its Estimated Standard Error

Covariance Parameter Estimates				
Cov Parm	Estimate	Standard Error	Z Value	Pr Z
blk	0.1926	0	.	.
Residual	0.05000	0	.	.

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		0.7842	0.3460	7	2.27	0.0578
trt	0	-0.5535	0.1147	20	-4.83	0.0001
trt	10	-0.07084	0.1188	20	-0.60	0.5577
trt	20	-0.08991	0.1162	20	-0.77	0.4480
trt	30	0
iwt		0.002827	0.000830	20	3.41	0.0028

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	20	10.19	0.0003
iwt	1	20	11.60	0.0028

Interpretation

The HOLD=1,2 specification in the PARMS statement forces PROC MIXED to carry out the fixed effects analysis based on the specified set of variance components. The estimate of the common slope obtained based on the method of moments estimates of the variance components from the intra- and inter-block models provides the combined estimate of the common slope described above, as shown in the iwt row of the “Solution for Fixed Effects” table. The analysis provides no standard errors for the variance components since their values have been fixed as shown in the “Covariance Parameter Estimates” table. The combined estimate of the common slope using the REML estimates of the variance components from PROC MIXED is

$\hat{\beta}_c = 0.002780$ and $\hat{\sigma}_{\hat{\beta}_c} = 0.0008423$, the same as the estimate and estimated standard error of the slope in the “Solution for Fixed Effects” table in Output 7.3. Thus, the mixed model approach to the analysis of this model extracts more information about the common slope than a non-mixed models approach that provides only the intra-block analysis.

7.4 Example: One-Way Treatment Structure in an Incomplete Block Design Structure—Time to Boil Water

A study was conducted to determine the relationship between the amount of water put into a container and the time in seconds it takes for the water to boil. An electric cook range was used as the heat source, where the front left-hand burner was used for each of the runs. It was also of interest to determine if the type or shape of the container had an effect on the time to boil water, so three ceramic containers were used. The dimensions of the containers were as follows: container A had a top diameter of 6.5 inches, was 1.25 inches deep, and had a bottom diameter of 4.75 inches; container B had a top diameter of 5.5 inches, was 1.75 inches deep, and had a bottom diameter of 4.0 inches; the corresponding dimensions for container C were 5.75, 2.25, and 3.5 inches. Four amounts of water (2 oz., 4 oz., 6 oz., and 8 oz.) were used to help establish the relationship between the amount of water and the time it took for it to reach the boiling point. The process consisted of putting a specified amount of water into one of the containers, putting it on the range, starting the stopwatch, and stopping the watch when the water reached what was judged as complete boiling. A thermometer was not used in the process. The treatment structure consisted of three types of containers and four times. The treatment combinations were randomly ordered and one replication of the twelve treatments was obtained. Then a second randomization was carried out for replication two. The two replications for two of the treatment combinations were somewhat dissimilar, so a third block was constructed involving just those two treatments. This leads to two complete blocks and one incomplete block. The listing of the data is in Data Set 7.4, “Cooking Times,” in Appendix 2, “Data Sets.”

Model

It seems reasonable that there could be a linear relationship between the amount of water put into a container and the time required for the water to reach the boiling point (this assumption will be checked later). A model to describe these data is

$$T_{ij} = \alpha_i + \beta_i W_{ij} + b_j + e_{ij} \quad (7.6)$$

where

- T_{ij} denotes the time in seconds for the water to boil
- W_{ij} denotes the amount of water put into the container
- $i = 1, 2, 3$
- $j = 1, 2, 3$
- α_i denotes the intercept of the i^{th} container model
- β_i denotes the slope of the i^{th} container model
- $b_j \sim \text{iid } N(0, \sigma_b^2)$ denotes the effect of the j^{th} block
- $e_{ij} \sim \text{iid } N(0, \sigma_e^2)$ denotes the experimental unit or run-to-run error
- b_j and e_{ij} are independent random variables

7.4.1 Step 1: Fit the Model to Test the Equal Slopes Hypothesis

Program

A plot of the data indicates that there is a relationship between the amount of water and the amount of time for the water to reach the boiling point (Figure 7.7). The first step in this analysis is to determine if the slopes are equal. The following code fits a model similar to the one for model (7.4), including both amount and amount \times container in the model.

```
proc mixed data=long;
  class container block;
  model time=container amount container*amount /
    solution outp=pred ddfm=kr;
  random block;
run;
```

Results

The results for fitting the model to test for equality of slopes are in Output 7.7. Since we are interested only in the hypothesis of equal slopes, only the estimates of the variance components and the Type 3 tests of fixed effects are included.

Output 7.7 Fit of the Model to Test Equality of Slopes

Covariance Parameter Estimates	
Cov Parm	Estimate
block	94.3235
Residual	16.2662

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	18	12.08	0.0005
amount	1	18	4464.08	<.0001
amount*container	2	17.9	35.09	<.0001

Interpretation

The *p*-value associated with the amount \times container effect from Output 7.7 is less than 0.0001. There is sufficient information to conclude that the slopes are not equal. The analysis continues, using the unequal slopes model.

7.4.2 Step 2: Fit the Unequal Slopes Model

Once it is decided that the slopes are not equal, the next step is to compare the treatments by comparing the regression lines or characteristics of the regression lines. The strategy here is to evaluate the regression lines at several amounts of water and to perform pairwise comparisons among the predicted values. The following PROC MIXED program fits model (7.6).

Program

```
proc mixed data=long;
  class container block;
  model time = container container*amount /
    noint solution outp=pred ddfm=kr;
  lsmeans container / at amount=2 diff;
  lsmeans container / at amount=4 diff;
  lsmeans container / at amount=6 diff;
  lsmeans container / at amount=8 diff;
  random block;
run;
```

The NOINT option in the MODEL statement prevents the inclusion of an intercept in the model. This yields estimates of the intercepts, and the solutions pertaining to the amount \times container effect are estimates of the slopes.

Results

The results are presented in Output 7.8, which includes estimates of the variance components, estimates of the intercepts and slopes for each model, tests of the fixed effects, least-squares means at four different amounts of water, and pairwise comparisons among the container mean times within each amount of water.

Output 7.8 Fit of the Model to Test Equality of Slopes

Covariance Parameter Estimates	
Cov Parm	Estimate
block	94.3235
Residual	16.2662

Solution for Fixed Effects						
Effect	container	Estimate	Standard Error	DF	t Value	Pr > t
container	A	41.5075	6.3752	2.92	6.51	0.0080
container	B	17.9980	6.6974	3.49	2.69	0.0634
container	C	31.5436	6.5501	3.23	4.82	0.0143
amount*container	A	20.3037	0.6104	18.1	33.26	<.0001

Solution for Fixed Effects						
Effect	container	Estimate	Standard Error	DF	t Value	Pr > t
amount*container	B	26.8463	0.6377	17.9	42.10	<.0001
amount*container	C	26.3032	0.6348	17.9	41.44	<.0001

Type 3 Tests of Fixed Effects					
Effect	Num DF	Den DF	F Value	Pr > F	
container	3	6.99	15.43	0.0018	
amount*container	3	18	1501.62	<.0001	

Least Squares Means							
Effect	container	amount	Estimate	Standard Error	DF	t Value	Pr > t
container	A	2.00	82.1149	5.9765	2.26	13.74	0.0032
container	B	2.00	71.6905	6.1926	2.57	11.58	0.0028
container	C	2.00	84.1500	6.0701	2.4	13.86	0.0024
container	A	4.00	122.72	5.8115	2.01	21.12	0.0022
container	B	4.00	125.38	5.9241	2.15	21.16	0.0015
container	C	4.00	136.76	5.8320	2.04	23.45	0.0016
container	A	6.00	163.33	5.9000	2.12	27.68	0.0009
container	B	6.00	179.08	5.9241	2.15	30.23	0.0007
container	C	6.00	189.36	5.8654	2.08	32.28	0.0008
container	A	8.00	203.94	6.2311	2.62	32.73	0.0002
container	B	8.00	232.77	6.1926	2.57	37.59	0.0001
container	C	8.00	241.97	6.1658	2.53	39.24	0.0001

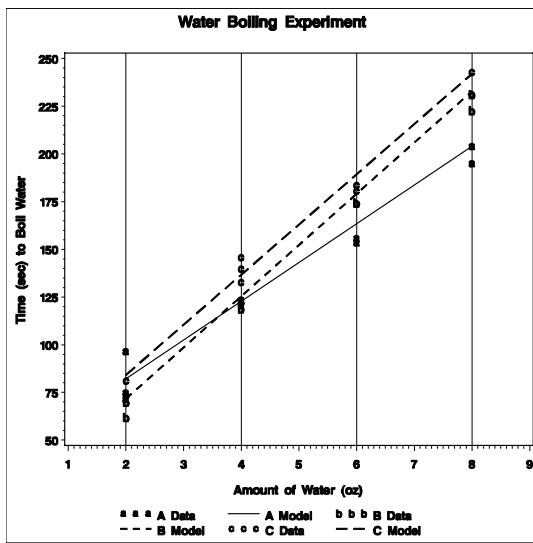
Differences of Least Squares Means								
Effect	container	_container	amount	Estimate	Standard Error	DF	t Value	Pr > t
container	A	B	2.00	10.4244	3.2626	18.1	3.20	0.0050
container	A	C	2.00	-2.0351	3.0699	17.9	-0.66	0.5158
container	B	C	2.00	-12.4595	3.3393	18	-3.73	0.0015
container	A	B	4.00	-2.6607	2.1535	18	-1.24	0.2325
container	A	C	4.00	-14.0341	2.0207	17.9	-6.95	<.0001
container	B	C	4.00	-11.3734	2.1789	18	-5.22	<.0001
container	A	B	6.00	-15.7458	2.2056	17.9	-7.14	<.0001
container	A	C	6.00	-26.0331	2.1761	17.9	-11.96	<.0001
container	B	C	6.00	-10.2873	2.1957	17.9	-4.69	0.0002
container	A	B	8.00	-28.8309	3.3654	17.9	-8.57	<.0001

Differences of Least Squares Means								
Effect	container	_container	amount	Estimate	Standard Error	DF	t Value	Pr > t
container	A	C	8.00	-38.0320	3.3736	17.9	-11.27	<.0001
container	B	C	8.00	-9.2011	3.3722	17.9	-2.73	0.0138

Interpretation

The variance component estimates in the “Covariance Parameter Estimates” table indicate that there was more block-to-block variation than within-block variation. The “Solution for the Fixed Effects” table provides the estimates of the slopes and intercepts. Remember that the intercepts represent predictions of the amount of time to boil the water when 0 oz. of water are used, an uninteresting prediction. The F statistics in the “Type 3 Tests of Fixed Effects” table test the hypotheses that all intercepts are equal to zero (container) and that all slopes are equal to zero (amount \times container). Neither of these hypotheses is of interest at this point in the analysis. The “Least Squares Means” table displays predicted mean times to boiling for each of the container models at 2, 4, 6, and 8 oz. of water. The predicted means and the data are plotted in Figure 7.7; vertical lines occur at the four levels of water. The “Differences of Least Squares Means” table shows pairwise comparisons among the container models at the four amounts of water. Without adjustment for multiplicity, there are no significant differences between A and C at 2 oz. and between A and B at 4 oz. A multiple comparison procedure, such as Tukey’s method, should be used for those comparisons within a level of amount of water and possibly for making comparisons across the four levels of water.

Figure 7.7 Plot of Data and LS-Means (Time (sec) to Boil Water) as a Function of Water Amount (oz)



7.4.3 Step 3: Test for Lack of Fit of the Simple Linear Regression Model

The last stage in this analysis (but the first step in the process) is to carry out a test for lack of fit of the simple linear regression model to the data for each of the containers. A test for lack of fit cannot be carried out for all data sets, but when some or all of the covariates have more than one observation, a model-free estimate of the variance can be obtained and the test for lack of fit can

be computed. This can be accomplished by using the amount effect as both a class and a continuous variable in the same model. The following code uses a DATA step to generate the variable CLASS_AMT to be equal to the AMOUNT variable.

Program

```
data linear;
  set long;
  class_amt = amount;
run;
proc mixed data=linear;
  class container class_amt block;
  model time=container amount container*amount
    container*class_amt / ddfm=kr;
  random block;
run;
```

Results

Selected results are shown in Output 7.9.

Output 7.9 Results to Test for Lack of Fit of the Simple Linear Regression Models

Class Level Information		
Class	Levels	Values
container	3	A B C
class_amt	4	2 4 6 8
block	3	1 2 3

Covariance Parameter Estimates	
Cov Parm	Estimate
block	85.1868
Residual	15.5052

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	11.9	13.52	0.0009
amount	0	.	.	.
amount*container	0	.	.	.
container*class_amt	6	11.9	1.18	0.3791

Interpretation

The “Class Level Information” table is included to show that the variable CLASS_AMT has four values. The model is constructed with terms for the intercepts (CONTAINER), terms for the slopes (AMOUNT \times CONTAINER), and terms for the container by amount cell means (CONTAINER \times CLASS_AMT). The “Type 3 Tests of Fixed Effects” table provides information about the question “Given all of the other terms are in the model, do I need this term, or is there sufficient variability remaining that this term is important in the model?” The question asked by the *F*-statistic for the CONTAINER \times CLASS_AMT effect is this: given that

variation about the intercepts and the slopes has been accounted for by the model, is there sufficient information for the cell means to still be important? The *F*-statistic for this effect provides the lack-of-fit test for this model. The degrees of freedom for lack of fit are computed as the number of cells with data (12) minus the number of parameters in the model (3 models each with an intercept and a slope, 6), leaving 6 numerator degrees of freedom for the lack-of-fit test. The *p*-value for this test is 0.3791, indicating that there is insufficient information to conclude that the simple linear regression models do not fit the data for each of the treatments.

7.4.4 Step 4: Testing the Equality of the Models at a Preselected Value of the Covariate

The design of the experiment used four amounts of water, and it might be of interest to construct a test of the hypothesis that the three container models are responding equally at selected values of amount of water. This can be accomplished by using CONTRAST statements, as in the following code.

Program

```
proc mixed data=linear;
  class container block;
  model time=container container*amount /ddfm=kr NOINT solution;
  random block;
  contrast 'Models = at 2oz'
    container 1 -1 0 container*amount 2 -2 0,
    container 1 0 -1 container*amount 2 0 -2;
run;
```

The CONTRAST statement tests the equality of container A at 2 oz. to container B at 2 oz. and the equality of container A at 2 oz. to container C at 2 oz. in a simultaneous test with two degrees of freedom.

Results

The results are in Output 7.10.

Output 7.10 Results for Comparing the Container Models at 2 oz. Amounts of Water

Covariance Parameter Estimates	
Cov Parm	Estimate
block	94.3235
Residual	16.2662

Solution for Fixed Effects						
Effect	container	Estimate	Standard Error	DF	t Value	Pr > t
container	A	41.5075	6.3752	2.92	6.51	0.0080
container	B	17.9980	6.6974	3.49	2.69	0.0634
container	C	31.5436	6.5501	3.23	4.82	0.0143
amount*container	A	20.3037	0.6104	18.1	33.26	<.0001

Solution for Fixed Effects						
Effect	container	Estimate	Standard Error	DF	t Value	Pr > t
amount*container	B	26.8463	0.6377	17.9	42.10	<.0001
amount*container	C	26.3032	0.6348	17.9	41.44	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	3	6.99	15.43	0.0018
amount*container	3	18	1501.62	<.0001

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
Models = at 2oz	2	18	7.82	0.0036

Interpretation

The “Contrasts” table in Output 7.10 provides the F value for testing the equality of the three regression models at 2 oz. of water—that is, the hypothesis $H_0: a_A + 2\beta_A = a_B + 2\beta_B = a_C + 2\beta_C$ versus H_a : (not H_0). The p -value of 0.0036 indicates that when 2 oz. of water is used, the models are not likely to be equal.

The comparisons of the models at a specific value of the covariate can also be done by shifting the origin of the covariate to that value and then performing an analysis of covariance. If you want to test the equality of the regression models at $x = 6$, then compute a new variable, $x^* = x - 6$, and carry out the analysis of covariance using x^* as the covariate. The F statistic corresponding to container is a test that the container models are equal at $x = 6$. The following code demonstrates this process by providing comparisons of the container models at 2, 4, 6, and 8 oz. of water.

Program

```

data test; set long;
amt_2 = amount-2;
amt_4 = amount-4;
amt_6 = amount-6;
amt_8 = amount-8;
run;

title2 "Compare the lines at 2 oz.";
proc mixed data=test;
class container block;
model time=container container*amt_2 /
solution outp=pred ddfm=kr;
random block;
run;
title2 "Compare the lines at 4 oz.";
proc mixed data=test;
class container block;

```

```

model time = container container*amt_4 /
            solution outp=pred ddfm=kr;
      random block;
run;
title2 "Compare the lines at 6 oz.";
proc mixed data=test;
  class container block;
  model time = container container*amt_6 /
            solution outp=pred ddfm=kr;
      random block;
run;
title2 "Compare the lines at 8 oz.";
proc mixed data=test;
  class container block;
  model time = container container*amt_8 /
            solution outp=pred ddfm=kr;
      random block;
run;

```

The DATA step computes the variables Amt_2, Amt_4, Amt_6, and Amt_8 by shifting the value of the AMOUNT variable. The DATA step is followed by four calls to PROC MIXED to produce the respective comparisons. The results are shown in Output 7.11.

Results

Output 7.11 Results for Comparing the Container Models at Four Amounts of Water

Covariance Parameter Estimates	
Cov Parm	Estimate
block	94.3235
Residual	16.2662

Compare the lines at 2 oz.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	18	7.82	0.0036
amt_2*container	3	18	1501.62	<.0001

Compare the lines at 4 oz.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	18	26.48	<.0001
amt_4*container	3	18	1501.62	<.0001

Compare the lines at 6 oz.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	17.9	72.49	<.0001
amt_6*container	3	18	1501.62	<.0001

Compare the lines at 8 oz.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
container	2	17.9	69.30	<.0001
amt_8*container	3	18	1501.62	<.0001

Interpretation

Only one “Covariance Parameter Estimates” table is presented; they are identical for the four models. Also, the *F*-statistics for the equal slopes hypothesis are identical for each of the analyses. The analyses differ only in the *F*-statistic for the test of equality of the intercepts. Using Amt_6 as the covariate, the intercept of the model has been moved to an amount of 6 oz., and thus the test of the equality of the intercepts is actually a test of the equality of the models at 6 oz. The result from the “Contrasts” table in Output 7.10 is also given in the “Type 3 Tests of Fixed Effects” table used to compare the models at 2 oz. of water.

7.5 Example: One-Way Treatment Structure in a Balanced Incomplete Block Design Structure

The data in Data Set 7.5, “Balanced Incomplete Block,” in Appendix 2, “Data Sets,” are from four treatments in a one-way treatment structure in a balanced incomplete block design structure with blocks of size three. The objectives of this example are to provide the methodology to work with the unequal slopes model and to demonstrate the use of PROC MIXED with a balanced incomplete block design structure. The response variable is *Y* and the covariate is *X*.

7.5.1 Step 1: Fit the Unequal Slopes Model

Model

A model that describes the data is

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + b_j + e_{ij}, \quad (i,j) \in B \quad (7.7)$$

where

α_i denotes the intercept of the i^{th} diet model

β_i denotes the slope of the i^{th} diet model

$b_j \sim iid N(0, \sigma_b^2)$ denotes the effect of the j^{th} block

$e_{ij} \sim iid N(0, \sigma_e^2)$ denotes the experimental unit error

b_j and e_{ij} are independent random variables

B is the index set of observed treatment-block combinations (i,j) or

$$B = [(1,1), (2,1), (3,1), (1,2), (2,2), (4,2), (1,3), (3,3), (4,3), (2,4), (3,4), (4,4), (1,5), (2,5), (3,5), (1,6), (2,6), (4,6), (1,7), (3,7), (4,7), (2,8), (3,8), (4,8)]$$

Program

The following program fits the unequal slopes model (7.7) to the data:

```
proc mixed data=bib;
  class blk trt;
  model y=trt x*trt/solution ddfm=kr;
  random blk;
run;
```

The NOINT option is not used in this MODEL statement. Thus, the estimates of the intercepts satisfy the set-to-zero restrictions. The DDFM=KR option is included to provide approximate degrees of freedom for the denominators of each test statistic and to adjust the estimated standard errors. The approximate degrees of freedom are needed because the estimates of the slopes and intercepts are combined inter- and intra-block estimates.

Results

The PROC MIXED results are in Output 7.12.

Output 7.12 Fit of the Unequal Slope Model and Test of the Slopes-Equal-to-Zero Hypothesis

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	18.2495
Residual	1.2004

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		22.3679	3.1260	14.2	7.16	<.0001
trt	1	4.4295	3.4043	9.51	1.30	0.2238
trt	2	-0.4374	2.9584	9.39	-0.15	0.8856
trt	3	6.2786	3.3244	9.56	1.89	0.0896
trt	4	0
x*trt	1	0.2188	0.06451	9.47	3.39	0.0074
x*trt	2	0.4959	0.05509	9.35	9.00	<.0001
x*trt	3	0.2634	0.05767	9.51	4.57	0.0012
x*trt	4	0.4425	0.08794	9.45	5.03	0.0006

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	9.32	4.43	0.0343
x*trt	4	9.38	31.75	<.0001

Interpretation

The set-to-zero restrictions are demonstrated by reviewing the “Solution for Fixed Effects” table in Output 7.12: the estimate corresponding to trt 4 is 0. The Type 3 F -statistic for the $x \times$ trt effect tests the hypotheses $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus H_a : (not H_0), and indicates there is evidence to reject the null hypothesis ($p < 0.0001$). (If the covariate X is not in the model, the Type 3 F for the $x \times$ trt effect tests the slopes-equal-to-zero hypothesis. If x is in the model along with $x \times$ trt, the Type 3 F for the $x \times$ trt effect tests the equal slopes hypothesis.)

7.5.2 Step 2: Test the Equal Slopes Hypothesis

Program

The following program fits the unequal slopes model to the data, but includes x as well as $x \times$ trt so that the Type 3 F for the $x \times$ trt effect tests

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta \text{ versus } H_a: (\text{not } H_0)$$

where β is not specified.

```
proc mixed data=bib;
  class blk trt;
  model y = trt x x*trt / solution ddfm=kr;
  estimate 'b1-b2' x*trt 1 -1 0 0;
  estimate 'b1-b3' x*trt 1 0 -1 0;
  estimate 'b1-b4' x*trt 1 0 0 -1;
  estimate 'b2-b3' x*trt 0 1 -1 0;
  estimate 'b2-b4' x*trt 0 1 0 -1;
  estimate 'b3-b4' x*trt 0 0 1 -1;
  random blk;
run;
```

The results are in Output 7.13.

Results

Output 7.13 Fit of the Unequal Slopes Model and Test of the Equal Slopes Hypothesis: ESTIMATE Statements Provide Pairwise Comparison among Slopes

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	18.2495
Residual	1.2004

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	9.32	4.43	0.0343
x	1	9.52	97.32	<.0001
x*trt	3	9.34	5.05	0.0242

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
b1-b2	-0.2771	0.07560	9.25	-3.67	0.0050
b1-b3	-0.04459	0.07508	9.27	-0.59	0.5668
b1-b4	-0.2238	0.1072	9.46	-2.09	0.0649
b2-b3	0.2326	0.07814	9.39	2.98	0.0149
b2-b4	0.05338	0.09787	9.35	0.55	0.5982
b3-b4	-0.1792	0.1173	9.6	-1.53	0.1589

Interpretation

The Type 3 *F*-statistic for the *x* × *trt* effect indicates that there is evidence to reject the equal slopes hypothesis (*p* = 0.0242). The ESTIMATE statements carry out multiple comparisons on the slopes. Based on these results (and based on the slope estimates in Output 7.12), treatments 1 and 3 group together and treatments 2 and 4 group together.

A new variable, GRP, is constructed to denote the two sets of treatments for which the slopes are equal. The value of GRP is determined as GRP=13 for treatments 1 and 3 and GRP=24 for treatments 2 and 4. The slopes are different between groups and homogeneous within groups. The grouping variable enables you to take advantage of a simpler model with only two (instead of four) slopes. This leads to fewer comparisons among the treatments, since you compare treatments with a common slope at only one value of *X* whereas you compare treatments with unequal slopes at a minimum of three values of *X*.

7.5.3 Step 3: Fit Model with Unequal Slopes for Each Level of GRP

The following program fits a model with unequal slopes for the two groups. Both trt and grp are included in the CLASS statement, but only trt is used in the MODEL statement to denote the four possible intercepts and x * grp is used to denote the two different slopes.

Program

```

data bib2; set bib;
  grp = 13;
  if trt=2 or trt=4 then grp = 24;
run;

proc mixed data=bib2;
  class blk trt grp;
  model y = trt x*grp / solution ddfm=kr;
  random blk;
  estimate 'trt1 at xx%= 10.0' intercept 1 trt 1 0 0 0 x*grp 10 0;
  estimate 'trt1 at xx%= 40.0' intercept 1 trt 1 0 0 0 x*grp 40 0;
  estimate 'trt1 at 25%= 17.0' intercept 1 trt 1 0 0 0 x*grp 17 0;
  estimate 'trt1 at 50%= 28.5' intercept 1 trt 1 0 0 0 x*grp 28.5 0;
  estimate 'trt1 at 75%= 37.0' intercept 1 trt 1 0 0 0 x*grp 37 0;
  estimate 'trt1 at mean=26.0' intercept 1 trt 1 0 0 0 x*grp 26 0;

  estimate 'trt3 at xx%= 10.0' intercept 1 trt 0 0 1 0 x*grp 10 0;
  estimate 'trt3 at xx%= 40.0' intercept 1 trt 0 0 1 0 x*grp 40 0;
  estimate 'trt3 at 25%= 17.0' intercept 1 trt 0 0 1 0 x*grp 17 0;
  estimate 'trt3 at 50%= 28.5' intercept 1 trt 0 0 1 0 x*grp 28.5 0;
  estimate 'trt3 at 75%= 37.0' intercept 1 trt 0 0 1 0 x*grp 37 0;
  estimate 'trt3 at mean=26.0' intercept 1 trt 0 0 1 0 x*grp 26 0;

  estimate 'trt2 at xx%= 10.0' intercept 1 trt 0 1 0 0 x*grp 0 10;
  estimate 'trt2 at xx%= 40.0' intercept 1 trt 0 1 0 0 x*grp 0 40;
  estimate 'trt2 at 25%= 17.0' intercept 1 trt 0 1 0 0 x*grp 0 17;
  estimate 'trt2 at 50%= 28.5' intercept 1 trt 0 1 0 0 x*grp 0 28.5;
  estimate 'trt2 at 75%= 37.0' intercept 1 trt 0 1 0 0 x*grp 0 37;
  estimate 'trt2 at mean=26.0' intercept 1 trt 0 1 0 0 x*grp 0 26;

  estimate 'trt4 at xx%= 10.0' intercept 1 trt 0 0 0 1 x*grp 0 10;
  estimate 'trt4 at xx%= 40.0' intercept 1 trt 0 0 0 1 x*grp 0 40;
  estimate 'trt4 at 25%= 17.0' intercept 1 trt 0 0 0 1 x*grp 0 17;
  estimate 'trt4 at 50%= 28.5' intercept 1 trt 0 0 0 1 x*grp 0 28.5;
  estimate 'trt4 at 75%= 37.0' intercept 1 trt 0 0 0 1 x*grp 0 37;
  estimate 'trt4 at mean=26.0' intercept 1 trt 0 0 0 1 x*grp 0 26;

  ***comparisons of means at 25%, 75%, and 50%;
  estimate 't1-t2 75%=37' trt 1 -1 0 0 x*grp 37 -37 ;
  estimate 't1-t2 50%=28.5' trt 1 -1 0 0 x*grp 28.5 -28.5;
  estimate 't1-t2 25%=17' trt 1 -1 0 0 x*grp 17 -17 ;
  estimate 't1-t4 75%=37' trt 1 0 0 -1 x*grp 37 -37 ;
  estimate 't1-t4 50%=28.5' trt 1 0 0 -1 x*grp 28.5 -28.5;
  estimate 't1-t4 25%=17' trt 1 0 0 -1 x*grp 17 -17 ;
  estimate 't3-t2 75%=37' trt 0 -1 1 0 x*grp 37 -37 ;
  estimate 't3-t2 50%=28.5' trt 0 -1 1 0 x*grp 28.5 -28.5;
  estimate 't3-t2 25%=17' trt 0 -1 1 0 x*grp 17 -17 ;
  estimate 't3-t4 75%=37' trt 0 0 1 -1 x*grp 37 -37 ;
  estimate 't3-t4 50%=28.5' trt 0 0 1 -1 x*grp 28.5 -28.5;
  estimate 't3-t4 25%=17' trt 0 0 1 -1 x*grp 17 -17 ;

  ***comparison of LSMEANS at X=26***;
  estimate 't1-t2 at mean' trt 1 -1 0 0 x*grp 26 -26;
  estimate 't1-t3 at mean' trt 1 0 -1 0 ;
  estimate 't1-t4 at mean' trt 1 0 0 -1 x*grp 26 -26;

```

```

estimate 't2-t3 at mean' trt 0 1 -1 0 x*grp 26 -26;
estimate 't2-t4 at mean' trt 0 1 0 -1 ;
estimate 't3-t4 at mean' trt 0 0 1 -1 x*grp 26 -26;
lsmeans trt / diff e at mean;
lsmeans trt / diff e at x=13;
run;

```

Results

The results from all of the preceding statements are shown in Output 7.14. The interpretation will refer to those sections needed for each inquiry.

Output 7.14 Fit of the Model for Four Treatments with Unequal Slope for Two Groups of Treatments

Class Level Information		
Class	Levels	Values
blk	8	1 2 3 4 5 6 7 8
trt	4	1 2 3 4
grp	2	13 24

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	18.5255
Residual	1.0378

Solution for Fixed Effects							
Effect	trt	grp	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			20.9452	2.0674	16	10.13	<.0001
trt	1		5.3414	1.9911	11.4	2.68	0.0207
trt	2		1.1356	0.7159	11.2	1.59	0.1406
trt	3		8.1810	1.7819	11.4	4.59	0.0007
trt	4		0
x*grp		13	0.2395	0.04339	11.5	5.52	0.0002
x*grp		24	0.4892	0.04440	11.3	11.02	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	3	11.2	15.49	0.0003
x*grp	2	11.4	73.43	<.0001

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
trt1 at xx% = 10.0	28.6818	1.7888	11.6	16.03	<.0001
trt1 at xx% = 40.0	35.8674	1.6507	9.08	21.73	<.0001
trt1 at 25% = 17.0	30.3584	1.6691	9.4	18.19	<.0001
trt1 at 50% = 28.5	33.1129	1.5831	7.81	20.92	<.0001
trt1 at 75% = 37.0	35.1488	1.6186	8.48	21.72	<.0001
trt1 at mean=26.0	32.5141	1.5889	7.92	20.46	<.0001
trt3 at xx% = 10.0	31.5214	1.6911	9.82	18.64	<.0001
trt3 at xx% = 40.0	38.7070	1.7352	10.7	22.31	<.0001
trt3 at 25% = 17.0	33.1980	1.6099	8.3	20.62	<.0001
trt3 at 50% = 28.5	35.9525	1.5976	8.08	22.50	<.0001
trt3 at 75% = 37.0	37.9884	1.6862	9.75	22.53	<.0001
trt3 at mean=26.0	35.3537	1.5870	7.89	22.28	<.0001
trt2 at xx% = 10.0	26.9730	1.6779	9.61	16.08	<.0001
trt2 at xx% = 40.0	41.6499	1.7624	11.2	23.63	<.0001
trt2 at 25% = 17.0	30.3977	1.6018	8.16	18.98	<.0001
trt2 at 50% = 28.5	36.0238	1.6048	8.22	22.45	<.0001
trt2 at 75% = 37.0	40.1823	1.7080	10.2	23.53	<.0001
trt2 at mean=26.0	34.8007	1.5902	7.95	21.88	<.0001
trt4 at xx% = 10.0	25.8375	1.8151	12.2	14.24	<.0001
trt4 at xx% = 40.0	40.5144	1.6496	9.05	24.56	<.0001
trt4 at 25% = 17.0	29.2621	1.6862	9.77	17.35	<.0001
trt4 at 50% = 28.5	34.8882	1.5879	7.9	21.97	<.0001
trt4 at 75% = 37.0	39.0467	1.6183	8.46	24.13	<.0001
trt4 at mean=26.0	33.6652	1.5960	8.05	21.09	<.0001
t1-t2 75% = 37	-5.0334	0.9404	11.2	-5.35	0.0002
t1-t2 50% = 28.5	-2.9109	0.6702	11.1	-4.34	0.0011
t1-t2 25% = 17	-0.03920	0.8415	11.3	-0.05	0.9637
t1-t4 75% = 37	-3.8978	0.7197	11.1	-5.42	0.0002
t1-t4 50% = 28.5	-1.7753	0.6456	11.2	-2.75	0.0187
t1-t4 25% = 17	1.0964	1.0619	11.4	1.03	0.3233
t3-t2 75% = 37	-2.1938	1.0988	11.3	-2.00	0.0705
t3-t2 50% = 28.5	-0.07128	0.7364	11.2	-0.10	0.9246
t3-t2 25% = 17	2.8004	0.7000	11.1	4.00	0.0020
t3-t4 75% = 37	-1.0583	0.8544	11.2	-1.24	0.2409
t3-t4 50% = 28.5	1.0643	0.6315	11.1	1.69	0.1199
t3-t4 25% = 17	3.9360	0.8936	11.2	4.40	0.0010
t1-t2 at mean	-2.2866	0.6500	11.1	-3.52	0.0047

Estimates							
Label		Estimate		Standard Error	DF	t Value	Pr > t
t1-t3 at mean		-2.8396		0.6734	11.2	-4.22	0.0014
t1-t4 at mean		-1.1510		0.6988	11.2	-1.65	0.1272
t2-t3 at mean		-13.5379		2.9888	11.4	-4.53	0.0008
t2-t4 at mean		1.1356		0.7159	11.2	1.59	0.1406
t3-t4 at mean		1.6886		0.6346	11.1	2.66	0.0221

Coefficients for trt Least Squares Means At x=26						
Effect	trt	grp	Row1	Row2	Row3	Row4
Intercept			1	1	1	1
trt	1		1			
trt	2			1		
trt	3				1	
trt	4					1
x*grp		13	13	13	13	13
x*grp		24	13	13	13	13

Coefficients for trt Least Squares Means At x=13						
Effect	trt	grp	Row1	Row2	Row3	Row4
Intercept			1	1	1	1
trt	1		1			
trt	2			1		
trt	3				1	
trt	4					1
x*grp		13	6.5	6.5	6.5	6.5
x*grp		24	6.5	6.5	6.5	6.5

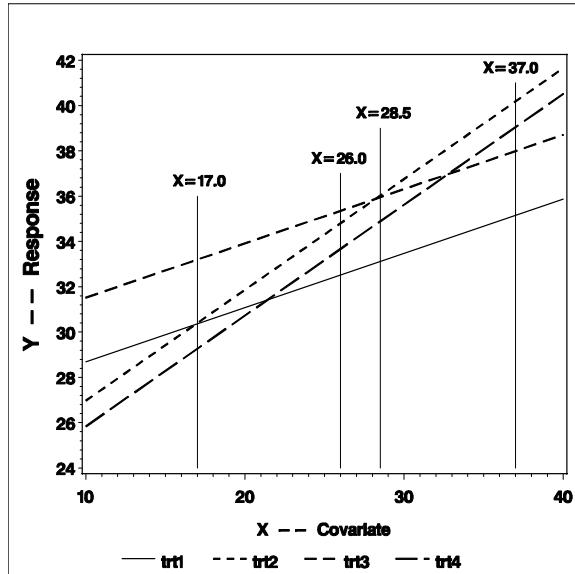
Least Squares Means							
Effect	trt	x	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	26.00	35.7604	1.8226	12.3	19.62	<.0001
trt	2	26.00	31.5545	1.7201	10.4	18.34	<.0001
trt	3	26.00	38.6000	1.7217	10.5	22.42	<.0001
trt	4	26.00	30.4189	1.8676	13.1	16.29	<.0001
trt	1	13.00	31.0235	1.8833	13.3	16.47	<.0001
trt	2	13.00	26.8176	1.7501	11	15.32	<.0001
trt	3	13.00	33.8631	1.7588	11.1	19.25	<.0001
trt	4	13.00	25.6820	1.9264	14.1	13.33	<.0001

Differences of Least Squares Means								
Effect	trt	_trt	x	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	26.00	4.2059	1.6961	11.4	2.48	0.0299
trt	1	3	26.00	-2.8396	0.6734	11.2	-4.22	0.0014
trt	1	4	26.00	5.3414	1.9911	11.4	2.68	0.0207
trt	2	3	26.00	-7.0455	1.4828	11.3	-4.75	0.0006
trt	2	4	26.00	1.1356	0.7159	11.2	1.59	0.1406
trt	3	4	26.00	8.1810	1.7819	11.4	4.59	0.0007
trt	1	2	13.00	4.2059	1.6961	11.4	2.48	0.0299
trt	1	3	13.00	-2.8396	0.6734	11.2	-4.22	0.0014
trt	1	4	13.00	5.3414	1.9911	11.4	2.68	0.0207
trt	2	3	13.00	-7.0455	1.4828	11.3	-4.75	0.0006
trt	2	4	13.00	1.1356	0.7159	11.2	1.59	0.1406
trt	3	4	13.00	8.1810	1.7819	11.4	4.59	0.0007

Interpretation

The “Class Level Information” table shows that the variable GRP can take on two values, 13 and 24, which indicate group membership of the treatments. Simplifying the model from four slopes to two slopes did not substantially increase the estimates of the variance components, as seen from the estimates of 18.52 and 1.04 in the “Covariance Parameter Estimates” table (compare to the estimates of 18.25 and 1.20 in Output 7.13). The “Solution for Fixed Effects” table contains estimates of the slopes and intercepts. The “Type 3 Tests of Fixed Effects” table displays the *F*-statistic for the test that the two slopes are equal to zero ($x \times \text{grp}$) and the trt *F*-statistic tests that the intercepts (models at $x = 0$) are equal (trt). Most likely these tests are not of interest. Of more interest is a comparison of the models at three or more values of x . The results in the “Estimates” table are adjusted means computed from the estimated regression models at $x = 17$ (25th percentile), $x = 26$ (the mean of the covariate values), $x = 28.5$ (the median of the covariate values), and $x = 37$ (the 75th percentile), as provided by ESTIMATE statements. The adjusted means were also computed at $x = 10$ and $x = 40$ to provide information for constructing graphs of the models; they do not correspond to any percentile of the distribution of X . The estimates labeled “ti-tj” are pairwise comparisons of the i^{th} and j^{th} treatments at the specified value of x . These comparisons are generally accomplished via the LSMEANS statement using, for example, the AT X=17.0 option. But in this case the LSMEANS statement does not provide the appropriate computations. The “Coefficients for trt Least Squares Means At x=26” (or 13) table shows the correct coefficients for the intercept and the levels of trt, but it multiplies 26/2=13 (6.5) times each slope without regard to the group to which the treatment belongs. Thus, the LSMEANS statement is not providing the appropriate information about the regression models evaluated at the average value of X . Treatments 1 and 3 and treatments 2 and 4 need to be compared at only one value of X , and they are compared at $x = 26$ (mean) only. By forming groups of treatments within which there are common slopes, the required number of comparisons is greatly reduced. The graph in Figure 7.8 displays the four regression lines indicating where the models are being evaluated and compared.

Figure 7.8 Graph of the Four Estimated Regression Lines;
Vertical Lines Drawn at Comparisons Values



7.6 Example: One-Way Treatment Structure in an Unbalanced Incomplete Block Design Structure

The data in Data Set 7.6, “Unbalanced Incomplete Block,” in Appendix 2, “Data Sets,” are from four treatments in a one-way treatment structure in an unbalanced incomplete block design structure with blocks of size two. This design illustrates one of the advantages of the mixed model analysis. The design is **not a connected block-treatment design**. But the mixed model analysis combines the intra-block and inter-block information about the fixed effects, thus extracting information about all of the treatments. The treatments can be compared using adjusted means such as those provided by the LSMEANS statement. The usual intra-block analysis, where blocks are considered as fixed effects, declares the comparisons 1 versus 3, 1 versus 4, 2 versus 3, and 2 versus 4 as nonestimable because the blocks and treatments are not connected. The response is Y , and the covariate is X . A model that can be used to describe these data is

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + b_j + e_{ij}, (i,j) \in B \quad (7.8)$$

where

α_i denotes the intercept of the model for the i^{th} diet

β_i denotes the slope of the model for the i^{th} diet

$b_j \sim iid N(0, \sigma_b^2)$ denotes the effect of the j^{th} block

$e_{ij} \sim iid N(0, \sigma_e^2)$ denotes the experimental unit error

b_j and e_{ij} are independent random variables

B is the index set of observed treatment-block combinations as described in Example 7.5

7.6.1 Step 1: Fit the Unequal Slopes Model and Test the Slopes-Equal-to-Zero Hypothesis

The following program fits the unequal slopes model to the data.

Program

```
proc mixed data=EX7_6_ubib;
  class blk trt;
  model y=trt x*trt/noint solution ddfm=kr;
  random blk;
run;
```

Results

The results are in Output 7.15, which include the estimates of the covariance parameters, the slopes and intercepts, and a test of the zero slopes hypothesis.

Output 7.15 Results for Fitting the Unequal Slopes Model and Test of the Slopes-Equal-to-Zero Hypothesis

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.02496
Residual	0.000856

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	0.5352	0.06806	11.7	7.86	<.0001
trt	2	0.6126	0.07055	13	8.68	<.0001
trt	3	0.3237	0.07977	15.6	4.06	0.0010
trt	4	0.3257	0.1096	13.4	2.97	0.0106
x*trt	1	-2.7296	0.6057	6.34	-4.51	0.0036
x*trt	2	-1.9179	0.7021	6.34	-2.73	0.0323
x*trt	3	-1.4062	0.9763	6.92	-1.44	0.1935
x*trt	4	-0.9902	1.3245	6.92	-0.75	0.4793

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	4	10.4	21.90	<.0001
x*trt	4	6.47	7.52	0.0135

Interpretation

The Type 3 F-test corresponding to $x \times \text{trt}$ in Output 7.15 tests the slopes-equal-to-zero hypothesis. There is evidence that the slopes are not all equal to zero ($p = 0.0135$).

7.6.2 Step 2: Test the Equal Slopes Hypothesis

Program

The following program is used to test the equal slopes hypothesis (x and x*trt are both in the model).

```
proc mixed data=EX7_6_ubib;
  class blk trt;
  model y=trt x x*trt/noint solution ddfm=kr;
  random blk;
run;
```

Results

The results are in Output 7.16, which include the estimates of the variance components, the estimates of the intercepts, the estimates of the slopes satisfying the set-to-zero restriction, and the test of the equal slopes hypothesis.

Output 7.16 Results from Fitting the Unequal Slopes Model and Test of the Equal Slopes Hypothesis

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.02496
Residual	0.000856

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	0.5352	0.06806	11.7	7.86	<.0001
trt	2	0.6126	0.07055	13	8.68	<.0001
trt	3	0.3237	0.07977	15.6	4.06	0.0010
trt	4	0.3257	0.1096	13.4	2.97	0.0106
x		-0.9902	1.3245	6.92	-0.75	0.4793
x*trt	1	-1.7394	1.4564	6.81	-1.19	0.2723
x*trt	2	-0.9277	1.4991	6.78	-0.62	0.5562
x*trt	3	-0.4160	1.0219	6.26	-0.41	0.6975
x*trt	4	0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	4	10.4	21.90	<.0001
x	1	6.96	9.50	0.0179
x*trt	3	6.48	0.61	0.6290

Interpretation

The results from the Type 3 F -statistic for the $x \times \text{trt}$ effect in Output 7.16 indicate that there is no evidence to believe the slopes are unequal ($p = 0.6290$). We conclude that a common slope model is adequate to describe the data.

7.6.3 Step 3: Fit a Common Slope Model

Program

The following program fits the common slope model to the data. The NOINT option is included in the MODEL statement so that the SOLUTION option provides estimates of the intercepts of the model. Without the NOINT option, the SOLUTION provides estimates of the intercepts that satisfy the set-to-zero restrictions. The LSMEANS statement is used to compare the treatments.

```
proc mixed data=EX7_6_ubib;
  class blk trt;
  model y=trt x /noint solution ddfm=kr;
  lsmeans trt / diff;
  random blk;
run;
```

Results

The results of fitting the common slope model to the data are given in Output 7.17, which includes the estimates of the model parameters, the least-squares means, and pairwise comparisons among the means.

Output 7.17 Results from Fitting the Common Slope Model and Results of the LS-Means to Compare Treatments at $X = 0.0449583$

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.02393
Residual	0.000793

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	0.5190	0.06529	10.9	7.95	<.0001
trt	2	0.6227	0.06585	11.3	9.46	<.0001
trt	3	0.3602	0.06680	11.9	5.39	0.0002
trt	4	0.4053	0.06939	13.4	5.84	<.0001
x		-2.1899	0.3970	9.47	-5.52	0.0003

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	4	13.2	32.04	<.0001
x	1	9.47	30.42	0.0003

Least Squares Means						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	0.4206	0.06447	10.5	6.52	<.0001
trt	2	0.5242	0.06427	10.3	8.16	<.0001
trt	3	0.2617	0.06420	10.3	4.08	0.0021
trt	4	0.3068	0.06475	10.6	4.74	0.0007

Differences of Least Squares Means							
Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	-0.1037	0.01649	9.02	-6.29	0.0001
trt	1	3	0.1589	0.09102	10.4	1.75	0.1104
trt	1	4	0.1138	0.09192	10.8	1.24	0.2421
trt	2	3	0.2625	0.09086	10.3	2.89	0.0156
trt	2	4	0.2174	0.09153	10.6	2.38	0.0376
trt	3	4	-0.04510	0.01806	9.09	-2.50	0.0338

Interpretation

The LSMEANS statement provides adjusted means or estimates from the common slope regression models at $x = 0.0449583$ (the mean of the covariate). It is interesting to note that the estimated standard errors of differences between treatments 1 and 2 and treatments 3 and 4 are much smaller than the estimated standard errors of the other differences. This difference in the estimated standard errors occurs because treatments 1 and 2 occur together in six blocks and treatments 3 and 4 occur together in six blocks, but all other pairs of treatments never occur together within a block (a characteristic of an unconnected design). Thus, the within-block comparisons of treatments have estimated standard errors that do not depend on the block variance component, and the between-block comparisons of treatments have estimated standard errors that do depend on the block variance component. Because the estimate of the block variance component is roughly 30 times larger than the estimate of the experimental unit variance component, those estimated standard errors involving block effects are larger than those not involving the block effects.

7.7 Example: Split-Plot Design with the Covariate Measured on the Large-Size Experimental Unit or Whole Plot

The data in Data Set 7.7, “Teaching Methods I,” in Appendix 2, “Data Sets,” are from a study designed to evaluate the effectiveness of three teaching methods (met). Four teachers were trained in each method. The teachers were randomly assigned to twelve classes of eight students (four females and four male students in each class). The experimental unit for the teaching method is the class of eight students. It was thought that years of experience in teaching might influence the effectiveness in delivery of the teaching method. Thus, the number of years of teaching experience for each teacher was used as a possible covariate. Part of the study was to determine if men and women responded differently to the teaching methods. Thus, gender (gen) is considered the treatment associated with experimental units at the student level. This experiment involves two sizes of experimental units: (1) the class of eight students to which the teaching method is applied and (2) the students who differ by gender. The covariate, years of teaching experience, is measured at the class level (the teacher level, to be exact, but teachers are confounded with classes). The treatment structure is a two-way factorial arrangement with three teaching methods and two genders. The design structure consists of two sizes of experimental units. Therefore, there are two variance components in the random effects part of the model. The responses are scores on a standardized test.

7.7.1 Step 1: Fit Unequal Slopes Model and Test Slopes-Equal-to-Zero Hypothesis

Model

A model involving the covariate to describe the test scores as a function of teaching method, class taught, gender of the students, and years of teaching experience is

$$Y_{ijkm} = \alpha_{ik} + \beta_{ik}x_{ij} + t_{j(i)} + e_{ijkm} \quad (7.9)$$

where

$$i = 1,2,3$$

$$j = 1,2,3,4$$

$$k = 1,2$$

$$m = 1,2,3,4$$

α_{ik} and β_{ik} are the intercept and slope associated with teaching method i and gender k

$t_{j(i)} \sim iid N(0, \sigma_t^2)$ is the random effect of the j^{th} class taught with the i^{th} method

$e_{ijkm} \sim iid N(0, \sigma_e^2)$ is the random effect associated with student m of gender k in class j taught by method i

The **fixed effects part** of the model is $\alpha_{ik} + \beta_{ik}x_{ij}$, the **random effects part** of the model is $t_{j(i)}$, and the residual part of the model is e_{ijkm} .

Program

The following program fits model (7.9) with separate intercepts (GEN*MET) and slopes (Y_EX*GEN*MET) for each gender-method combination.

```
title2 'Means model for intercepts and slopes';
proc mixed data=Ex_7_7sp;
  class teacher met gen;
  model score = gen*met y_ex*gen*met/solution noint ddfm=kr;
  random teacher(met);
run;
```

Results

The results are listed in Output 7.18, which include the estimates of the model parameters as well as tests of the all-slopes-equal-zero hypothesis.

Output 7.18 Fit of the Unequal Slopes Model and Test Whether Slopes Are Zero

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	1.7814
Residual	0.6405

Solution for Fixed Effects							
Effect	gen	met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	16.4231	2.3507	6.53	6.99	0.0003
met*gen	m	1	16.6538	2.3507	6.53	7.08	0.0003
met*gen	f	2	20.6993	2.0648	6.53	10.02	<.0001
met*gen	m	2	21.1351	2.0648	6.53	10.24	<.0001
met*gen	f	3	30.9821	1.8117	6.53	17.10	<.0001
met*gen	m	3	26.9345	1.8117	6.53	14.87	<.0001
y_ex*met*gen	f	1	-0.03205	0.1996	6.53	-0.16	0.8773
y_ex*met*gen	m	1	-0.03590	0.1996	6.53	-0.18	0.8627
y_ex*met*gen	f	2	0.09797	0.1620	6.53	0.60	0.5657
y_ex*met*gen	m	2	-0.09459	0.1620	6.53	-0.58	0.5788
y_ex*met*gen	f	3	-0.1518	0.1520	6.53	-1.00	0.3537
y_ex*met*gen	m	3	-0.1815	0.1520	6.53	-1.19	0.2740

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met*gen	6	15.6	71.52	<.0001
y_ex*met*gen	6	15.6	1.58	0.2174

The “Solution for Fixed Effects” table lists the estimates of the intercepts (e.g., met*gen f 1) and the slopes (e.g., y_ex*gen*met f 1). The Type 3 F-statistic for y_ex*gen*met tests the slopes-equal-to-zero hypothesis and fails to provide enough evidence to conclude nonzero slopes ($p = 0.2174$). For illustration purposes, we investigate the structure of the slopes further to see if by chance a simpler model involving the covariate may be appropriate.

7.7.2 Step 2: Fit Factorial Effects Model for Intercepts and Slopes

The intercepts and slopes are each expressed as factorial effects in the following model:

$$Y_{ijkm} = (\mu + \tau_i + \gamma_k + \delta_{ik}) + (\beta + \varphi_i + \theta_k + \rho_{ik})x_{ij} + t_{j(i)} + e_{ijkm} \quad (7.10)$$

$$i = 1, 2, 3, \quad j = 1, 2, 3, 4, \quad k = 1, 2, \quad m = 1, 2, 3, 4$$

where

$$\alpha_{ik} = (\mu + \tau_i + \gamma_k + \delta_{ik})$$

$$\beta_{ik} = (\beta + \varphi_i + \theta_k + \rho_{ik})$$

$$t_{j(i)} \sim iid N(0, \sigma_t^2)$$

$$e_{ijkm} \sim iid N(0, \sigma_e^2)$$

The additional parameters in model (7.10) compared to model (7.9) provide factorial effect representations for the slopes and intercepts. Thus, each slope or intercept is expressed through an overall mean, an effect due to teaching method, an effect due to gender, and an effect due to the interaction between teaching method and gender. Those parameters are as follows:

μ and β denote the overall means

τ_i and φ_i denote the effect of the i^{th} teaching method

γ_k and θ_k denote the k^{th} gender effect

δ_{ik} and ρ_{ik} denote the interaction effect between the i^{th} teaching method and the k^{th} gender

Program

The following program fits a model with unequal slopes, expressing intercepts and slopes as factorial effects:

```
title2 'Effects model for intercepts and slopes';
proc mixed data=Ex_7_7sp;
  class teacher met gen;
  model score = met gen gen*met
            y_ex y_ex*met y_ex*gen y_ex*gen*met/ddfm=kr;
  random teacher(met);
run;
```

Results

The analysis of variance results of the factorial tests for the slopes and intercepts are shown in Output 7.19, which also includes the estimates of the variance components.

Output 7.19 Fit of the Unequal Slopes Model Using Factorial Effects

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	1.7814
Residual	0.6405

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	6	10.09	0.0120
gen	1	78	5.30	0.0240
met*gen	2	78	10.30	0.0001
y_ex	1	6	0.46	0.5216
y_ex*met	2	6	0.33	0.7322
y_ex*gen	1	78	3.48	0.0660
y_ex*met*gen	2	78	2.24	0.1138

Interpretation

The Type 3 F -test associated with the $y_{ex} \times met \times gen$ effect is not significant ($p = 0.1138$). Thus, we removed the $y_{ex} \times met \times gen$ term from the model and refit the model with just a main effects partition of the slopes.

7.7.3 Step 3: Fit Model with Slopes Expressed as Main Effects of the Two Factors

Model

The model with slopes as a function of the main effects only is

$$Y_{ijkm} = (\mu + \tau_i + \gamma_k + \delta_{ik}) + (\beta + \varphi_i + \theta_k) x_{ij} + t_{j(i)} + e_{ijkm}$$

where

$$i = 1, 2, 3$$

$$j = 1, 2, 3, 4$$

$$k = 1, 2$$

$$m = 1, 2, 3, 4$$

$$\alpha_{ik} = (\mu + \tau_i + \gamma_k + \delta_{ik})$$

$$\beta_{ik} = (\beta + \varphi_i + \theta_k)$$

Program

The following program fits the model without the y_ex*met*gen term:

```
title2 'Effects model for intercepts and additive model for
slopes';
proc mixed data=Ex_7_7sp;
  class teacher met gen;
  model score = met gen gen*met y_ex y_ex*met y_ex*gen/ddfm=kr;
  random teacher(met);
run;
```

Results

The results are listed in Output 7.20, which includes the analysis of variance tests for the remaining terms in the model.

Output 7.20 Fit of the Model with Unequal Slopes Expressed as Main Effects

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	1.7789
Residual	0.6602

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	6	10.09	0.0120
gen	1	80	5.00	0.0281
met*gen	2	80	63.90	<.0001
y_ex	1	6	0.46	0.5216
y_ex*met	2	6	0.33	0.7322
y_ex*gen	1	80	4.20	0.0436

Interpretation

The Type 3 F-test for the $y_{ex} \times met$ effect is not significant. Thus, we remove the term from the model and refit a model with unequal slopes for each gender.

7.7.4 Step 4: Fit Model with Unequal Slopes for Each Gender

Model

The model with slopes as a function of the gender effects only is

$$Y_{ijkm} = (\mu + \tau_i + \gamma_k + \delta_{ik}) + (\beta + \theta_k) x_{ij} + t_{j(i)} + e_{ijkm}$$

where

$$i = 1, 2, 3$$

$$j = 1, 2, 3, 4$$

$$k = 1, 2$$

$$m = 1, 2, 3, 4$$

$$\alpha_{ik} = (\mu + \tau_i + \gamma_k + \delta_{ik})$$

$$\beta_{ik} = (\beta + \theta_k)$$

Program

The following program fits the model with unequal slopes for each gender:

```
title2 'two-way for intercepts and one-way for slopes';
proc mixed data=Ex_7_7sp;
  class teacher met gen;
  model score = met gen gen*met y_ex y_ex*gen/ddfm=kr;
  random teacher(met);
run;
```

Results

The results are in Output 7.21, which includes the analysis of variance for the remaining effects in the model.

Output 7.21 Fit of the Unequal Slopes Model for the Levels of Gender

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	1.4665
Residual	0.6602

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	8	77.68	<.0001
gen	1	80	5.00	0.0281
met*gen	2	80	63.90	<.0001
y_ex	1	8	0.75	0.4107
y_ex*gen	1	80	4.20	0.0436

Interpretation

The Type 3 F -statistic for the $y_{ex} \times \text{gen}$ effect tests the equality of the gender slopes and presents evidence (at the 5% significance level) that the two slopes are unequal ($p = 0.0436$). The simplified model to investigate the relationship between teaching methods and gender involving the covariate with unequal slopes for each gender and a means model for the intercepts is

$$Y_{ijkm} = \alpha_{ik} + \beta_k x_{ij} + t_{j(i)} + e_{ijkm} \quad (7.11)$$

where

$$\begin{aligned} i &= 1,2,3 \\ j &= 1,2,3,4 \\ k &= 1,2 \\ m &= 1,2,3,4 \\ t_{j(i)} &\sim iid N(0, \sigma_t^2) \\ e_{ijkm} &\sim iid N(0, \sigma_e^2) \end{aligned}$$

Program

The following program fits model (7.11) to the data:

```
title2 "Means model for intercepts and slopes for y_ex for each gender";
proc mixed data=Ex_7_7sp;
  class teacher met gen;
  model score =gen*met y_ex*gen/noint solution ddfm=kr;
  random teacher(met);

  lsmeans gen*met / diff at means;
  lsmeans gen*met / diff at y_ex=5;
  lsmeans gen*met / diff at y_ex=10;
  lsmeans gen*met / diff at y_ex=20;

  estimate 'f-m m=1 y_ex=5'      gen*met 1 -1 0 0 0 0 y_ex*gen 5 -5;
  estimate 'f-m m=1 y_ex=10'     gen*met 1 -1 0 0 0 0 y_ex*gen 10 -10;
  estimate 'f-m m=1 y_ex=20'     gen*met 1 -1 0 0 0 0 y_ex*gen 20 -20;
  estimate 'f-m m=2 y_ex=5'      gen*met 0 0 1 -1 0 0 y_ex*gen 5 -5;
  estimate 'f-m m=2 y_ex=10'     gen*met 0 0 1 -1 0 0 y_ex*gen 10 -10;
  estimate 'f-m m=2 y_ex=20'     gen*met 0 0 1 -1 0 0 y_ex*gen 20 -20;
  estimate 'f-m m=3 y_ex=5'      gen*met 0 0 0 0 1 -1 y_ex*gen 5 -5;
  estimate 'f-m m=3 y_ex=10'     gen*met 0 0 0 0 1 -1 y_ex*gen 10 -10;
  estimate 'f-m m=3 y_ex=20'     gen*met 0 0 0 0 1 -1 y_ex*gen 20 -20;

  estimate 'male m=1 y_ex=5'     gen*met 1 0 0 0 0 0 y_ex*gen 5 0;
  estimate 'male m=1 y_ex=10'    gen*met 1 0 0 0 0 0 y_ex*gen 10 0;
  estimate 'male m=1 y_ex=20'    gen*met 1 0 0 0 0 0 y_ex*gen 20 0;
  estimate 'male m=2 y_ex=5'     gen*met 0 0 1 0 0 0 y_ex*gen 5 0;
  estimate 'male m=2 y_ex=10'    gen*met 0 0 1 0 0 0 y_ex*gen 10 0;
  estimate 'male m=2 y_ex=20'    gen*met 0 0 1 0 0 0 y_ex*gen 20 0;
  estimate 'male m=3 y_ex=5'     gen*met 0 0 0 0 1 0 y_ex*gen 5 0;
  estimate 'male m=3 y_ex=10'    gen*met 0 0 0 0 1 0 y_ex*gen 10 0;
  estimate 'male m=3 y_ex=20'    gen*met 0 0 0 0 1 0 y_ex*gen 20 0;

  estimate 'female m=1 y_ex=5'   gen*met 0 1 0 0 0 0 y_ex*gen 0 5;
  estimate 'female m=1 y_ex=10'  gen*met 0 1 0 0 0 0 y_ex*gen 0 10;
  estimate 'female m=1 y_ex=20'  gen*met 0 1 0 0 0 0 y_ex*gen 0 20;
  estimate 'female m=2 y_ex=5'   gen*met 0 0 0 1 0 0 y_ex*gen 0 5;
  estimate 'female m=2 y_ex=10'  gen*met 0 0 0 1 0 0 y_ex*gen 0 10;
  estimate 'female m=2 y_ex=20'  gen*met 0 0 0 1 0 0 y_ex*gen 0 20;
  estimate 'female m=3 y_ex=5'   gen*met 0 0 0 0 0 1 y_ex*gen 0 5;
  estimate 'female m=3 y_ex=10'  gen*met 0 0 0 0 0 1 y_ex*gen 0 10;
  estimate 'female m=3 y_ex=20'  gen*met 0 0 0 0 0 1 y_ex*gen 0 20;
run;
```

ESTIMATE statements compare genders within each teaching method at 5, 10, and 20 years of teaching experience and provide adjusted means for combinations of gender and teaching method at 5, 10, and 20 years of teaching experience. The first LSMEANS statement compares teaching methods within each gender and compare all models at the average years of teaching experience, 11.41667. The second through fourth LSMEANS statements also provide comparisons at 5, 10, and 20 years of teaching experience for comparison purposes using the AT option.

Results

The estimates of the variance components and of the fixed effects are shown in Output 7.22, least-squares means results are given in Output 7.23, pairwise comparisons of least-squares means are provided in Output 7.24, and the results of the ESTIMATE statements can be found in Output 7.25.

Output 7.22 Fit of Model with Intercepts for the Method-by-Gender Combinations and Separate Gender Slopes

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	1.4665
Residual	0.6602

Solution for Fixed Effects							
Effect	gen	met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	16.4468	1.1860	8.87	13.87	<.0001
met*gen	m	1	17.5559	1.1860	8.87	14.80	<.0001
met*gen	f	2	22.2849	1.2427	8.87	17.93	<.0001
met*gen	m	2	21.3930	1.2427	8.87	17.22	<.0001
met*gen	f	3	29.6883	1.1674	8.87	25.43	<.0001
met*gen	m	3	26.2144	1.1674	8.87	22.46	<.0001
y_ex*gen	f		-0.03416	0.08883	8.87	-0.38	0.7096
y_ex*gen	m		-0.1161	0.08883	8.87	-1.31	0.2241

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met*gen	6	21	120.44	<.0001
y_ex*gen	2	18.7	2.39	0.1187

Output 7.23 LS-Means Comparing Method-by-Gender Adjusted Means at the Average Value of the Covariate, Y_Ex = 11.41667, and at Y_Ex = 5, 10, and 20

Least Squares Means									
Effect	gen	met	y_ex	Estimate	Standard Error	DF	t Value	Pr > t	
met*gen	f	1	11.42	16.0568	0.6388	8.87	25.13	<.0001	
met*gen	m	1	11.42	16.2307	0.6388	8.87	25.41	<.0001	
met*gen	f	2	11.42	21.8949	0.6408	8.87	34.17	<.0001	
met*gen	m	2	11.42	20.0677	0.6408	8.87	31.32	<.0001	
met*gen	f	3	11.42	29.2983	0.6397	8.87	45.80	<.0001	
met*gen	m	3	11.42	24.8891	0.6397	8.87	38.91	<.0001	
met*gen	f	1	5.00	16.2760	0.8462	8.87	19.23	<.0001	
met*gen	m	1	5.00	16.9755	0.8462	8.87	20.06	<.0001	
met*gen	f	2	5.00	22.1141	0.8914	8.87	24.81	<.0001	
met*gen	m	2	5.00	20.8126	0.8914	8.87	23.35	<.0001	
met*gen	f	3	5.00	29.5175	0.8318	8.87	35.48	<.0001	
met*gen	m	3	5.00	25.6340	0.8318	8.87	30.82	<.0001	
met*gen	f	1	10.00	16.1052	0.6482	8.87	24.84	<.0001	
met*gen	m	1	10.00	16.3951	0.6482	8.87	25.29	<.0001	
met*gen	f	2	10.00	21.9433	0.6629	8.87	33.10	<.0001	
met*gen	m	2	10.00	20.2322	0.6629	8.87	30.52	<.0001	
met*gen	f	3	10.00	29.3467	0.6448	8.87	45.51	<.0001	
met*gen	m	3	10.00	25.0536	0.6448	8.87	38.85	<.0001	
met*gen	f	1	20.00	15.7636	1.0060	8.87	15.67	<.0001	
met*gen	m	1	20.00	15.2343	1.0060	8.87	15.14	<.0001	
met*gen	f	2	20.00	21.6017	0.9555	8.87	22.61	<.0001	
met*gen	m	2	20.00	19.0713	0.9555	8.87	19.96	<.0001	
met*gen	f	3	20.00	29.0051	1.0233	8.87	28.35	<.0001	
met*gen	m	3	20.00	23.8928	1.0233	8.87	23.35	<.0001	

Output 7.24 Needed Comparison of the LS-Means for the Method-by-Gender Adjusted Means at the Average Value of the Covariate, Y_Ex= 11.41667, and at Y_Ex = 5, 10, and 20

Differences of Least Squares Means										
Effect	gen	met	_gen	_met	y_ex	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	m	1	11.42	-0.1738	0.2874	80	-0.60	0.5469
met*gen	f	1	f	2	11.42	-5.8381	0.9056	8.87	-6.45	0.0001
met*gen	f	1	f	3	11.42	-13.2415	0.9035	8.87	-14.66	<.0001

Differences of Least Squares Means										
Effect	gen	met	_gen	_met	y_ex	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	m	1	m	2	11.42	-3.8371	0.9056	8.87	-4.24	0.0023
met*gen	m	1	m	3	11.42	-8.6585	0.9035	8.87	-9.58	<.0001
met*gen	f	2	m	2	11.42	1.8272	0.2882	80	6.34	<.0001
met*gen	f	2	f	3	11.42	-7.4033	0.9076	8.87	-8.16	<.0001
met*gen	m	2	m	3	11.42	-4.8214	0.9076	8.87	-5.31	0.0005
met*gen	f	3	m	3	11.42	4.4091	0.2878	80	15.32	<.0001
met*gen	f	1	m	1	5.00	-0.6995	0.3807	80	-1.84	0.0698
met*gen	f	1	f	2	5.00	-5.8381	0.9056	8.87	-6.45	0.0001
met*gen	f	1	f	3	5.00	-13.2415	0.9035	8.87	-14.66	<.0001
met*gen	m	1	m	2	5.00	-3.8371	0.9056	8.87	-4.24	0.0023
met*gen	m	1	m	3	5.00	-8.6585	0.9035	8.87	-9.58	<.0001
met*gen	f	2	m	2	5.00	1.3015	0.4010	80	3.25	0.0017
met*gen	f	2	f	3	5.00	-7.4033	0.9076	8.87	-8.16	<.0001
met*gen	m	2	m	3	5.00	-4.8214	0.9076	8.87	-5.31	0.0005
met*gen	f	3	m	3	5.00	3.8835	0.3742	80	10.38	<.0001
met*gen	f	1	m	1	10.00	-0.2899	0.2916	80	-0.99	0.3231
met*gen	f	1	f	2	10.00	-5.8381	0.9056	8.87	-6.45	0.0001
met*gen	f	1	f	3	10.00	-13.2415	0.9035	8.87	-14.66	<.0001
met*gen	m	1	m	2	10.00	-3.8371	0.9056	8.87	-4.24	0.0023
met*gen	m	1	m	3	10.00	-8.6585	0.9035	8.87	-9.58	<.0001
met*gen	f	2	m	2	10.00	1.7112	0.2982	80	5.74	<.0001
met*gen	f	2	f	3	10.00	-7.4033	0.9076	8.87	-8.16	<.0001
met*gen	m	2	m	3	10.00	-4.8214	0.9076	8.87	-5.31	0.0005
met*gen	f	3	m	3	10.00	4.2931	0.2900	80	14.80	<.0001
met*gen	f	1	m	1	20.00	0.5293	0.4525	80	1.17	0.2456
met*gen	f	1	f	2	20.00	-5.8381	0.9056	8.87	-6.45	0.0001
met*gen	f	1	f	3	20.00	-13.2415	0.9035	8.87	-14.66	<.0001
met*gen	m	1	m	2	20.00	-3.8371	0.9056	8.87	-4.24	0.0023
met*gen	m	1	m	3	20.00	-8.6585	0.9035	8.87	-9.58	<.0001
met*gen	f	2	m	2	20.00	2.5304	0.4298	80	5.89	<.0001
met*gen	f	2	f	3	20.00	-7.4033	0.9076	8.87	-8.16	<.0001
met*gen	m	2	m	3	20.00	-4.8214	0.9076	8.87	-5.31	0.0005
met*gen	f	3	m	3	20.00	5.1123	0.4603	80	11.11	<.0001

Output 7.25 Results of the ESTIMATE Statements to Compute and Compare Adjusted Means across Genders at 5, 10, and 20 Years of Experience

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
f-m m=1 y_ex=5	-0.6995	0.3807	80	-1.84	0.0698
f-m m=1 y_ex=10	-0.2899	0.2916	80	-0.99	0.3231
f-m m=1 y_ex=20	0.5293	0.4525	80	1.17	0.2456
f-m m=2 y_ex=5	1.3015	0.4010	80	3.25	0.0017
f-m m=2 y_ex=10	1.7112	0.2982	80	5.74	<.0001
f-m m=2 y_ex=20	2.5304	0.4298	80	5.89	<.0001
f-m m=3 y_ex=5	3.8835	0.3742	80	10.38	<.0001
f-m m=3 y_ex=10	4.2931	0.2900	80	14.80	<.0001
f-m m=3 y_ex=20	5.1123	0.4603	80	11.11	<.0001
male m=1 y_ex=5	16.2760	0.8462	8.87	19.23	<.0001
male m=1 y_ex=10	16.1052	0.6482	8.87	24.84	<.0001
male m=1 y_ex=20	15.7636	1.0060	8.87	15.67	<.0001
male m=2 y_ex=5	22.1141	0.8914	8.87	24.81	<.0001
male m=2 y_ex=10	21.9433	0.6629	8.87	33.10	<.0001
male m=2 y_ex=20	21.6017	0.9555	8.87	22.61	<.0001
male m=3 y_ex=5	29.5175	0.8318	8.87	35.48	<.0001
male m=3 y_ex=10	29.3467	0.6448	8.87	45.51	<.0001
male m=3 y_ex=20	29.0051	1.0233	8.87	28.35	<.0001
female m=1 y_ex=5	16.9755	0.8462	8.87	20.06	<.0001
female m=1 y_ex=10	16.3951	0.6482	8.87	25.29	<.0001
female m=1 y_ex=20	15.2343	1.0060	8.87	15.14	<.0001
female m=2 y_ex=5	20.8126	0.8914	8.87	23.35	<.0001
female m=2 y_ex=10	20.2322	0.6629	8.87	30.52	<.0001
female m=2 y_ex=20	19.0713	0.9555	8.87	19.96	<.0001
female m=3 y_ex=5	25.6340	0.8318	8.87	30.82	<.0001
female m=3 y_ex=10	25.0536	0.6448	8.87	38.85	<.0001
female m=3 y_ex=20	23.8928	1.0233	8.87	23.35	<.0001

Interpretation

The Type 3 F-statistics in Output 7.22 test that all intercepts are equal to zero and that all slopes are equal to zero. Most likely these tests are not of much interest, but this form of the model is most convenient for constructing ESTIMATE statements because the resulting model is nonsingular and has parameters as described in model (7.11). But it is very important to realize that the p-value associated with the F-test for the slopes is 0.1187, and one might be tempted to

exclude the covariate from the analysis. This is also true for the initial model in Output 7.17. But there is information about the slopes from the within-teacher comparisons and from the between-teacher comparisons. When within-teacher comparisons are made, the information about the covariate becomes important. The moral of the story is to always start with the full factorial representation of the slope part of the model before doing any simplification.

The LSMEANS results are in Output 7.23, the comparisons of the LSMEANS are in Output 7.24, and the results of the ESTIMATE statements are in Output 7.25. For teaching experience of 5, 10, and 20 years, only gender comparisons within a teaching method are presented. The results from the LSMEANS and ESTIMATE statements are identical, but both are presented to help understand the computation of the LSMEANS.

The importance of using the mixed models approach for this example is that the classes of eight students are blocks for gender and (like the RCB in Section 7.2) there is information about the slopes from within blocks and between blocks. The mixed model equations combine the information from the various parts of the model into combined estimators of the slopes. Using the combined estimators should generally provide estimators with smaller variances than those based on intra-block information alone.

7.8 Example: Split-Plot Design with the Covariate Measured on the Small-Size Experimental Unit or Subplot

This example is similar to the example in Section 7.7, except that the covariate is measured at the student level instead of at the teacher level. The data in Data Set 7.8, “Teaching Methods II,” in Appendix 2, “Data Sets,” are from a study designed by a researcher to evaluate the effectiveness of three teaching methods. Four teachers were trained in each method. The twelve teachers were randomly assigned to twelve classes of eight students consisting of four males and four females.

The **experimental unit** for the teaching method is a class of eight students, the entity to which the teaching methods were randomly assigned. The researcher was interested in the possible difference in gender response to the three teaching methods. Thus, gender is the treatment for the small-size experimental unit, the student.

The researcher hypothesized that the IQ of the individual might influence the student’s ability to respond to the teaching method. The IQ of each of the 96 students was determined and investigated as a possible covariate.

This experiment again involves two sizes of experimental units: (1) the class of eight students to which the teaching method is applied and (2) the student nested within class and gender. The covariate, IQ, is measured at the student level, the small-size experimental unit.

The treatment structure consists of the two-way set of treatment combinations, three teaching methods by two genders. Because the design structure consists of two sizes of experimental units (the class and the student within class), there are two variance components in the random effects part of the model. The responses are scores on a standardized test (SCORE).

Model

A model to describe the test scores involving the covariate is

$$Y_{ijkm} = \alpha_{ik} + \beta_{ik}w_{ijkm} + t_{j(i)} + e_{ijkm} \quad (7.12)$$

where

$$i = 1, 2, 3$$

$$j = 1, 2, 3, 4$$

$$k = 1, 2$$

$$m = 1, 2, 3, 4$$

$$t_{j(i)} \sim iid N(0, \sigma_t^2)$$

$$e_{ijkm} \sim iid N(0, \sigma_e^2)$$

The intercept and slope of the model for the i^{th} teaching method and the k^{th} gender are α_{ik} and β_{ik} , respectively. The fixed effects part of the model is $\alpha_{ik} + \beta_{ik}w_{ijkm}$, where w_{ijkm} denotes the IQ of the m^{th} student of the k^{th} gender in the j^{th} class taught by the i^{th} teaching method, the random effects part of the model is $t_{j(i)}$, and the residual part of the model is e_{ijkm} .

7.8.1 Step 1: Fit Model with Factorial Effects for Both Intercepts and Slopes

Program

The following program fits model (7.12), where intercept and slopes are expressed as factorial effects:

```
title2 "Factorial effects for both intercepts and slopes";
proc mixed data=EX_7_8_spcv;
  class teacher met gen;
  model score = met gen gen*met iq iq*gen iq*met
    iq*gen*met / ddfm=kr;
  random teacher(met);
run;
```

Results

The results are shown in Output 7.26.

Output 7.26 Fit of the Unequal Slopes Model with Intercepts and Slopes as Factorial Effects

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	24.2621
Residual	18.1347

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	83.4	0.41	0.6662
gen	1	75.8	0.09	0.7693
met*gen	2	75.8	0.23	0.7971
iq	1	76.4	148.76	<.0001
iq*gen	1	75.8	0.01	0.9128
iq*met	2	76.4	0.39	0.6792
iq*met*gen	2	75.8	0.19	0.8251

Interpretation

The *p*-value of the iq × met × gen term is 0.8251, indicating it is possible to remove the term from the model.

7.8.2 Step 2: Fit the Factorial Effects for Intercepts and a Main Effects Representation for the Slopes

The model without the iq × met × gen term uses a main effects representation for the slopes, and is fit with the following statements.

Program

```
title2 "Factorial effects for intercepts and main effects for
slopes";
proc mixed data=EX_7_8_spcv;
  class teacher met gen;
  model score = met gen gen*met iq iq*gen iq*met/ ddfm=kr;
  random teacher(met);
run;
```

Results

The results are shown in Output 7.27.

Output 7.27 Fit of the Unequal Slopes Models Where Slopes Are Expressed as Main Effects of Gender and Teaching Method

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	24.4066
Residual	17.7473

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	85	0.41	0.6680
gen	1	77.8	0.08	0.7793
met*gen	2	77.1	5.35	0.0067
iq	1	78.4	153.19	<.0001
iq*gen	1	77.8	0.01	0.9245
iq*met	2	78.4	0.40	0.6725

Interpretation

The *p*-value corresponding to *iq* × *gen* term is 0.9245, indicating that the term can be deleted from the model.

7.8.3 Step 3: Fit the Factorial Effects Model for Intercepts and the Slopes as a Function of Teaching Method

The model with the slope as a function of teaching method is fit with the following program.

Program

```
proc mixed data=EX_7_8_spcv;
  class teacher met gen;
  model score = met gen gen*met iq  iq*met / ddfm=kr;
  random teacher(met);
run;
```

Results

The results of the model above are shown in Output 7.28.

Output 7.28 Fit of Model Where Slopes Are a Function of the Levels of Teaching Method

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	24.4142
Residual	17.5235

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	85.9	0.41	0.6644
gen	1	78.1	2.67	0.1064
met*gen	2	78.1	5.58	0.0054
iq	1	79.4	155.49	<.0001
iq*met	2	79.3	0.41	0.6675

Interpretation

The p -value associated with $\text{iq} \times \text{met}$ is 0.6675, indicating that a common slope model is adequate to describe the covariate part of the model.

7.8.4 Step 4: Fit a Common Slope Model

Model

The following model describes the test scores as a function of IQ but with a common slope and investigates the relationship between the teaching methods and gender:

$$Y_{ijkm} = \alpha_{ik} + \beta_{W_{ijkm}} + t_{j(i)} + e_{ijkm} \quad (7.13)$$

where

$$i = 1, 2, 3$$

$$j = 1, 2, 3, 4$$

$$k = 1, 2$$

$$m = 1, 2, 3, 4$$

$$t_{j(i)} \sim \text{iid } N(0, \sigma_t^2)$$

$$e_{ijkm} \sim \text{iid } N(0, \sigma_e^2)$$

The following program fits model (7.13) to the data.

Program

```
proc mixed data=EX_7_8_spcv;
  class teacher met gen;
  model score = met gen gen*met iq / ddfm=kr;
  random teacher(met);
run;
```

Results

The estimates of the variance components and tests for the fixed effects are displayed in Output 7.29.

Output 7.29 Fit of the Common Slope Model

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	24.4529
Residual	17.2634

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met	2	9.03	10.05	0.0051
gen	1	80	2.65	0.1074
met*gen	2	80	6.09	0.0034
iq	1	81.3	158.52	<.0001

Interpretation

The *p*-value associated with the iq effect in Output 7.29 indicates that the covariate is needed in the model. The *p*-value associated with the met \times gen effect indicates that there is a significant interaction between teaching methods and gender, and thus the gender by teaching method regression lines need to be compared.

Program

Since there is an important teaching method by gender interaction in the intercept part of the model, the means model representation of the intercepts for the teaching method by gender combinations is the simplest model to work with and to carry out the comparisons. The following program uses a means model for the intercepts, uses a common slope for the covariate, and requests the least-squares means for the gen \times met combinations. These are predicted values evaluated at the average IQ value in the data set.

```
title2 "Means model for intercepts and common slope";
proc mixed data=EX_7_8_spcv;
  class teacher met gen;
  model score = gen*met iq / noint solution ddfm=kr;
  random teacher(met);
  lsmeans gen*met/diff;
run;
```

Results

The results are given in Output 7.30, which includes the estimates of the variance components, the estimates of the intercepts and common slope, and the tests of fixed effects. The results of the LSMEANS statement are given in Output 7.31.

Output 7.30 Fit of Common Slope Model with Intercepts Represented as a Means Model

Covariance Parameter Estimates	
Cov Parm	Estimate
teacher(met)	24.4529
Residual	17.2634

Solution for Fixed Effects							
Effect	gen	met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	12.8142	4.6692	59.8	2.74	0.0080
met*gen	m	1	15.1127	4.4970	55.5	3.36	0.0014
met*gen	f	2	26.0002	4.7961	62.8	5.42	<.0001
met*gen	m	2	24.5332	4.5742	57.5	5.36	<.0001
met*gen	f	3	32.3922	4.7171	61	6.87	<.0001
met*gen	m	3	27.3677	4.7884	62.6	5.72	<.0001
iq			0.4701	0.03734	81.3	12.59	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
met*gen	6	36.1	10.97	<.0001
iq	1	81.3	158.52	<.0001

Interpretation

The results in Output 7.30 indicate that there is similar variation between teachers within a teaching method and between students within a class. The “Solution for Fixed Effects” table provides the estimates of the intercepts and common slope. For example, the estimated regression model for teaching method 1 with female students is Score = 12.8142 + 0.4701(IQ).

Output 7.31 Least-Squares Means and Comparisons from Common Slopes Model

Least Squares Means							
Effect	gen	met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	60.9963	2.6818	10.5	22.74	<.0001
met*gen	m	1	63.2948	2.6906	10.6	23.52	<.0001
met*gen	f	2	74.1822	2.6860	10.6	27.62	<.0001
met*gen	m	2	72.7152	2.6846	10.6	27.09	<.0001
met*gen	f	3	80.5743	2.6824	10.5	30.04	<.0001
met*gen	m	3	75.5497	2.6855	10.6	28.13	<.0001

Differences of Least Squares Means									
Effect	gen	met	_gen	_met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	1	m	1	-2.2985	1.4843	80	-1.55	0.1254
met*gen	f	1	f	2	-13.1860	3.7958	10.5	-3.47	0.0055
met*gen	f	1	f	3	-19.5780	3.7931	10.5	-5.16	0.0004
met*gen	m	1	m	2	-9.4204	3.7939	10.5	-2.48	0.0313
met*gen	m	1	m	3	-12.2550	3.8094	10.7	-3.22	0.0085

Differences of Least Squares Means									
Effect	gen	met	_gen	_met	Estimate	Standard Error	DF	t Value	Pr > t
met*gen	f	2	m	2	1.4670	1.4937	80.1	0.98	0.3290
met*gen	f	2	f	3	-6.3921	3.7939	10.5	-1.68	0.1214
met*gen	m	2	m	3	-2.8345	3.8017	10.6	-0.75	0.4721
met*gen	f	3	m	3	5.0245	1.4715	80	3.41	0.0010

Interpretation

The Type 3 F-test for the met \times gen effect indicates an interaction between gender and teaching method. The least-squares means in Output 7.31 are adjusted means or are estimated from the parallel regression lines evaluated at the average IQ value of 102.5. Because there is a significant met \times gen interaction, you need to compare the interaction adjusted means rather than the main-effect adjusted means to complete the analysis. PROC MIXED provides comparisons of least-squares means for each pair, but only some of the comparisons are important. The only differences displayed are for pairs that have either a common level of gender or a common level of teaching method. Without adjusting for multiple testing, male and female means are not different for teaching method 2, and method 2 and method 3 means are not different for males and not different for females.

A more convenient way to perform the comparison of interaction least-squares means is to use the SLICEDIFF= option in the GLIMMIX procedure. It allows you to filter comparisons by holding the level of a factor constant. This avoids comparisons where both factors are varied at the same time. The following statements produce the results in Output 7.32:

```
ods select CovParms SliceDiffs;
proc glimmix data=EX_7_8_spcv;
  class teacher met gen;
  model score = gen*met iq / ddfm=kr;
  random teacher(met);
  lsmeans gen*met / slicediff=(gen met);
run;
```

Notice that the statements are exactly the same as those used with the MIXED procedure. The NOINT and SOLUTION options were dropped from the MODEL statement since they are not important for least-squares means comparisons. The SLICEDIFF=(gen met) option in the LSMEANS statement requests two sets of simple effects comparisons. In the first set the GEN effect is held fixed at each of its levels in turn and comparisons of the MET effect are performed. In the second set the roles of the effects are reversed: MET is held fixed and the genders are compared.

Output 7.32 Least-Squares Means Slice Differences from PROC GLIMMIX

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
teacher(met)	24.4682	12.5615
Residual	17.2623	2.7292

Simple Effect Comparisons of met*gen Least Squares Means By gen							
Simple Effect Level	met	_met	Estimate	Standard Error	DF	t Value	Pr > t
gen f	1	2	-13.1860	3.7968	10.54	-3.47	0.0055
gen f	1	3	-19.5780	3.7941	10.51	-5.16	0.0004
gen f	2	3	-6.3921	3.7949	10.51	-1.68	0.1215
gen m	1	2	-9.4204	3.7949	10.51	-2.48	0.0313
gen m	1	3	-12.2550	3.8104	10.68	-3.22	0.0085
gen m	2	3	-2.8345	3.8027	10.6	-0.75	0.4722

Simple Effect Comparisons of met*gen Least Squares Means By met							
Simple Effect Level	gen	_gen	Estimate	Standard Error	DF	t Value	Pr > t
met 1	f	m	-2.2985	1.4842	80.04	-1.55	0.1254
met 2	f	m	1.4670	1.4937	80.05	0.98	0.3290
met 3	f	m	5.0245	1.4715	80.01	3.41	0.0010

Interpretation

The REML estimates in the “Covariance Parameter Estimates” table of PROC GLIMMIX are slightly different from the estimates of PROC MIXED in Output 7.30. The value of the –2 Res Log Likelihood achieved by PROC GLIMMIX is slightly smaller than that of PROC MIXED. The GLIMMIX procedure arrives at better estimates. You can achieve the same covariance parameter estimates with PROC MIXED by forcing the procedure to add one more iteration (add the option CONVG=1E-7 to your PROC MIXED statement). This difference in the covariance parameter estimates is also the reason for the slight discrepancy between the standard errors in Outputs 7.32 and 7.31. You can see easily how the simple effects results in Output 7.32 provide a filtering and reorganization of those least-squares means comparisons in Output 7.31, where one of the factors is at the same level in both comparisons. In the first “Simple Effects” table, rows 1–3 represent method comparisons among female students, and rows 4–6 represent method comparisons among male students. Teaching method 1 is significantly different from the other two methods for both female and male students. For example, the *p*-value for comparing teaching method 1 to method 2 is *p* = 0.0055 for female students and *p* = 0.0313 for male students. The second “Simple Effects” table contains one row for a comparison of female and male students at each method. There is a significant gender effect for the third method (*p* = 0.0010), but not for the other methods.

7.8.5 Comparison with PROC GLM

As described in Section 7.2, mixed model analysis using PROC MIXED extracts information from all structures of the design, whereas the intra-block analysis obtained using PROC GLM does not. PROC GLM assumes that teacher effects are fixed effects and extracts the information about the slopes from the within-block information (from the smallest size of experimental unit or student), which in this case is the classroom. PROC GLM does not use information about the slopes contained in the block or classroom totals, whereas the mixed models analysis of PROC MIXED does. If you use PROC GLM or PROC MIXED with a teacher(met) effect in the

MODEL instead of the RANDOM statement and there is more than one size of experimental unit in the design, you will not obtain full information about the slopes. It is particularly critical when the model involves unequal slopes, as the slopes are involved in all comparisons between the regression models. An additional problem with the intra-block analysis is that the estimate of the standard errors of the fixed effects is generally not correct when more than one variance component is in the model. The following code fits the model with a means model representation for the intercepts and a common slope where teacher(met) is specified in the RANDOM statement. Least-squares means are requested and the ESTIMATE statement provides the estimate of the intercepts for teaching method 1 applied to female students.

Program

```
title2 "Means model for intercepts and common slope";
proc glm data=EX_7_8_spcv ;
  class teacher met gen;
  model score = gen*met teacher(met) iq / noint solution;
  random teacher(met) / test;
  lsmeans gen*met / pdiff stderr;
  estimate 'met 1 female'
    met*gen 4 0 0 0 0 0
    teacher(met) 1 1 1 1 0 0 0 0 0 0 0 0 / divisor=4;
run;
```

Results

The results are displayed in Output 7.33.

Output 7.33 Intra-block Analysis for Equal Slopes Model and Means Model Representation of the Intercepts Using PROC GLM

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	497635.9355	31102.2460	1801.64	<.0001
Error	80	1381.0645	17.2633		
Uncorrected Total	96	499017.0000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
met*gen	3	260.971437	86.990479	5.04	0.0030
teacher(met)	9	1907.814375	211.979375	12.28	<.0001
iq	1	2751.279220	2751.279220	159.37	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > t
met*gen 1 f	15.16720709	B	4.17962730	3.63	0.0005
met*gen 1 m	17.47875814	B	3.98481408	4.39	<.0001
met*gen 2 f	28.01300332	B	4.31873792	6.49	<.0001
met*gen 2 m	26.56261785	B	4.06915445	6.53	<.0001
met*gen 3 f	39.48739544	B	3.99547044	9.88	<.0001
met*gen 3 m	34.45756150	B	4.07450417	8.46	<.0001
teacher(met) 1 1	-8.65954922	B	2.07798321	-4.17	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > t
teacher(met) 2 1	-6.88567617	B	2.07864105	-3.31	0.0014
teacher(met) 3 1	5.19472125	B	2.07880417	2.50	0.0145
teacher(met) 4 1	0.00000000	B	.	.	.
teacher(met) 1 2	-6.34297149	B	2.08809109	-3.04	0.0032
teacher(met) 2 2	-2.69409844	B	2.07956169	-1.30	0.1989
teacher(met) 3 2	0.00943055	B	2.09449277	0.00	0.9964
teacher(met) 4 2	0.00000000	B	.	.	.
teacher(met) 1 3	-9.93843422	B	2.09275307	-4.75	<.0001
teacher(met) 2 3	-12.70978656	B	2.11427996	-6.01	<.0001
teacher(met) 3 3	-6.68529070	B	2.10532678	-3.18	0.0021
teacher(met) 4 3	0.00000000	B	.	.	.
iq	0.47236062		0.03741695	12.62	<.0001

Source	Type III Expected Mean Square
met*gen	Var(Error) + Q(met*gen)
teacher(met)	Var(Error) + 7.9154 Var(teacher(met))
iq	Var(Error) + Q(iq)

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type III SS	Mean Square	F Value	Pr > F
met*gen	3	260.971437	86.990479	5.04	0.0030
teacher(met)	9	1907.814375	211.979375	12.28	<.0001
iq	1	2751.279220	2751.279220	159.37	<.0001
Error: MS(Error)	80	1381.064530	17.263307		

Least Squares Means

met	gen	score LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	f	60.9965451	1.0387389	<.0001	1
1	m	63.3080961	1.0612525	<.0001	2
2	f	74.1730575	1.0494556	<.0001	3
2	m	72.7226720	1.0458224	<.0001	4
3	f	80.5709816	1.0401200	<.0001	5
3	m	75.5411477	1.0481625	<.0001	6

Least Squares Means for effect met*gen Pr > t for H0: LSMean(i)=LSMean(j)						
Dependent Variable: score						
i/j	1	2	3	4	5	6
1		0.1233	<.0001	<.0001	<.0001	<.0001
2	0.1233		<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		0.3345	<.0001	0.3545
4	<.0001	<.0001	0.3345		<.0001	0.0625
5	<.0001	<.0001	<.0001	<.0001		0.0010
6	<.0001	<.0001	0.3545	0.0625	0.0010	

Parameter	Estimate	Standard Error	t Value	Pr > t
met 1 female	12.5795811	3.96889733	3.17	0.0022

Interpretation

There are slight differences between the results of the mixed model and the within-block model analyses because PROC GLM does not utilize the between-block information about the slopes. For example, the estimate of the common slope is 0.4701 for PROC MIXED (Output 7.30) and 0.4723 for PROC GLM (Output 7.33). But the main differences are in the estimated standard errors for the least-squares means. The estimated standard errors for the least-squares means from PROC MIXED are a little greater than 2.68 (Output 7.31), while the estimated standard errors of the least-squares means from PROC GLM are a little less than 1.07. PROC GLM does not include the teacher(met) variance component appropriately for a model with more than one variance component, and thus the estimated standard errors are too small. The estimate of the intercept for teaching method 1 applied to females from PROC GLM is 12.5796 with estimated standard error 3.9689 (Output 7.33), while the estimate from PROC MIXED is 12.8142 with estimated standard error 4.6692 (Output 7.30).

The analysis of a model with more than one error term requires the use of a mixed models process (such as PROC MIXED) to provide an appropriate analysis, as the within-block analysis provided by PROC GLM is not as efficient and many estimated standard errors are not correct.

7.9 Example: Complex Strip-Plot Design with the Covariate Measured on an Intermediate-Size Experimental Unit

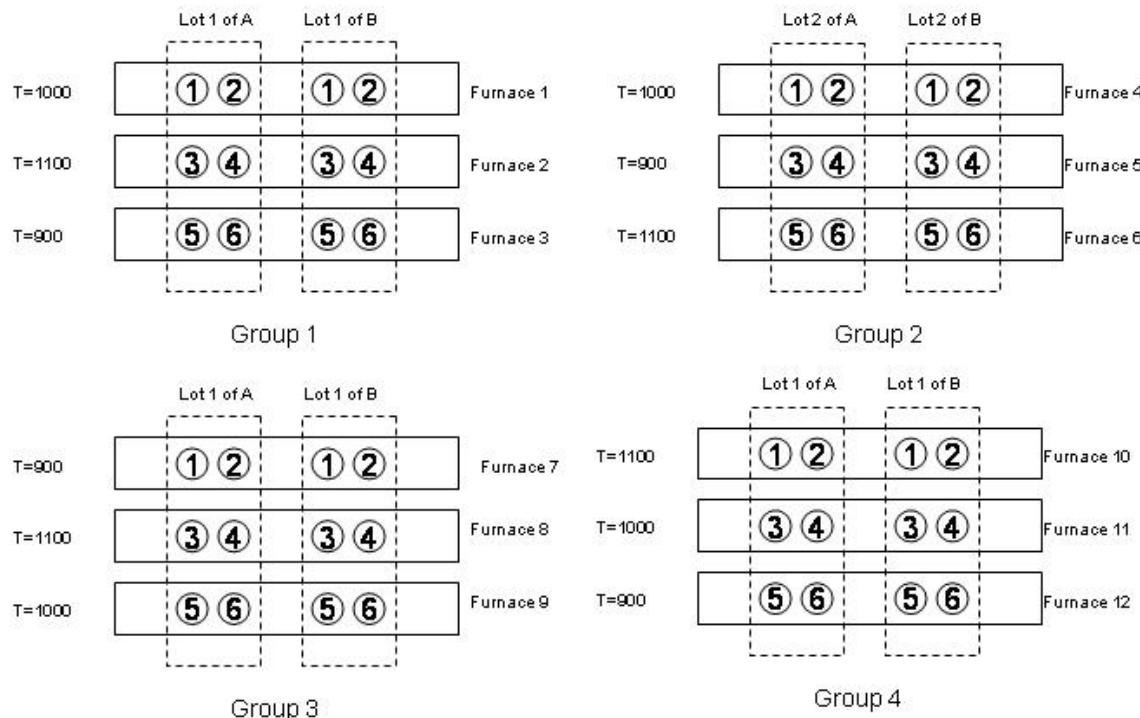
The data in Data Set 7.9, “Wafer Types,” in Appendix 2, “Data Sets,” are from an experiment designed to study the effect of temperature (TEMP at three levels, 900°F, 1000°F, and 1100°F) on the deposition rate (DELTA) of a layer of polysilicon in the fabrication of wafers in the semiconductor industry. The experiment includes two wafer types (A and B) in order to study the possibility of interaction between temperature and wafer type. The experiment consists of putting two wafers of each type into a cassette and then inserting the cassette into a furnace for treatment at a given level of temperature. The furnace is the experimental unit for the levels of temperature. Wafers of a given type are produced in groups of six wafers called **lots**. Four lots of each type of wafer were available for use in the experiment. The six wafers from one randomly selected lot of type A wafers are randomly divided into three groups of two wafers.

These three groups of two type A wafers are randomly assigned to the three furnaces in the group 1 part of the experiment. Similarly, the six wafers from one randomly selected lot of type B wafers are randomly divided into three groups of two wafers. These three groups of two type B wafers are randomly assigned to the three furnaces in the group 1 part of the experiment. The diagram in Figure 7.8 exhibits the randomization process of assigning temperatures to furnaces and assigning wafer types from the wafer lots to the furnaces.

The lot is the experimental unit for wafer type. (Note that the type A wafer lots are different from the type B wafer lots.) Therefore each furnace has two wafers from each of two lots, and the wafers from a given lot occur in three different furnaces, one at each temperature. This process is repeated four times using a total of four lots of type A wafers, four lots of type B wafers, and twelve furnaces or furnace runs. (Actually, the same furnace was used each time.)

The experiment was conducted by randomly assigning the three temperatures to the three furnaces in each group, thus creating a blocking factor, GROUP. The measurements are the amount of deposited material at three randomly chosen sites on each wafer. It was thought that the wafer thickness before the deposition process may have an effect on the deposition rate. Therefore, the average thickness of each wafer was determined and used as a possible covariate. The covariate is measured on the wafer, the next-to-smallest size of experimental unit. The engineer also wanted to estimate the variance components associated with lots, wafers(lots), and sites(wafers lots).

Figure 7.9 Layout of the Wafer Experiment



The design structure consists of six sizes experimental units, the group, the furnace, the lot, the lot-furnace combination, the wafer, and the site. Consequently, there are six variance components in the random effects part of the model. The furnaces and the lots within a group form strip plots. The wafers are nested within lots, and the sites are nested within wafers.

Model

A model to describe the amount of material deposited as it relates to the fixed effects and the random effects is

$$Y_{ijkmn} = \alpha_{jk} + \beta_{jk}t_{ijkm} + g_i + f_{ij} + l_{ik} + fl_{ijk} + w_{ijkm} + e_{ijkmn} \quad (7.14)$$

where

$$i = 1, 2, 3, 4$$

$$j = 1, 2, 3$$

$$k = 1, 2$$

$$m = 1, 2$$

$$n = 1, 2, 3$$

$$g_i \sim iid N(0, \sigma_g^2)$$

$$f_{ij} \sim iid N(0, \sigma_f^2)$$

$$l_{ik} \sim iid N(0, \sigma_l^2)$$

$$fl_{ijk} \sim iid N(0, \sigma_{fl}^2)$$

$$w_{ijkm} \sim iid N(0, \sigma_w^2)$$

$$e_{ijkmn} \sim iid N(0, \sigma_e^2)$$

The intercept and slope of the model for the j^{th} temperature and the k^{th} wafer type are α_{jk} and β_{jk} , where the index i denotes the group, j denotes the level of temperature, k denotes the wafer type, m denotes the wafer, and n denotes the site. The covariate is denoted by t_{ijkm} . Therefore, the fixed effects part of the model is $\alpha_{jk} + \beta_{jk}t_{ijkm}$. The random effects part of the model consists of the group effect, g_i , the furnace effect, f_{ij} , the lot effect, l_{ik} , the lot-furnace effect, fl_{ijk} , and the wafer effect, w_{ijkm} , and is represented by $g_i + f_{ij} + l_{ik} + fl_{ijk} + w_{ijkm}$. The residual part of the model corresponds to the site of a wafer effect, e_{ijkmn} . The preliminary analysis using the strategy as employed in Section 7.6 indicates that a model where the slopes depend on temperature adequately describes these data:

$$Y_{ijkmn} = \alpha_{jk} + \beta_{jk}t_{ijkm} + g_i + f_{ij} + l_{ik} + fl_{ijk} + w_{ijkm} + e_{ijkmn} \quad (7.15)$$

Program

The following program fits model (7.15) to the data:

```
title2 "intercept is means model and slope is fct of temp";
proc mixed data=ex_7_9_lots;
  class grp temp type;
  model delta=temp*type thick*temp / ddfm=kr noint solution;
  random grp grp*temp grp*type
    grp*type*temp wafer(grp temp type);
  lsmeans temp*type / at thick=1828;
  lsmeans temp*type / at thick=2000;
  lsmeans temp*type / at thick=2170;
run;
```

The main difficulty with constructing an appropriate MODEL statement for PROC MIXED is deciding how to code terms for the five variance components (in addition to the residual). The group-to-group variance component is extracted using the GRP effect. Because the levels of temperature and levels of wafer type form a strip plot (see Milliken and Johnson 1992, Chapter 25), the furnace variance component is extracted by using grp \times temp, the lot variance component is extracted by using grp \times type, and the furnace-by-lot variance component is extracted by using grp \times temp \times type. The wafer-to-wafer variance component corresponds to the term wafer(grp type temp), and the site-to-site variance component is provided by the residual component. When the design involves many structures, break the design down into components to determine the appropriate code to compute the desired variance components before carrying out the analysis.

To compare the least-squares means at the three values of the thickness covariate, we again call upon the SLICEDIFF= option in the LSMEANS statement in the GLIMMIX procedure. Alternatively, you can add the DIFF option to the LSMEANS statements in the MIXED code above (or in PROC GLIMMIX) and ignore the comparisons where both type and temp are varied.

```
ods select SliceDiffs;
proc glimmix data=ex_7_9_lots;
  class grp temp type;
  model delta=temp*type thick*temp / ddfm=kr noint solution;
  random grp grp*temp grp*type
    grp*type*temp wafer(grp temp type);
  lsmeans temp*type / slicediff=(temp type) at thick=1828;
  lsmeans temp*type / slicediff=(temp type) at thick=2000;
  lsmeans temp*type / slicediff=(temp type) at thick=2170;
run;
```

Results

The PROC MIXED results are given in Output 7.34. The PROC GLIMMIX multiple comparisons of least-squares means at various thicknesses are given in Output 7.35.

Output 7.34 Estimates of the Variance Components, Analysis of the Fixed Effects and Least-Squares Means for Model (7.15)

Covariance Parameter Estimates	
Cov Parm	Estimate
grp	16.0435
grp*temp	76.0051
grp*type	46.7612
grp*temp*type	1.7377
wafer(grp*temp*type)	10.2578
Residual	3.8296

Solution for Fixed Effects							
Effect	type	temp	Estimate	Standard Error	DF	t Value	Pr > t
temp*type	A	900	98.0756	12.1876	38.5	8.05	<.0001
temp*type	B	900	131.33	13.3081	34.9	9.87	<.0001
temp*type	A	1000	58.2337	13.7262	25.8	4.24	0.0003
temp*type	B	1000	97.6445	14.8970	25.1	6.55	<.0001
temp*type	A	1100	126.64	12.2418	37.4	10.35	<.0001
temp*type	B	1100	184.15	13.4609	37.2	13.68	<.0001
thick*temp		900	0.1073	0.005903	27.1	18.18	<.0001
thick*temp		1000	0.1170	0.006525	18.7	17.94	<.0001
thick*temp		1100	0.07929	0.005711	31	13.89	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
temp*type	6	10.1	51.11	<.0001
thick*temp	3	28.5	301.79	<.0001

Least Squares Means									
Effect	type	temp	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
temp*type	A	900	1828.0	1.50	294.25	6.0444	9.95	48.68	<.0001
temp*type	B	900	1828.0	1.50	327.50	6.1357	10.5	53.38	<.0001
temp*type	A	1000	1828.0	1.50	272.15	6.0554	10	44.94	<.0001
temp*type	B	1000	1828.0	1.50	311.56	6.2741	11	49.66	<.0001
temp*type	A	1100	1828.0	1.50	271.59	6.0456	9.95	44.92	<.0001
temp*type	B	1100	1828.0	1.50	329.10	6.2477	10.9	52.68	<.0001
temp*type	A	900	2000.0	1.50	312.71	6.1631	10.7	50.74	<.0001
temp*type	B	900	2000.0	1.50	345.96	6.0426	9.93	57.25	<.0001
temp*type	A	1000	2000.0	1.50	292.28	6.0856	10.1	48.03	<.0001
temp*type	B	1000	2000.0	1.50	331.69	6.0689	10.1	54.65	<.0001
temp*type	A	1100	2000.0	1.50	285.23	6.0915	10.2	46.82	<.0001
temp*type	B	1100	2000.0	1.50	342.74	6.0726	10.1	56.44	<.0001
temp*type	A	900	2170.0	1.50	330.95	6.4376	12.4	51.41	<.0001
temp*type	B	900	2170.0	1.50	364.20	6.1170	10.4	59.54	<.0001
temp*type	A	1000	2170.0	1.50	312.17	6.3145	11.2	49.44	<.0001
temp*type	B	1000	2170.0	1.50	351.58	6.0666	10	57.95	<.0001
temp*type	A	1100	2170.0	1.50	298.71	6.2891	11.1	47.50	<.0001
temp*type	B	1100	2170.0	1.50	356.22	6.0531	9.99	58.85	<.0001

Interpretation

The Type 3 F-test for the temp \times type effect indicates a significant interaction ($p < 0.0001$), and thus the temperature-wafer type models need to be used to compare the treatment combinations. The least-squares means were computed for five values of the thickness variable, 1828, 2000, and 2170 (10, 50, and 90 percentiles of the distribution of thickness). The least-squares means were also computed for 1800 and 2200 to provide predicted values for graphics. The six regression lines are displayed in Figure 7.9. Since the slopes are different for each temperature, the regression lines for the same temperature are parallel and those for different temperatures are not parallel.

This model involves six different sizes of experimental units, and each has information about the slope in the model. The mixed models analysis extracts the information from each part and then combines the information into a combined slope estimate.

Output 7.35 Comparisons of Temp \times Type Least-Squares Means at Fixed Levels of Temp and Type for Thicknesses of 1828, 2000, and 2170 (PROC GLIMMIX)

Simple Effect Comparisons of temp*type Least Squares Means By temp									
Simple Effect Level	type	_type	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
temp 900	A	B	1828.0	1.50	-33.2512	5.3475	3.872	-6.22	0.0038
temp 1000	A	B	1828.0	1.50	-39.4108	5.3538	3.812	-7.36	0.0022
temp 1100	A	B	1828.0	1.50	-57.5074	5.3752	3.853	-10.70	0.0005

Simple Effect Comparisons of temp*type Least Squares Means By type									
Simple Effect Level	temp	_temp	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
type A	900	1000	1828.0	1.50	22.0981	6.4652	6.392	3.42	0.0128
type A	900	1100	1828.0	1.50	22.6557	6.4561	6.368	3.51	0.0115
type A	1000	1100	1828.0	1.50	0.5576	6.4680	6.397	0.09	0.9339
type B	900	1000	1828.0	1.50	15.9385	6.7579	7.312	2.36	0.0489
type B	900	1100	1828.0	1.50	-1.6005	6.7299	7.225	-0.24	0.8186
type B	1000	1100	1828.0	1.50	-17.5391	6.8976	7.677	-2.54	0.0357

Simple Effect Comparisons of temp*type Least Squares Means By temp									
Simple Effect Level	type	_type	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
temp 900	A	B	2000.0	1.50	-33.2512	5.3475	3.872	-6.22	0.0038
temp 1000	A	B	2000.0	1.50	-39.4108	5.3538	3.812	-7.36	0.0022
temp 1100	A	B	2000.0	1.50	-57.5074	5.3752	3.853	-10.70	0.0005

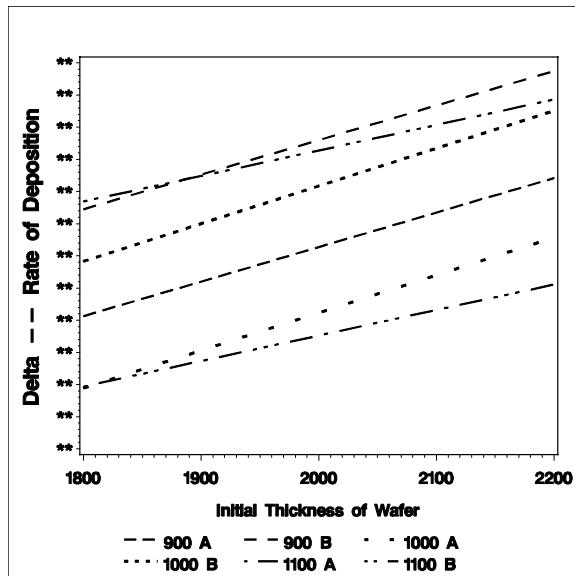
Simple Effect Comparisons of temp*type Least Squares Means By type									
Simple Effect Level	temp	_temp	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
type A	900	1000	2000.0	1.50	20.4286	6.6062	6.88	3.09	0.0179
type A	900	1100	2000.0	1.50	27.4753	6.6104	6.887	4.16	0.0044
type A	1000	1100	2000.0	1.50	7.0467	6.5465	6.614	1.08	0.3194
type B	900	1000	2000.0	1.50	14.2690	6.4763	6.415	2.20	0.0669
type B	900	1100	2000.0	1.50	3.2191	6.4798	6.425	0.50	0.6359
type B	1000	1100	2000.0	1.50	-11.0499	6.5098	6.51	-1.70	0.1366

Simple Effect Comparisons of temp*type Least Squares Means By temp									
Simple Effect Level	type	_type	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
temp 900	A	B	2170.0	1.50	-33.2512	5.3475	3.872	-6.22	0.0038
temp 1000	A	B	2170.0	1.50	-39.4108	5.3538	3.812	-7.36	0.0022
temp 1100	A	B	2170.0	1.50	-57.5074	5.3752	3.853	-10.70	0.0005

Simple Effect Comparisons of temp*type Least Squares Means By type									
Simple Effect Level	temp	_temp	thick	wafer	Estimate	Standard Error	DF	t Value	Pr > t
type A	900	1000	2170.0	1.50	18.7784	7.0741	8.56	2.65	0.0274
type A	900	1100	2170.0	1.50	32.2389	7.0445	8.447	4.58	0.0016
type A	1000	1100	2170.0	1.50	13.4604	6.9795	7.938	1.93	0.0902
type B	900	1000	2170.0	1.50	12.6189	6.5451	6.67	1.93	0.0973
type B	900	1100	2170.0	1.50	7.9826	6.5312	6.629	1.22	0.2633
type B	1000	1100	2170.0	1.50	-4.6362	6.4874	6.453	-0.71	0.4999

Interpretation

The fact that the slopes of the regressions vary by temperature, and, as a result, regressions for the same temperature have the same slope can be gleaned as follows from Output 7.35. The first, third, and fifth “Simple Effect Comparisons” tables in Output 7.35 are least-squares means comparisons sliced by temperature (holding temp fixed). The estimates of the type A versus type B differences are the same in all three tables, regardless of the value of the thickness covariate. For example, the A–B difference is -33.2512 at 900°F, -39.,4108 at 1000° F, and -57.5704 at 1100°F, regardless of thickness. The “Simple Effect Comparisons” tables sliced by type compare the deposition rates at two different temperatures. These comparisons do depend on the wafer thickness (Figure 7.10).

Figure 7.10 Graph of Six Regression Models for Wafer Type Data

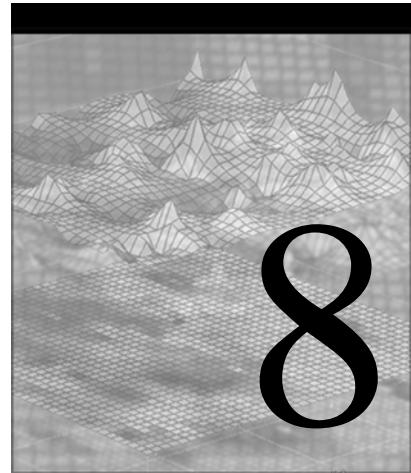
7.10 Summary

An analysis of covariance strategy was described and demonstrated via several mixed model examples. When the model involves more than one factor in the design structure, information about the fixed effects part of the model can be extracted from several parts of the design structure. The mixed models analysis combines the information from the various parts of the design structure and provides combined estimators of the fixed effects. In particular, the estimates of the slopes corresponding to the covariates are computed from the combined information. If the design structure involves only one block—i.e., a completely randomized design structure—the usual within-block analysis extracts all of the information about the slopes from the data. When the design structure is more complex, the mixed models analysis extracts more information from the data than the usual analysis provided by an analysis where the blocks are considered as fixed effects.

The examples in this chapter employed mostly PROC MIXED for model fitting and its ESTIMATE, CONTRAST, and LSMEANS statements for post-processing. In two instances the SLICEDIFF= option of PROC GLIMMIX was helpful to simplify and organize least-squares means comparisons. In general, the post-processing facilities of PROC GLIMMIX exceed those of PROC MIXED. For example, PROC GLIMMIX allows you to specify multiple-row ESTIMATE statements (akin to standard multiple-row CONTRAST statements) and to adjust those for multiplicity. The LSMESTIMATE statement in PROC GLIMMIX enables you to test linear combinations of least-squares means (rather than linear combinations of the model

coefficients via ESTIMATE statements). The LSMEANS statement of PROC GLIMMIX offers facilities for graphing least-squares means (see Section 5.3.4) and least-squares means comparisons, producing slice differences, performing analysis of means (ANOM, Ott 1967; Nelson 1982, 1991, 1993), applying enhanced multiplicity adjustments (e.g., stepdown p -values; Westfall 1997), producing lines (lettering) output, and so forth. PROC GLIMMIX makes the post-fitting analysis process easier.

A large variety of applications of analysis of covariance models including repeated measures and random coefficient models are discussed in detail in Milliken and Johnson (2002).



Random Coefficient Models

8.1	Introduction	317
8.2	Example: One-Way Random Effects Treatment Structure in a Completely Randomized Design Structure	320
8.3	Example: Random Student Effects	326
8.4	Example: Repeated Measures Growth Study	330
8.4.1	Repeated Measures Analysis	332
8.4.2	Random Coefficient Analysis.....	334
8.4.3	Test of Lack of Fit.....	339
8.5	Summary	341

8.1 Introduction

Data that have a nested or hierarchical structure are common in a wide variety of disciplines, and similar methods for analyzing such data are found in these disciplines under different guises. The analyses considered here fall under the headings of **random coefficient models** and **empirical Bayes models** in the statistics literature (Laird and Ware 1982, Strenio, Weisberg, and Bryk 1983, Rutter and Elashoff 1994, Wolfinger, 1996). Analogous terms in the educational and social science literature are **hierarchical linear models** and **multilevel linear models** (see, for example, Goldstein 1987, Bryk and Raudenbush 1992, and contributions in a special issue of the *Journal of Educational and Behavioral Statistics*, 1995). A primary objective of this chapter is to describe these models and illustrate how to fit random coefficient models using PROC MIXED.

The basic structure of random coefficient models builds on the analysis of covariance models discussed in Chapter 7. There linear regression models are used to include continuous variables (covariates) as independent variables. In Chapter 7 the regression coefficients for the covariates were assumed to be fixed effects—that is, unknown fixed parameters that were estimated from the data.

In this chapter the regression coefficients for one or more covariates are assumed to be a random sample from some population of possible coefficients; hence the term **random coefficients**. Random coefficient models are sensible whenever the data arise from independent

subjects or clusters and the regression model for each subject or cluster can be assumed to be a random deviation from some population regression model.

The standard random coefficient model involves a random intercept and slope for each subject. Let Y_{ij} denote the measurement of the j^{th} observation on the i^{th} subject, and the random coefficient model can be written as

$$Y_{ij} = a_i + x_{ij}b_i + e_{ij} \quad (8.1)$$

where

$$\begin{aligned} i &= 1, 2, \dots, t \\ j &= 1, 2, \dots, n_i \\ \begin{bmatrix} a_i \\ b_i \end{bmatrix} &\sim \text{iid } N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{G}\right) \\ \mathbf{G} &= \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} \\ e_{ij} &\sim \text{iid } N(0, \sigma^2) \end{aligned}$$

Model (8.1) can be expressed as

$$Y_{ij} = \alpha + a_i^* + \beta x_{ij} + b_i^* x_{ij} + e_{ij} \quad (8.2)$$

where

$$\begin{aligned} i &= 1, 2, \dots, t \\ j &= 1, 2, \dots, n_i \\ a_i^* &= a_i - \alpha \\ b_i^* &= b_i - \beta \\ \begin{bmatrix} a_i^* \\ b_i^* \end{bmatrix} &\sim \text{iid } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{G}\right) \\ e_{ij} &\sim \text{iid } N(0, \sigma^2) \end{aligned}$$

Model (8.2) can be expressed in terms of a mixed model as

$$Y_{ij} = \alpha + \beta x_{ij} + a_i^* + b_i^* x_{ij} + e_{ij} \quad (8.3)$$

where

$$i = 1, 2, \dots, t \quad j = 1, 2, \dots, n_i$$

$\alpha + \beta x_{ij}$ is the fixed effects part of the model

$a_i^* + b_i^* x_{ij}$ is the random effects part of the model

e_{ij} is the residual part of the model

Finally, model (8.3) can be expressed as

$$Y_{ij} = \alpha + \beta x_{ij} + e_{ij}^* \quad (8.4)$$

where

$$i = 1, 2, \dots, t$$

$$j = 1, 2, \dots, n_i$$

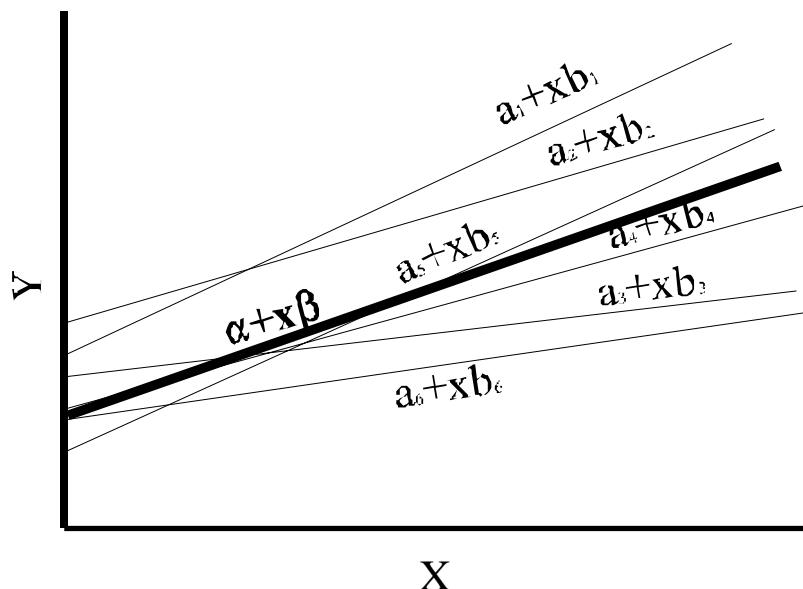
$$E[Y_{ij}] = \alpha + \beta x_{ij}$$

$$e_{ij}^* = a_i^* + b_i^* x_{ij} + e_{ij}$$

$$\text{Var}[Y_{ij}] = [1, x_{ij}] \mathbf{G} \begin{bmatrix} 1 \\ x_{ij} \end{bmatrix} + \sigma_e^2$$

A graphical representation of this model is displayed in Figure 8.1. The random regression lines for each subject deviate about the overall population regression line, $\mu(y | x) = \alpha + \beta x$. For this graph the covariance between intercept and slope, σ_{ab} , is small, as there is no real relationship between a subject's intercept and slope.

Figure 8.1 Several Simple Linear Regression Models from a Random Sample of Treatments with the Population Model



The random coefficient model is a conditional hierarchical linear model (Bryk and Raudenbush 1992) where the experimental units are nested within the randomly selected levels of the subjects. It is possible to introduce additional levels of hierarchy as well as more than two random coefficients at each level. The first two examples in the next section consider only the standard random intercept-slope model just described, and the third example consists of a quadratic random coefficient model with three random coefficients. Using the notation from the previous section, the analyses of the examples consist of estimating α , β , \mathbf{G} , and σ^2 and of testing hypotheses about the parameters.

8.2 Example: One-Way Random Effects Treatment Structure in a Completely Randomized Design Structure

The data in Data Set 8.2, “Winter Wheat,” in Appendix 2, “Data Sets,” are from ten varieties of wheat that were randomly selected from the population of varieties of hard red winter wheat adapted to dry climate conditions. The experimental units are 1-acre plots of land in a 60-acre field. The varieties were randomly assigned to six 1-acre plots of land. It was thought that the preplanting moisture content of the plots could have an influence on the germination rate and hence on the eventual yield of the plots. Thus, the amount of preplanting moisture in the top 36 inches of the soil was determined for each plot. The response is yield in bushels per acre (YIELD), and the covariate is the measured amount of moisture (MOIST), which was measured before planting the varieties on the plots. Because the varieties were randomly selected, the resulting regression model for each variety is a random model selected from the population of variety models. The fixed effects of the model are the population intercept and the slope, which are the expected values of the population of the intercepts and slopes of the varieties.

Model

The following model, akin to model (8.1), describes these data:

$$Y_{ij} = a_i + b_i x_{ij} + e_{ij} \quad (8.5)$$

where

$$i = 1, 2, \dots, 10$$

$$j = 1, 2, \dots, 6$$

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} \sim iid N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{G}\right)$$

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$$

$$e_{ij} \sim iid N(0, \sigma^2)$$

Program

The following program fits model (8.5) to the data:

```
proc mixed data=wheat covtest cl scoring=8;
  class variety;
  model yield = moist / solution;
  random int moist / type=un subject=variety solution G Gcorr;
  ods output solutionf=fixed solutionr=random;
run;
```

The SUBJECT=VARIETY option in the RANDOM statement specifies that the intercept and slope of one variety are independently distributed from the intercepts and slopes of other varieties. The option TYPE=UN in the RANDOM statement specifies the covariance structure for a subject's random effects. TYPE=UN calls for an unstructured (2×2) covariance matrix, comprising the variance of the random slopes (σ_a^2), the variance of the random intercepts (σ_b^2), and their covariance (σ_{ab}). The effects INT and MOIST in the RANDOM statement, together with the SUBJECT=VARIETY option, instruct PROC MIXED to add a random intercept and a random MOIST slope for each variety to the fixed effects part of the model. The fixed effects part consists of a fixed intercept (automatically included through the MODEL statement) and a fixed slope in MOIST. The option SCORING=8 requests that PROC MIXED use Fisher's scoring for the first eight iterations. The COVTEST and CL options in the PROC MIXED statement request estimated standard errors and 95% confidence intervals for the covariance parameters. The confidence intervals for variances are constructed using the Satterthwaite type approximation with degrees of freedom equal to $2 \times (z\text{-score})^2$. The confidence interval about the covariance is computed using the Wald interval. The ODS statement provides two data sets, one with the fixed effects parameter estimates and one with the predictions of the intercepts and slopes for each of the random effects.

Results

The results are given in Outputs 8.1–8.2.

Output 8.1 Results for Covariance Parameters and Fixed Effects

Estimated G Matrix				
Row	Effect	variety	Col1	Col2
1	Intercept	1	18.8954	-0.07272
2	moist	1	-0.07272	0.002395

Estimated G Correlation Matrix				
Row	Effect	variety	Col1	Col2
1	Intercept	1	1.0000	-0.3419
2	moist	1	-0.3419	1.0000

Covariance Parameter Estimates								
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
UN(1,1)	variety	18.8954	9.1117	2.07	0.0191	0.05	8.8169	65.2880
UN(2,1)	variety	-0.07272	0.08244	-0.88	0.3777	0.05	-0.2343	0.08885
UN(2,2)	variety	0.002395	0.001350	1.77	0.0380	0.05	0.001011	0.01103
Residual		0.3520	0.07901	4.46	<.0001	0.05	0.2370	0.5775

Fit Statistics	
-2 Res Log Likelihood	186.1
AIC (smaller is better)	194.1
AICC (smaller is better)	194.9
BIC (smaller is better)	195.4

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	33.4339	1.3985	9	23.91	<.0001
moist	0.6617	0.01679	9	39.42	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
moist	1	9	1553.84	<.0001

Interpretation

The “Solution for the Fixed Effects” table gives $\hat{\alpha} = 33.4339$ and $\hat{\beta} = 0.6617$. Based on the RANDOM statement, REML estimates of the variance and covariance components are obtained from the “Estimated G Matrix” and “Covariance Parameters Estimates” table as

$$\hat{\mathbf{G}} = \begin{bmatrix} 18.8954 & -0.07272 \\ -0.07272 & 0.002395 \end{bmatrix}$$

$$\hat{\sigma}_e^2 = 0.3520$$

The SOLUTION option in the RANDOM statement provides the EBLUPs of the variety effects for the intercepts and slopes. The “Solution for Random Effects” table in Output 8.2 contains the predicted values for the deviations of the varieties’ intercepts from the population mean intercept (the first variety’s intercept is labeled “Intercept 1”; this is a prediction of $a_1 - \alpha$) and the deviations of the varieties’ slopes from the population mean intercept (the first variety’s slope is labeled “moist 1”; this is a prediction of $b_1 - \beta$) for each variety.

Output 8.2 Results for Random Effects

Solution for Random Effects						
Effect	variety	Estimate	Std Err Pred	DF	t Value	Pr > t
Intercept	1	0.9578	1.5102	40	0.63	0.5295
moist	1	-0.04921	0.02114	40	-2.33	0.0250
Intercept	2	-2.2842	1.5269	40	-1.50	0.1425
moist	2	-0.06670	0.02742	40	-2.43	0.0196
Intercept	3	-0.4082	1.5159	40	-0.27	0.7891
moist	3	0.06723	0.02231	40	3.01	0.0045
Intercept	4	0.6960	1.4819	40	0.47	0.6411
moist	4	-0.02331	0.02386	40	-0.98	0.3346
Intercept	5	1.1159	1.6929	40	0.66	0.5136
moist	5	-0.01990	0.02599	40	-0.77	0.4483
Intercept	6	4.6391	1.4671	40	3.16	0.0030
moist	6	0.02389	0.01989	40	1.20	0.2367
Intercept	7	-10.7301	1.4554	40	-7.37	<.0001
moist	7	0.05643	0.02220	40	2.54	0.0150
Intercept	8	2.4011	1.4842	40	1.62	0.1136
moist	8	0.02244	0.02213	40	1.01	0.3168
Intercept	9	-0.1763	1.5168	40	-0.12	0.9081
moist	9	0.02336	0.02227	40	1.05	0.3005
Intercept	10	3.7887	1.8340	40	2.07	0.0454
moist	10	-0.03421	0.02872	40	-1.19	0.2407

These deviations of intercepts and deviations of slopes can be used with the estimates of α and β to construct a plot of the family of simple linear regression lines. The values of α and β are in the data set FIXED, and the predicted deviations for each variety's intercept and slope are in the data set RANDOM (generated with the ODS OUTPUT statement in the preceding program). The following program can be used to calculate the values of the intercepts (a_i) and the values of the slopes (b_i) (Output 8.3). The variables FINT and FSLOPE contain the fixed effect parameter estimates, and the variables RINT and RSLOPE contain the predicted deviations.

Program

```

data lines;
merge random(where=(effect='Intercept') rename=(estimate=rint))
      random(where=(effect='moist'      ) rename=(estimate=rslope));
if _n_ = 1 then merge
   fixed(where=(effect='Intercept') rename=(estimate=fint))
   fixed(where=(effect='moist'      ) rename=(estimate=fslope));
intercept = fint + rint;
slope     = fslope + rslope;
keep variety fint fslope rint rslope intercept slope;
run;

```

```

proc print data=lines;
  var variety fint rint intercept fslope rslope slope;
run;

```

Results

The predicted intercepts (variable INTERCEPT) and predicted slopes (variable SLOPE) for each variety are shown in Output 8.3.

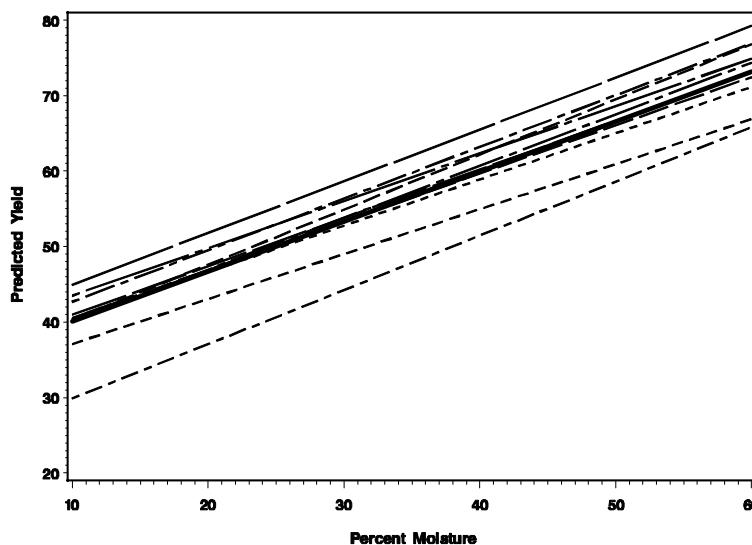
Output 8.3 Predicted Slopes and Intercepts for Each Variety

Obs	variety	fint	rint	intercept	fslope	rslope	slope
1	1	33.4339	0.9578	34.3917	0.6617	-0.04921	0.61244
2	2	33.4339	-2.2842	31.1498	0.6617	-0.06670	0.59495
3	3	33.4339	-0.4082	33.0257	0.6617	0.06723	0.72888
4	4	33.4339	0.6960	34.1299	0.6617	-0.02331	0.63835
5	5	33.4339	1.1159	34.5498	0.6617	-0.01990	0.64175
6	6	33.4339	4.6391	38.0730	0.6617	0.02389	0.68554
7	7	33.4339	-10.7301	22.7038	0.6617	0.05643	0.71808
8	8	33.4339	2.4011	35.8350	0.6617	0.02244	0.68409
9	9	33.4339	-0.1763	33.2576	0.6617	0.02336	0.68501
10	10	33.4339	3.7887	37.2226	0.6617	-0.03421	0.62745

Interpretation

Regression lines were constructed by evaluating the model for each variety at MOIST=10 and 60 and then connecting the two points. Figure 8.2 displays the set of regression lines for the ten varieties and the estimate of the population mean (dark line in center of bundle).

Figure 8.2 Predicted Simple Linear Regression Models from a Random Sample of Varieties with the Estimated Population Model



Model (8.5) specifies a covariance between the intercepts and slopes. Sometimes a model with $\sigma_{ab} = 0$ describes the data as well as the model where the covariance is estimated. Fitting an unstructured covariance matrix can also be numerically difficult, and a model without covariances between the random effects provides less difficulty. Notice that even if all entries of the (2×2) unstructured covariance matrix are positive, the resulting covariance matrix may not be positive definite. The following code fits the model with $\sigma_{ab} = 0$. Specifying TYPE=UN(1) requests that unequal variances (diagonal values) be estimated from the data but that all covariances are set to zero.

Program

```
proc mixed data=wheat scoring=8;
  class variety;
  model yield = moist/solution;
  random int moist/subject=variety type=un(1) solution G Gcorr;
run;
```

Results

The results are given in Output 8.4.

Output 8.4 Results of Fitting the Random Coefficients Model with Zero Covariance

Estimated G Matrix				
Row	Effect	variety	Col1	Col2
1	Intercept	1	18.2824	
2	moist	1		0.002291

Estimated G Correlation Matrix				
Row	Effect	variety	Col1	Col2
1	Intercept	1	1.0000	
2	moist	1		1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	variety	18.2824
UN(2,1)	variety	0
UN(2,2)	variety	0.002291
Residual		0.3543

Fit Statistics	
-2 Res Log Likelihood	187.1
AIC (smaller is better)	193.1
AICC (smaller is better)	193.6
BIC (smaller is better)	194.0

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	33.4108	1.3761	9	24.28	<.0001
moist	0.6619	0.01647	9	40.19	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
moist	1	9	1615.29	<.0001

Interpretation

The estimates of the variance components (with $\sigma_{ab} = 0$) from the “Estimated G Matrix” and “Covariance Parameter Estimates” tables are

$$\hat{\mathbf{G}} = \begin{bmatrix} 18.2824 & 0 \\ 0 & 0.002291 \end{bmatrix}$$

$$\hat{\sigma}_e^2 = 0.3543$$

The estimates of the fixed effects intercept and slope have changed slightly compared to Output 8.1. There was little evidence in Output 8.1 that $\sigma_{ab} \neq 0$ (see the Z score corresponding to UN(2,1), $p = 0.3776$ in the “Covariance Parameter Estimates” table). Because the two models are nested with respect to the covariance parameters, we can perform a likelihood ratio test. The $-2 \text{ Res Log Likelihood}$ for the model with $\sigma_{ab} = 0$ is 187.1129. The $-2 \text{ Res Log Likelihood}$ for the unstructured covariance matrix is 186.1463, a difference of 0.9667, which corresponds (asymptotically) to a chi-square distribution with 1 degree of freedom. Thus, there is insufficient evidence to reject the notion that $\sigma_{ab} = 0$.

8.3 Example: Random Student Effects

Kreft et al. (1994) analyze data from the Second International Mathematics Study (SIMS). Here 3,691 eighth-grade students are measured on mathematics achievement tests, and the hierarchical structure of the data is due to the fact that these students are grouped into 190 classes. The raw data are given in Data Set 8.3, “Mathematics Study,” in Appendix 2, “Data Sets.”

Model

A multilevel analysis of these data begins by constructing a model at the student level involving any explanatory variables measured on the students. For this example, Kreft et al. (1994) use the standard regression model

$$(GAIN)_{ij} = \beta_{0j} + \beta_{1j}(PRETOT)_{ij} + e_{ij}$$

where $(GAIN)_{ij}$ is the gain on the score of a particular achievement test of the i^{th} student in the j^{th} class and $(PRETOT)_{ij}$ is the sum of some pretest score items for the same student. The

residual errors e_{ij} are assumed to be independent and identically distributed Gaussian random variables with mean zero and variance σ^2 .

Next, the regression coefficients β_{0j} and β_{1j} are assumed to arise from a model at the class level. Assuming initially that there are no class-level variables, the basic model to consider here is

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + d_{0j} \\ \beta_{1j} &= \gamma_{10} + d_{1j}\end{aligned}$$

where the class-level disturbance terms (d_{0j}, d_{1j}) are assumed to be independent and identically distributed bivariate Gaussian random variables with zero mean and variance-covariance matrix \mathbf{G} or

$$\begin{bmatrix} d_{0j} \\ d_{1j} \end{bmatrix} \sim iid N(\mathbf{0}, \mathbf{G}), \text{ where } \mathbf{G} = \begin{bmatrix} \sigma_{00} & \sigma_{10} \\ \sigma_{10} & \sigma_{11} \end{bmatrix}$$

Substituting the expressions for β_{0j} and β_{1j} into the student-level model produces the following single-equation formulation:

$$(GAIN)_{ij} = \gamma_{00} + \gamma_{10}(PRETOT)_{ij} + d_{0j} + (PRETOT)_{ij}d_{1j} + e_{ij}$$

This equation reveals that the hierarchical linear model is actually the same as the random coefficient model considered in previous sections. The mean model consists of the two parameters γ_{00} and γ_{10} , and the variance model has a random intercept d_{0j} , a random slope d_{1j} , and a residual error e_{ij} . The fixed effects part of the model is $\gamma_{00} + \gamma_{10}(PRETOT)_{ij}$, the random effects part of the model is $d_{0j} + (PRETOT)_{ij}d_{1j}$, and the residual part of the model is e_{ij} .

Program

The single-equation formulation is probably easiest to translate into PROC MIXED code because all of the fixed effects must be placed together in the MODEL statement and all of the random effects are specified in the RANDOM statement. For this example, an appropriate PROC MIXED program is as follows:

```
proc mixed data=sims covtest cl;
  class class;
  model gain = pretot / s ddfm=kr;
  random int pretot / subject=class type=un s;
  ods output solutionf=est solutionr=rand;
run;
```

The DATA= option in the PROC MIXED statement specifies the SAS data set to use for the analysis. The COVTEST option requests the estimated standard error and z-score, and the CL option indicates that confidence intervals about the covariance parameters are to be included.

The CLASS statement looks a bit unusual here because the variable CLASS shares its same name. The purpose of the CLASS statement is to treat the listed variables as classification rather than continuous variables. This is done for the variable CLASS, which indicates the class of a particular student.

The MODEL statement specifies the mean model effects. An intercept is included by default, corresponding to γ_{00} , and the variable PRETOT models the γ_{10} term. Because PRETOT is not in the CLASS statement, it is treated as a continuous effect with a single degree-of-freedom

regression. The S (=SOLUTION) option requests that the estimates of γ_0 and γ_1 be printed along with their estimated standard errors and corresponding *t*-statistics.

The RANDOM statement is the mechanism for specifying the terms involving d_{0j} and d_{1j} . The INTERCEPT effect corresponds to the former; INTERCEPT is a keyword automatically interpreted by PROC MIXED as an effect with all 1's. PRETOT corresponds to the $(\text{PRETOT})_{ij}d_{1j}$ term in the model. The SUBJECT=CLASS option is important because it instructs PROC MIXED regarding the index *j* and when it changes. Observations from different classes are assumed to be independent. The TYPE=UN option sets up **G** as an unstructured (2×2) matrix with the three parameters σ_{00} , σ_{10} , and σ_{11} .

PROC MIXED includes the homogeneous residual error e_{ij} in the model by default.

Results

The results of this analysis are shown in Output 8.5.

Output 8.5 Results of Random Student Effects Model with DDFM=KR

Covariance Parameter Estimates								
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
UN(1,1)	class	14.4797	2.6993	5.36	<.0001	0.05	10.3642	21.6578
UN(2,1)	class	-0.2337	0.1016	-2.30	0.0214	0.05	-0.4328	-0.03456
UN(2,2)	class	0.009192	0.004940	1.86	0.0314	0.05	0.004004	0.03847
Residual		22.2362	0.5519	40.29	<.0001	0.05	21.1930	23.3588

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	7.0595	0.3665	165	19.26	<.0001
pretot	-0.1860	0.01620	126	-11.48	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
pretot	1	126	131.86	<.0001

Interpretation

The “Covariance Parameter Estimates” table displays the estimates of σ_{00} , σ_{10} , and σ_{11} , and σ^2 in the Estimate column. The Standard Error column displays approximate standard errors of the estimates based upon the fact that they are asymptotically normal. A corresponding *z*-score and *p*-value are also printed, testing whether the parameter is different from zero. These tests can, however, be unreliable in small samples.

The “Solution for Fixed Effects” table prints estimates of the mean model parameters γ_0 and γ_1 . The “Type 3 Tests of Fixed Effects” table prints an *F*-test for $\gamma_0 = 0$; this hypothesis is strongly rejected. Because this *F*-statistic has only 1 numerator degree of freedom, it is equal to

the square of the t -statistic for PRETOT in the “Solution for Fixed Effects” table and the p -values are the same.

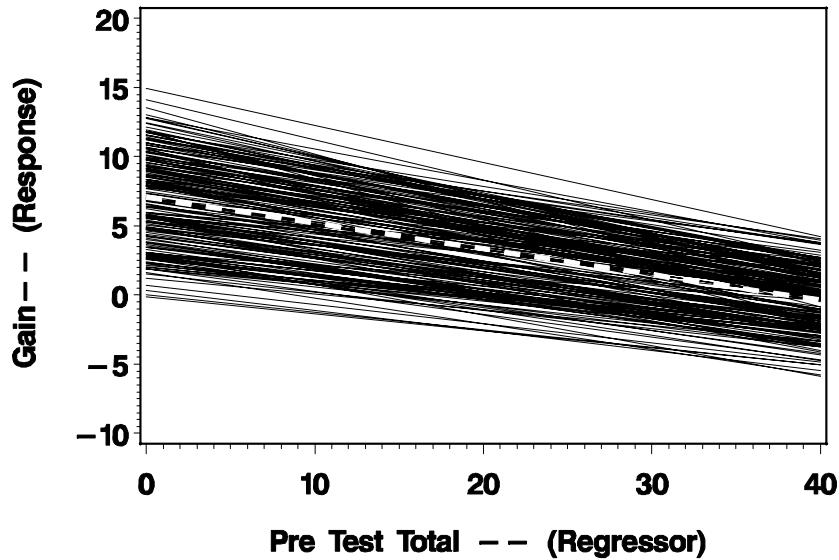
For comparison, Table 8.1 reproduces a portion of Table 1 from Kreft et al. (1994) as well as the results from PROC MIXED. It displays REML results for the preceding model from three other software packages. They are seen to agree closely with the PROC MIXED results.

Table 8.1 Comparison of REML Results for SIMS Data from Other Software Packages and PROC MIXED

Parameter	GENMOD	HLM	ML3	MIXED
γ_{00}	7.060	7.060	7.060	7.060
γ_{10}	-0.186	-0.186	-0.186	-0.186
σ_{00}	14.52	14.53	14.49	14.48
σ_{10}	-0.234	-0.237	-0.234	-0.234
σ_{11}	0.009	0.009	0.009	0.009
σ^2	22.23	22.23	22.24	22.24

The intercepts and slopes for each class were put together as for the previous examples, and the regression lines were evaluated at pretest total of 0 and 40. The resulting predicted regression lines as well as the estimated population linear regression model are displayed in Figure 8.3, where the white-black dashed line denotes the population model.

Figure 8.3 Predicted Simple Linear Regression Models from a Random Sample of Classes with the Estimated Population Model



The final analysis uses the Satterthwaite adjustment for degrees of freedom and compares the results to those with DDFM=KR.

Program

```
proc mixed data=sims info;
  model gain = pretot / s ddfm=satterth;
  random int pretot / sub=class type=un;
run;
```

Results

The results are given in Output 8.6.

Output 8.6 Results of Random Student Effects Model with DDFM=SATTERTH

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	class	14.4797
UN(2,1)	class	-0.2337
UN(2,2)	class	0.009192
Residual		22.2362

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	7.0595	0.3658	165	19.30	<.0001
pretot	-0.1860	0.01610	126	-11.56	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
pretot	1	126	133.57	<.0001

Interpretation

The main difference between the results in Output 8.5 and Output 8.6 is in the computation of the estimated standard errors for the fixed effects. The method used to compute the estimated standard errors for the fixed effects in Output 8.6 (with DDFM=SATTERTH) is the model-based method, while that used in Output 8.5 (with DDFM=KR) is the Prasad-Rao-Jeske-Kackar-Harville method. The estimated standard errors in Output 8.5 are a little larger than those in Output 8.6, thus demonstrating the increasing effect of the Prasad-Rao-Jeske-Kackar-Harville method. This adjustment is also evident in the values of the *F*-statistics, which are smaller with the Prasad-Rao-Jeske-Kackar-Harville method.

8.4 Example: Repeated Measures Growth Study

The growth of animals is determined by measuring the same animal at several times during the growth period of concern. This process generates repeated measures data. The data for this example come from studying the growth of pigs after weaning for 30 days. The weight of each pig is measured at days 0, 6, 12, 18, 24, and 30 after weaning. A plot of the data shows the growth exhibits a quadratic response as a function of days. The objective of the experiment was

to determine if any of three treatments had an effect on the growth response. The data appear as Data Set 8.4, “Pig Growth,” in Appendix 2, “Data Sets.” Data Set 8.4 is in multivariate form, where repeated measurements of an animal correspond to different variables. A SAS DATA set in univariate form—as required by PROC MIXED—is produced with the following program:

```
data pigs;
  set pigsx;
  array days{6} day:;
  do i=1 to 6;
    day = (i-1)*6;
    weight = days{i};
    output;
  end;
  keep trt pig weight day;
run;
```

Figures 8.4, 8.5, and 8.6 show the growth curves (connect the dots between measurements) of the 20 pigs assigned to each of the three treatments.

Figure 8.4 Weight Data for Pigs Assigned to Treatment 1

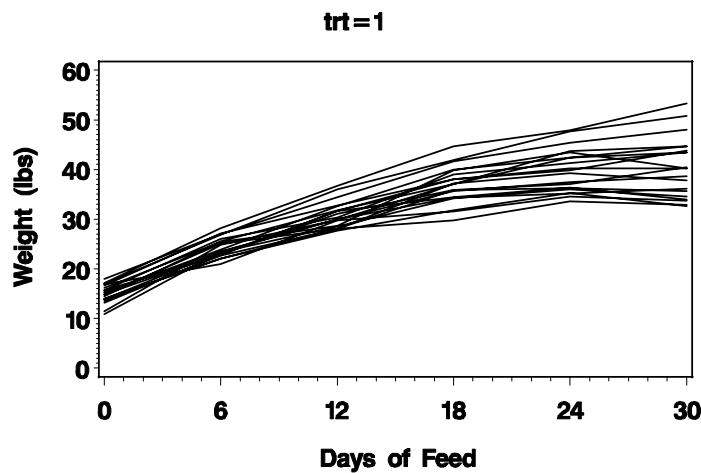


Figure 8.5 Weight Data for Pigs Assigned to Treatment 2

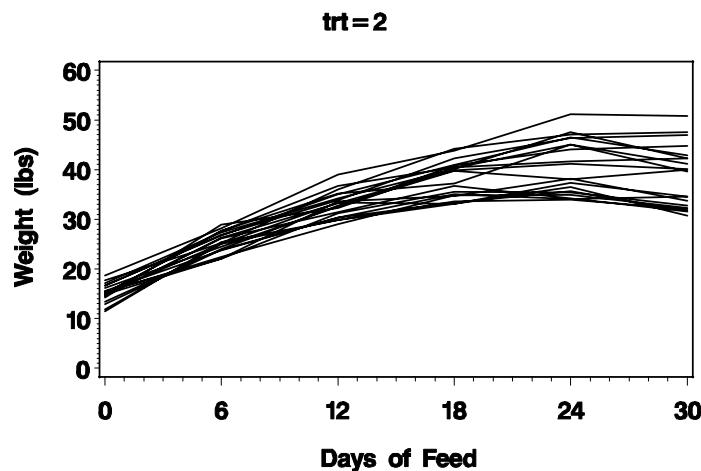
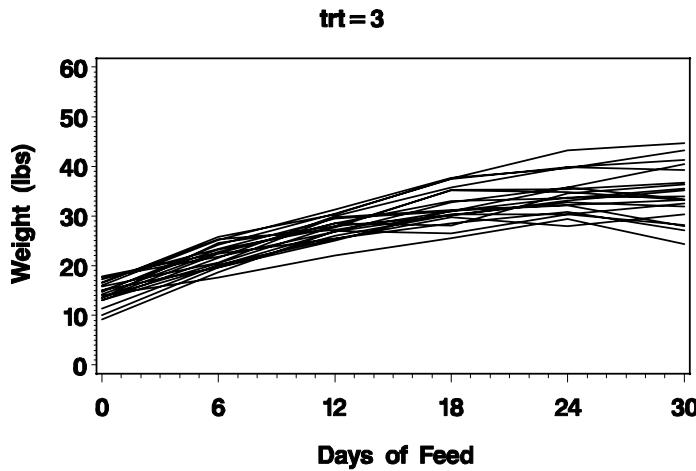


Figure 8.6 Weight Data for Pigs Assigned to Treatment 3

8.4.1 Repeated Measures Analysis

One approach to the analysis of this data set is to carry out a repeated measures analysis of variance where a REPEATED statement is used to model the covariance structure of the repeated measures. Since the measurements are made at equally spaced time points, an AR(1) covariance structure is a possibility. The repeated measures model is

$$wt_{ijk} = \mu + \tau_i + \delta_k + (\tau\delta)_{ik} + p_{ij} + e_{ijk}$$

$$i=1,2,3 \quad j=1,2,\dots,20 \quad k=1,2,\dots,6$$

where it is assumed that

$$p_{ij} \sim iid N(0, \sigma_{pig}^2)$$

and

$$\begin{bmatrix} e_{ij1} \\ e_{ij2} \\ e_{ij3} \\ e_{ij4} \\ e_{ij5} \\ e_{ij6} \end{bmatrix} \sim iid N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \sigma_e^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \right)$$

The following program fits the model above.

Program

```
proc mixed data=pigs;
  class trt day pig;
  model weight=trt day day*trt / ddfm=kr;
  random pig / subject=trt;
  repeated day / subject=pig(trt) type=ar(1);
  lsmeans day*trt;
run;
```

The RANDOM statement specifies $p_{ij} \sim iid N(0, \sigma_{pig}^2)$ and the REPEATED statement specifies that DAY is the repeated measurement, that pig(trt) is the subject effect, and that the covariance structure of the repeated measurements has AR(1) structure. The results of the analysis are given in Output 8.7.

Results

Output 8.7 Results of Repeated Measures Analysis with TYPE=AR(1)

Estimated R Matrix for pig(trt) 1 1						
Row	Col1	Col2	Col3	Col4	Col5	Col6
1	14.5604	11.9035	9.7315	7.9558	6.5041	5.3173
2	11.9035	14.5604	11.9035	9.7315	7.9558	6.5041
3	9.7315	11.9035	14.5604	11.9035	9.7315	7.9558
4	7.9558	9.7315	11.9035	14.5604	11.9035	9.7315
5	6.5041	7.9558	9.7315	11.9035	14.5604	11.9035
6	5.3173	6.5041	7.9558	9.7315	11.9035	14.5604

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
pig	trt	1.7574
AR(1)	pig(trt)	0.8175
Residual		14.5604

Fit Statistics	
-2 Res Log Likelihood	1635.4
AIC (smaller is better)	1641.4
AICC (smaller is better)	1641.5
BIC (smaller is better)	1638.7

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	2	56.4	10.06	0.0002
day	5	222	384.12	<.0001
trt*day	10	241	3.93	<.0001

Least Squares Means							
Effect	trt	day	Estimate	Standard Error	DF	t Value	Pr > t
trt*day	1	0	14.9450	0.9033	93.5	16.55	<.0001
trt*day	1	6	24.3300	0.9033	93.5	26.94	<.0001
trt*day	1	12	30.9000	0.9033	93.5	34.21	<.0001
trt*day	1	18	36.8950	0.9033	93.5	40.85	<.0001
trt*day	1	24	39.7600	0.9033	93.5	44.02	<.0001
trt*day	1	30	40.4450	0.9033	93.5	44.78	<.0001
trt*day	2	0	15.1250	0.9033	93.5	16.74	<.0001
trt*day	2	6	25.4550	0.9033	93.5	28.18	<.0001
trt*day	2	12	32.9800	0.9033	93.5	36.51	<.0001
trt*day	2	18	37.9700	0.9033	93.5	42.04	<.0001
trt*day	2	24	40.6850	0.9033	93.5	45.04	<.0001
trt*day	2	30	38.6050	0.9033	93.5	42.74	<.0001
trt*day	3	0	14.3950	0.9033	93.5	15.94	<.0001
trt*day	3	6	21.9850	0.9033	93.5	24.34	<.0001
trt*day	3	12	27.7050	0.9033	93.5	30.67	<.0001
trt*day	3	18	31.8900	0.9033	93.5	35.31	<.0001
trt*day	3	24	34.2500	0.9033	93.5	37.92	<.0001
trt*day	3	30	34.4200	0.9033	93.5	38.11	<.0001

Interpretation

The estimate of the autoregressive correlation in the “Covariance Parameter Estimates” table is 0.8175, the estimate of the pig-to-pig variability within a treatment is 1.7574, and the estimate of the residual variance is 14.5604. From the “Type 3 Tests of Fixed Effects” table we see that there is a significant day \times treatment interaction ($p < 0.0001$), so only the day by treatment means were requested. Pairwise comparisons of treatment within a day are of interest, but these are not displayed here.

8.4.2 Random Coefficient Analysis

A quadratic model that can be used to describe the weight on the k^{th} day of the j^{th} pig from the i^{th} treatment is

$$wt_{ijk} = a_{ij}^* + b_{ij}^* \text{day}_k + c_{ij}^* \text{day}_k^2 + e_{ijk}$$

$$i=1,2,3, \quad j=1,2,\dots,20, \quad k=1,2,\dots,6$$

The animals in the study represent a random sample of pigs from a population of pigs, so the quadratic model used to represent a pig's growth represents a realization from the population of quadratic models. This can be expressed through the following distributional assumptions:

$$\begin{bmatrix} a_{ij}^* \\ b_{ij}^* \\ c_{ij}^* \end{bmatrix} \sim N \left(\begin{bmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{bmatrix} \right) \text{ and } e_{ijk} \sim iid N(0, \sigma_e^2)$$

If you let $a_{ij}^* = \alpha_i + a_{ij}$, $b_{ij}^* = \beta_i + b_{ij}$, $c_{ij}^* = \gamma_i + c_{ij}$, the model can be expressed as

$$\begin{aligned} wt_{ijk} &= [\alpha_i + a_{ij}] + [\beta_i + b_{ij}] \text{day}_k + [\gamma_i + c_{ij}] \text{day}_k^2 + e_{ijk} \\ &= \alpha_i + \beta_i \text{day}_k + \gamma_i \text{day}_k^2 + a_{ij} + b_{ij} \text{day}_k + c_{ij} \text{day}_k^2 + e_{ijk} \\ i &= 1, 2, 3, \quad j = 1, 2, \dots, 20, \quad k = 1, 2, \dots, 6 \end{aligned}$$

and

$$\begin{bmatrix} a_{ij} \\ b_{ij} \\ c_{ij} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{bmatrix} \right) \text{ and } e_{ijk} \sim iid N(0, \sigma_e^2)$$

$\alpha_i + \beta_i \text{day}_k + \gamma_i \text{day}_k^2$ is the fixed effects part of the model, $a_{ij} + b_{ij} \text{day}_k + c_{ij} \text{day}_k^2$ is the random effects part of the model, and e_{ijk} is the residual part of the model. The following program fits the model above.

Program

```
proc mixed data=pigs;
  class trt pig;
  model weight=trt day day*trt day*day day*day*trt/ddfm=kr;
  random int day day*day / subject=pig(trt) type=un;

  lsmeans trt / at day=0 diff;
  lsmeans trt / at day=6 diff;
  lsmeans trt / at day=12 diff;
  lsmeans trt / at day=18 diff;
  lsmeans trt / at day=24 diff;
  lsmeans trt / at day=30 diff;
  ods output lsmeans=lsm diffs=diffs;
run;
```

The MODEL statement is used to specify the fixed effects part of the model, $\alpha_0 + \alpha_i + (\beta_0 + \beta_i) \text{day}_k + (\gamma_0 + \gamma_i) \text{day}_k^2$. This representation enables you to test the equality of the α_i , the equality of the β_i , and the equality of the γ_i . The RANDOM statement is used to specify the random effects part of the model, $a_{ij} + b_{ij} \text{day}_k + c_{ij} \text{day}_k^2$, where INT, DAY, and DAY \times DAY specify a_{ij} , b_{ij} , and c_{ij} , respectively, and TYPE=UN specifies the unstructured covariance matrix of the random coefficients. The LSMEANS statements provide predicted values of each of the treatments at the selected values of days. The DIFF option in the LSMEANS statements carries out the pairwise comparisons among the treatment means within a day.

Results

The results are given in Output 8.8 and Output 8.9.

Output 8.8 Results of Quadratic Random Coefficient Model

Estimated G Matrix						
Row	Effect	trt	pig	Col1	Col2	Col3
1	Intercept	1	1	1.6756	0.03795	0.003901
2	day	1	1	0.03795	0.000093	0.000366
3	day*day	1	1	0.003901	0.000366	2.752E-6

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	pig(trt)	1.6756
UN(2,1)	pig(trt)	0.03795
UN(2,2)	pig(trt)	0.000093
UN(3,1)	pig(trt)	0.003901
UN(3,2)	pig(trt)	0.000366
UN(3,3)	pig(trt)	2.752E-6
Residual		2.1562

Fit Statistics	
-2 Res Log Likelihood	1549.7
AIC (smaller is better)	1563.7
AICC (smaller is better)	1564.0
BIC (smaller is better)	1578.3

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	2	57	0.65	0.5280
day	1	57	4065.92	<.0001
day*trt	2	57	39.37	<.0001
day*day	1	57	1253.47	<.0001
day*day*trt	2	57	26.58	<.0001

Interpretation

The “Estimated G Matrix” table provides the estimated variances and covariances associated with individual pig regression models within a treatment,

$$\hat{\mathbf{G}} = \begin{bmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{ab} & \hat{\sigma}_{ac} \\ \hat{\sigma}_{ab} & \hat{\sigma}_b^2 & \hat{\sigma}_{bc} \\ \hat{\sigma}_{ac} & \hat{\sigma}_{bc} & \hat{\sigma}_c^2 \end{bmatrix} = \begin{bmatrix} \text{un}(1,1) & \text{un}(2,1) & \text{un}(3,1) \\ \text{un}(2,1) & \text{un}(2,2) & \text{un}(3,2) \\ \text{un}(3,1) & \text{un}(3,2) & \text{un}(3,3) \end{bmatrix}$$

$$= \begin{bmatrix} 1.6756 & 0.03795 & 0.0039 \\ 0.03795 & 0.000094 & 0.00037 \\ 0.0039 & 0.00037 & 2.752E-6 \end{bmatrix}$$

The “Type 3 Tests of Fixed Effects” table indicate that you would fail to reject $H_0: \alpha_1 = \alpha_2 = \alpha_3$ versus H_a : (not H_0) ($p = 0.5280$), reject $H_0: \beta_1 = \beta_2 = \beta_3$ versus H_a : (not H_0) ($p < 0.0001$ from the day \times trt source), and reject $H_0: \gamma_1 = \gamma_2 = \gamma_3$ versus H_a : (not H_0) ($p < 0.0001$ from the day \times day \times trt source). Thus, the population models for the three treatments could have a common intercept, different coefficients for DAY, and different coefficients for DAY \times DAY. The treatments can be compared at any value of DAY. The LSMEANS statements are used to provide estimates of the regression models at selected values of DAY. The results of the LSMEANS statements are given in Output 8.9.

Results

Output 8.9 Least-Squares Means and Pairwise Comparisons of Differences between the Quadratic Random Coefficient Models

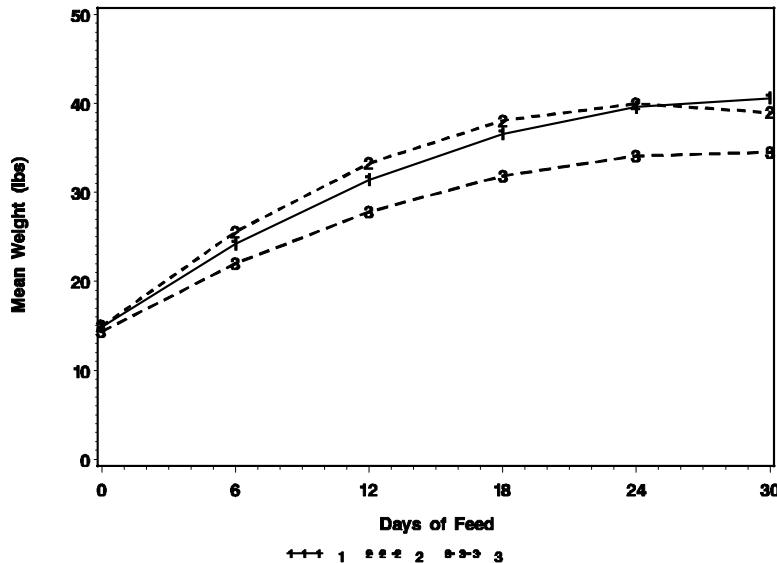
Least Squares Means							
Effect	trt	day	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	0.00	14.8993	0.4151	57	35.89	<.0001
trt	2	0.00	14.9589	0.4151	57	36.03	<.0001
trt	3	0.00	14.3536	0.4151	57	34.58	<.0001
trt	1	6.00	24.2017	0.4024	57	60.14	<.0001
trt	2	6.00	25.5676	0.4024	57	63.53	<.0001
trt	3	6.00	21.9951	0.4024	57	54.65	<.0001
trt	1	12.00	31.4213	0.5407	57	58.11	<.0001
trt	2	12.00	33.2731	0.5407	57	61.54	<.0001
trt	3	12.00	27.8317	0.5407	57	51.48	<.0001
trt	1	18.00	36.5580	0.7403	57	49.38	<.0001
trt	2	18.00	38.0754	0.7403	57	51.43	<.0001
trt	3	18.00	31.8633	0.7403	57	43.04	<.0001
trt	1	24.00	39.6119	0.9936	57	39.87	<.0001
trt	2	24.00	39.9745	0.9936	57	40.23	<.0001
trt	3	24.00	34.0899	0.9936	57	34.31	<.0001
trt	1	30.00	40.5829	1.3197	57	30.75	<.0001

Least Squares Means							
Effect	trt	day	Estimate	Standard Error	DF	t Value	Pr > t
trt	2	30.00	38.9704	1.3197	57	29.53	<.0001
trt	3	30.00	34.5114	1.3197	57	26.15	<.0001

Differences of Least Squares Means								
Effect	trt	_trt	day	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	0.00	-0.05964	0.5871	57	-0.10	0.9194
trt	1	3	0.00	0.5457	0.5871	57	0.93	0.3565
trt	2	3	0.00	0.6054	0.5871	57	1.03	0.3068
trt	1	2	6.00	-1.3659	0.5691	57	-2.40	0.0197
trt	1	3	6.00	2.2066	0.5691	57	3.88	0.0003
trt	2	3	6.00	3.5725	0.5691	57	6.28	<.0001
trt	1	2	12.00	-1.8519	0.7646	57	-2.42	0.0186
trt	1	3	12.00	3.5896	0.7646	57	4.69	<.0001
trt	2	3	12.00	5.4414	0.7646	57	7.12	<.0001
trt	1	2	18.00	-1.5174	1.0469	57	-1.45	0.1527
trt	1	3	18.00	4.6947	1.0469	57	4.48	<.0001
trt	2	3	18.00	6.2121	1.0469	57	5.93	<.0001
trt	1	2	24.00	-0.3626	1.4051	57	-0.26	0.7973
trt	1	3	24.00	5.5220	1.4051	57	3.93	0.0002
trt	2	3	24.00	5.8846	1.4051	57	4.19	<.0001
trt	1	2	30.00	1.6125	1.8663	57	0.86	0.3912
trt	1	3	30.00	6.0714	1.8663	57	3.25	0.0019
trt	2	3	30.00	4.4589	1.8663	57	2.39	0.0202

Interpretation

The least-squares means are estimates of the three treatment regression models evaluated at 0, 6, 12, 18, 24, and 30 days after initiation of the treatments. Those means are graphed in Figure 8.7. One of the problems with using the quadratic model is that there is an estimated decline in weight from days 24 to 30 for treatment 2, which is probably not real. What is important is that the selected model adequately describes the data in the range of the data and the quadratic models seem to do a good job. The next section presents a test for lack of fit of the quadratic regression models. The “Differences of Least Squares Means” table provides pairwise comparisons among the three treatment models evaluated at a specific number of days. There are no significant differences among the three models at day 0 (as one would expect), all means are significantly different at day 6, the means of treatments 1 and 2 are not significantly different for days 12 to 30, but treatments 1 and 2 means are significantly different from the mean of treatment 3 for days 12 to 30. These conclusions can be visualized by the graph in Figure 8.7.

Figure 8.7 Models for the Pig Growth Data

8.4.3 Test of Lack of Fit

The final step in this analysis is to evaluate whether or not the quadratic models adequately describe the data. The variable DAY needs to be used as a class variable and as a continuous variable in the same model specification, so the first step is to generate a new variable XDAY=DAY. Then, construct a model with the desired regression part of the model and include the term XDAY*TRT, where both XDAY and TRT are in the CLASS statement. If you request the Type 1 analysis using the option HTYPE=1, then the *F*-statistic associated with the XDAY*TRT source provides the statistic to test the lack of fit of the regression model to describe the DAY*TRT means.

Program

```

data pigs; set pigs;
  xday=day;
run;
proc mixed data=pigs;
  class trt pig xday;
  model weight=trt day day*trt day*day day*day*trt xday*trt/
    ddfm=kr htype=1;
  random int day day*day/subject=pig(trt) type=un;
run;

```

Results

The results of the lack of fit test are given in Output 8.10.

Output 8.10 Test for Lack of Fit of Quadratic Random Coefficient Model

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	pig(trt)	1.7047
UN(2,1)	pig(trt)	0.03447
UN(2,2)	pig(trt)	0.000808
UN(3,1)	pig(trt)	0.003988
UN(3,2)	pig(trt)	0.000344
UN(3,3)	pig(trt)	3.483E-6
Residual		2.1208

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	2	57	11.56	<.0001
day	1	57	1292.49	<.0001
day*trt	2	57	6.38	0.0032
day*day	1	57	1253.47	<.0001
day*day*trt	2	57	26.58	<.0001
trt*xday	9	171	1.33	0.2226

Interpretation

The “Type 1 Tests of Fixed Effects” table provides sequential tests for the effects in the model. The question addressed by the source labeled “trt*xday” is: given all previously listed (=all other) terms in the model, do they describe all of the pattern in the “trt*xday” means or is a more complex model more appropriate? In this case, the *p*-value is 0.2226, indicating that the quadratic regression model is adequate to describe the xday \times trt means. A small *p*-value would indicate that the model does not adequately describe the data. The analysis can also be carried out using the Type 3 Tests of Fixed Effects, but in that table the sources for all terms except trt \times xday have zero degrees of freedom and zero *F*-statistics. This occurs because the Type 3 Tests are adjusted for all other terms in the model. So, when trt \times xday is in the model, all of the variation among the means is described, and there is no variability left for other terms in the model.

A nonlinear model would most likely be a better representation of the data, but that model was not considered here.

8.5 Summary

The data structure appropriate for this chapter consists of repeated measurements on independent subjects. The random coefficients model for such data is effectively an analysis of covariance for each subject; however, the coefficients from these regression models are assumed to have arisen from a normal probability distribution. Three examples were used to illustrate the important features of the model: one where the subjects are treatments, another where the subjects are students and a repeated measures growth study. Only linear random coefficient models were discussed in this chapter, but nonlinear random coefficient models are quite useful, as in the growth study in Section 8.4. Nonlinear random coefficient models can be fit using PROC NLMIXED, which is described in Chapter 15.



Heterogeneous Variance Models

9.1	Introduction	344
9.2	Example: Two-Way Analysis of Variance with Unequal Variances.....	345
9.2.1	Testing for Homogeneity of Variances	345
9.2.2	Fitting a Model with Heterogeneous Variances	350
9.2.3	Fitting a Model with Reduced Heterogeneous Variance Structure	352
9.3	Example: Simple Linear Regression Model with Unequal Variances ...	354
9.3.1	Fit Power-of-X Model Using REML	355
9.3.2	Fit Power-of-X Model Using ML	357
9.3.3	Fit Power-of-X Model Using ML from PROC NLMIXED	358
9.3.4	Fit Power-of-the-Mean Model Using REML.....	359
9.3.5	Fit Power-of-the-Mean Model Using ML from PROC MIXED	363
9.3.6	Fit Power-of-the-Mean Model Using ML from PROC NLMIXED	364
9.3.7	Comparison of Power-of-X and Power-of-the-Mean Model Covariance Structures	365
9.4	Example: Nested Model with Unequal Variances for a Random Effect	366
9.4.1	Model with Equal Variances	366
9.4.2	Testing for Equal District Variances across Regions	369
9.4.3	Test for Equal Residual Variances.....	371
9.4.4	Fitting Model with Unequal Random Effect Variances	372
9.5	Example: Within-Subject Variability	374
9.5.1	Basic Unstructured Covariance Model	375
9.5.2	Other Covariance Structures	379
9.5.3	Selecting a Covariance Structure.....	381
9.5.4	Power-of-the-Mean Model	386
9.6	Example: Combining Between- and Within-Subject Heterogeneity	393
9.6.1	Basic Unstructured Covariance Model	394
9.6.2	Incorporating Between-Subject Heterogeneity	396
9.6.3	Heterogeneous Random Coefficients	400

9.7 Example: Log-Linear Variance Models	402
9.7.1 Initial Model	403
9.7.2 Full Model with Power-of-X Dispersion Effects.....	404
9.7.3 Reduced Model with Power-of-X Dispersion Effects.....	407
9.7.4 Model with UN Repeated Measures Error Structure	408
9.8 Summary	411

9.1 Introduction

Heterogeneity of variances occurs in many situations and, with the development of PROC MIXED and PROC NLMIXED models with heterogeneous variances, can be fit to data sets enabling an analysis with unequal variances. Unequal variances can occur with a set of treatments where each treatment has its own variance. Unequal variances can occur at any set of levels of a random effect. Models to describe the variance as a function of independent variables in a regression model can be fit to data where the variance increases or decreases as the values of the independent variables change. One of the great advantages of the likelihood-based estimation approach to mixed models is the ability to fit a variety of covariance structures. This chapter considers structures modeling variability that changes across the data—that is, variability that is **heterogeneous**.

Heterogeneous variances occur in all types of data, yet many analysts do not consider them or fail to adequately account for them in their statistical inferences. A primary motivation for modeling heterogeneous variances is the ability to appropriately downweight portions of your data that are highly variable and extract more information from portions of your data that are more precise. Failure to account for heterogeneity when it is present can lead to inefficient and possibly misleading inferences about fixed effects in your model.

Heterogeneous variance models can be placed into two categories: **within-subject** or **R**-side of the covariance parameters, and **between-subject** or **G**-side of the covariance parameters. Within-subject heterogeneity occurs across data from the same subject. A typical example is variances that increase with time in a longitudinal data setting or where the variances of the residuals are different for different levels of a factor or combination of factors or where the variances of the residuals are a function of the mean of the model or of independent variables. Between-subject heterogeneity occurs when different groups of subjects display different variance patterns but are homogeneous within groups or when the variance components corresponding to a random effects part of the model are unequal for levels of another factor. Heterogeneous variances can be incorporated into the analysis by specifying different variances or covariance structures for different levels of a factor or combination of levels of factors or by specifying the variance as a function of a set of fixed effects or other independent variables.

This chapter discusses these types of heterogeneity in the context of six examples. Refer to Wolfinger (1996) for additional examples and discussion.

9.2 Example: Two-Way Analysis of Variance with Unequal Variances

This study was conducted to evaluate the effect of having a television set operating in a play area of a day-care facility. The facility had 27 children ages 2 to 4. The play activity of the children was videotaped during a three-hour session after lunch one afternoon. The videotapes were viewed by the researcher to determine the total number of minutes that each child stopped play activities and watched the television. The researcher was interested in evaluating the effect of age and gender on the amount of time the TV was watched. The data are given as Data Set 9.2, "TV," in Appendix 2, "Data Sets."

An equal variance model that can be used to describe these data is

$$\begin{aligned} \text{time}_{ijk} &= \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + e_{ijk} \\ i &= 1, 2, 3 \\ j &= 1, 2 \\ k &= 1, 2, \dots, n_{ij} \\ e_{ijk} &\sim \text{iid } N(0, \sigma_e^2) \end{aligned}$$

9.2.1 Testing for Homogeneity of Variances

The following program uses the facilities of PROC GLM to provide a test of the equality of variances. When the model consists of a single term, the HOVTEST option in the MEANS statement provides a test of the equality of variances across the combinations of levels of the single term in the model. The possible tests for homogeneity of variances are Bartlett's, Levene's (absolute value and squared residuals), Brown-Forsythe's, and O'Brien's. An additional option is to request WELCH, which provides a test of the equality of the means across the levels of the single term using Welch's statistic, which incorporates the unequal variances.

For this example, it is of interest to determine if the variances for the age \times sex combinations are equal; thus the single term in the model is age \times sex.

Program

```
proc glm data=tv;
  class age sex;
  model time=age*sex;
  means age*sex / hovtest=levene(type=abs) welch;
run;
```

Results

Output 9.1 Levene's Test Using PROC GLM for Time Watching TV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1979.129630	395.825926	9.47	<.0001
Error	21	877.833333	41.801587		
Corrected Total	26	2856.962963			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age*sex	5	1979.129630	395.825926	9.47	<.0001

Levene's Test for Homogeneity of time Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
age*sex	5	235.2	47.0437	5.08	0.0033
Error	21	194.5	9.2603		

Welch's ANOVA for time			
Source	DF	F Value	Pr > F
age*sex	5.0000	8.28	0.0039
Error	8.6668		

Level of age	Level of sex	N	time	
			Mean	Std Dev
2	F	4	9.0000000	2.1602469
2	M	6	12.3333333	1.9663842
3	F	5	16.0000000	3.6742346
3	M	4	19.5000000	3.5118846
4	F	4	26.2500000	11.7862915
4	M	4	35.2500000	10.5948101

Interpretation

The first two parts of Output 9.1 contain the analysis of variance for time using the equal variance model. The third part contains the results of Levene's test for equality of variances, where the *p*-value is 0.0033, indicating that there is a lot of evidence against the equal variance hypothesis. The fourth part of the output contains the results of the Welch test of equality of the age \times sex means using the unequal variance model. The last part provides the means and standard deviations for the age \times sex combinations.

Levene's test used above is obtained by performing an analysis of variance on the absolute values of the residuals. The residuals are included in the data set R obtained by using the following PROC MIXED code with the OUTP=R option in the MODEL statement. The variable RESID contains the residual for each of the observations.

Program

```
proc mixed data=tv ic;
  class age sex;
  model time = sex|age / outp=R;
  lsmeans age sex / diff adjust=Tukey;
run;
```

Results

Output 9.2 Equal Variance Analysis of the Data for Time Watching TV

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	41.8016

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
146.9	1	148.9	149.1	149.2	150.0	151.0

Type 3 Tests of Fixed Effects					
Effect	Num DF	Den DF	F Value	Pr > F	
sex	1	21	4.39	0.0485	
age	2	21	21.35	<.0001	
age*sex	2	21	0.52	0.6011	

Least Squares Means							
Effect	sex	age	Estimate	Standard Error	DF	t Value	Pr > t
age		2	10.6667	2.0867	21	5.11	<.0001
age		3	17.7500	2.1686	21	8.19	<.0001
age		4	30.7500	2.2859	21	13.45	<.0001
sex	F		17.0833	1.8031	21	9.47	<.0001
sex	M		22.3611	1.7597	21	12.71	<.0001

Differences of Least Squares Means											
Effect	sex	age	_sex	_age	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
age		2		3	-7.0833	3.0095	21	-2.35	0.0284	Tukey-Kramer	0.0700
age		2		4	-20.0833	3.0951	21	-6.49	<.0001	Tukey-Kramer	<.0001
age		3		4	-13.0000	3.1509	21	-4.13	0.0005	Tukey-Kramer	0.0013
sex	F		M		-5.2778	2.5195	21	-2.09	0.0485	Tukey-Kramer	0.0485

Interpretation

The estimate of the variance from the equal variance model is 41.8, as shown in the “Covariance Parameter Estimates” table. The “Type 3 Tests for Fixed Effects” table does not provide evidence of an interaction between the levels of age and levels of sex, but there are significant age and sex effects. The least-squares means are provided for the levels of age and

the levels of sex as well as all pairwise differences with Tukey-Kramer adjustment for multiplicity.

One way to investigate the assumptions of the model is to look at plots of the residuals for each of the treatment combinations. The following program provides a set of box plots of the residuals for each age \times sex combination using the ODS Graphics capabilities of PROC MIXED.

Program

```
ods html;
ods graphics on;

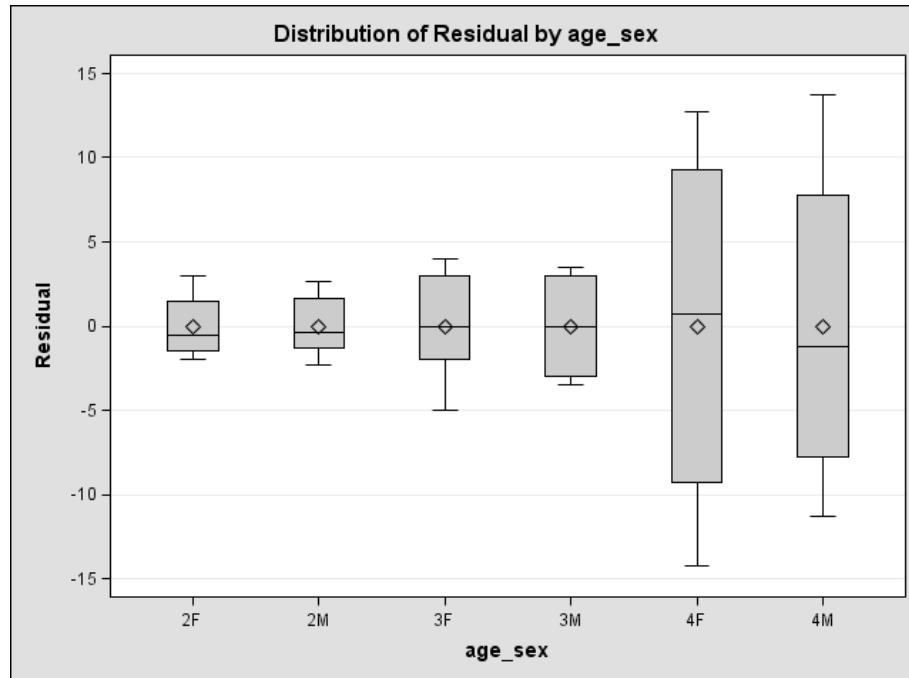
proc mixed data=tv boxplot;
  class age_sex;
  model time = age_sex;
run;

ods graphics off;
ods html close;
```

Interpretation

The set of box plots are shown in Figure 9.1. They indicate that the variances are increasing as age increases.

Figure 9.1 Box Plots of Residuals for the Age \times Sex Combinations



Levene's test for equality of variances is obtained by carrying out an analysis of variance on the absolute values of the residuals, as shown in the following program where the DATA step is used to compute the absolute values.

Program

```
data R1; set R; **Compute the absolute value of residuals;
    ABSR = ABS(RESID);
run;
proc glm data=R1;
    class age sex;
    model ABSR = age*sex;
run;
```

Results

Output 9.3 Levene's Test for Equality of Variances Using Analysis of Variance on the Absolute Values of the Residuals

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	235.2185185	47.0437037	5.08	0.0033
Error	21	194.4666667	9.2603175		
Corrected Total	26	429.6851852			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age*sex	5	235.2185185	47.0437037	5.08	0.0033

Interpretation

The value of the *F*-statistic is 5.08 with a *p*-value of 0.0033, the same as in Output 9.1.

An extension of Levene's test is to carry out a factorial effects analysis of variance on the absolute values of the residuals in an attempt to discover a simpler pattern among the variances. The following program provides an analysis of variance with sex, age, and age \times sex in the model.

Program

```
proc glm data=R1;
    class age sex;
    model ABSR=age sex age*sex;
run;
```

Results

Output 9.4 Levene's Test for Equality of Variances Using a Factorial Effects Analysis of Variance on the Absolute Values of the Residuals

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	235.2185185	47.0437037	5.08	0.0033
Error	21	194.4666667	9.2603175		
Corrected Total	26	429.6851852			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age	2	228.4774049	114.2387025	12.34	0.0003
sex	1	0.9398374	0.9398374	0.10	0.7532
age*sex	2	3.9539150	1.9769575	0.21	0.8095

Interpretation

The analysis of variance in Output 9.4 provides evidence that the variances are unequal for the levels of age, but there is no influence of sex on the variances.

9.2.2 Fitting a Model with Heterogeneous Variances

The result of Levene's test for homogeneity of variances indicates that the variances are not equal. A model with unequal variances, residual variances in this case, can be specified in PROC MIXED by using the REPEATED statement with the GROUP= option. In this case, using GROUP=age*sex specifies a different residual variance for each combination of age and sex. A model with unequal variances that can be used to describe the data is

$$\text{time}_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + e_{ijk}$$

$$i = 1, 2, 3$$

$$j = 1, 2$$

$$k = 1, 2, \dots, n_{ij}$$

$$e_{ijk} \sim iid N(0, \sigma_{ij}^2)$$

The difference between this model and the independent errors model in Section 9.2 is that there is a variance for each of the age \times sex combinations. The following program fits a model with six variances.

Program

```
proc mixed data=TV covtest cl ic;
  class age sex;
  model time = sex|age / ddfm=KR outp=R;
  repeated / group=sex*age;
  lsmeans age sex / diff adjust=Tukey;
run;
```

Results

Output 9.5 Results from Fitting Two-Way Analysis of Variance Model with Unequal Variances for Each Combination of Sex and Age

Covariance Parameter Estimates								
Cov Parm	Group	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
Residual	age*sex 2 F	4.6667	3.8103	1.22	0.1103	0.05	1.4976	64.8763
Residual	age*sex 2 M	3.8667	2.4455	1.58	0.0569	0.05	1.5066	23.2592
Residual	age*sex 3 F	13.5000	9.5459	1.41	0.0786	0.05	4.8460	111.47
Residual	age*sex 3 M	12.3333	10.0701	1.22	0.1103	0.05	3.9579	171.46

Covariance Parameter Estimates								
Cov Parm	Group	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
Residual	age*sex 4 F	138.92	113.42	1.22	0.1103	0.05	44.5798	1931.23
Residual	age*sex 4 M	112.25	91.6517	1.22	0.1103	0.05	36.0222	1560.51

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
126.8	6	138.8	144.8	141.1	146.6	152.6

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
sex	1	7.39	3.56	0.0988
age	2	9.58	22.23	0.0003
age*sex	2	9.58	0.23	0.7969

Least Squares Means							
Effect	sex	age	Estimate	Standard Error	DF	t Value	Pr > t
age		2	10.6667	0.6729	6.11	15.85	<.0001
age		3	17.7500	1.2024	6.7	14.76	<.0001
age		4	30.7500	3.9621	5.93	7.76	0.0003
sex	F		17.0833	2.0709	3.68	8.25	0.0017
sex	M		22.3611	1.8794	3.8	11.90	0.0004

Differences of Least Squares Means											
Effect	sex	age	_sex	_age	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
age		2		3	-7.0833	1.3779	10.4	-5.14	0.0004	Tukey-Kramer	0.0013
age		2		4	-20.0833	4.0188	6.28	-5.00	0.0022	Tukey-Kramer	0.0016
age		3		4	-13.0000	4.1405	7.02	-3.14	0.0163	Tukey-Kramer	0.0272
sex	F		M		-5.2778	2.7965	7.39	-1.89	0.0988	Tukey-Kramer	0.0988

Interpretation

The “Covariance Parameter Estimates” table in Output 9.5 contains the variances of the observations for each of the age \times sex combinations. The number of degrees of freedom associated with an estimate of a variance is computed as $df = 2 \times (z\text{-score})^2$. The estimates of the variances range from 3.8667 to 138.92, so it is no surprise that Levene’s test is significant ($p = 0.0033$). The AIC in the “Information Criteria” table is 138.8, which is much smaller than the value of 148.9 obtained from the equal variance model in Output 9.2. This indicates that the unequal variance model fits the data better than the equal variance model. The “Type 3 Tests of Fixed Effects” table provides tests for equal sex means, for equal age means, and for no interaction among the levels of sex and the levels of age using the unequal variance covariance structure, thus providing tests similar to the Welch’s test. The denominator degrees of freedom

for each effect are obtained from the generalization of the Satterthwaite approximation (see Appendix 1, “Linear Mixed Models Theory”). There is a highly significant age effect ($p = 0.0003$), a marginal sex effect ($p = 0.0988$), and no indication of an age \times sex interaction ($p = 0.7969$).

The estimates of the standard errors of the least-squares means is a combination of those variances involved in that mean, and thus the degrees of freedom are approximated using the Satterthwaite method as indicated by the decimal points. The differences among the age means and between the sex means are displayed in the “Differences of Least Squares Means” table. The estimated standard errors of the differences are combinations of the variances, and thus the degrees of freedom are approximated. The ADJUST=TUKEY option was used to provide error rate protection for carrying out multiple tests. There are significant differences ($p < 0.05$) among all of the age means, and the sex means are significantly different at $p = 0.0988$.

9.2.3 Fitting a Model with Reduced Heterogeneous Variance Structure

The analysis in Output 9.4 indicates there are differences in the variances for the levels of age, but there are no differences in the variances for the two levels of sex, nor is there an interaction. Thus, a model with unequal variances for each age was fit using the GROUP=AGE option in the REPEATED statement.

Program

```
proc mixed data=TV ic;
  class age sex;
  model time=sex|age/DDFM=KR OUTP=R;
  repeated / group=age;
  lsmeans age sex / diff adjust=Tukey;
run;
```

Results

The results are given in Output 9.6.

Output 9.6 Results from Fitting Two-Way Analysis of Variance Model with Unequal Variances for Each Age

Covariance Parameter Estimates		
Cov Parm	Group	Estimate
Residual	age 2	4.1667
Residual	age 3	13.0000
Residual	age 4	125.58

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
126.9	3	132.9	134.3	134.1	136.8	139.8

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
sex	1	7.48	3.56	0.0984
age	2	9.84	22.36	0.0002
age*sex	2	9.84	0.23	0.7962

Least Squares Means							
Effect	sex	age	Estimate	Standard Error	DF	t Value	Pr > t
age		2	10.6667	0.6588	8	16.19	<.0001
age		3	17.7500	1.2093	7	14.68	<.0001
age		4	30.7500	3.9621	6	7.76	0.0002
sex	F		17.0833	1.9731	7.42	8.66	<.0001
sex	M		22.3611	1.9816	7.53	11.28	<.0001

Differences of Least Squares Means											
Effect	sex	age	_sex	_age	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
age		2		3	-7.0833	1.3771	10.9	-5.14	0.0003	Tukey-Kramer	0.0012
age		2		4	-20.0833	4.0165	6.33	-5.00	0.0021	Tukey-Kramer	0.0015
age		3		4	-13.0000	4.1425	7.12	-3.14	0.0161	Tukey-Kramer	0.0265
sex	F		M		-5.2778	2.7964	7.48	-1.89	0.0984	Tukey-Kramer	0.0984

Interpretation

The advantage of using a simpler covariance structure is that there are more degrees of freedom available for estimating the standard errors of the fixed effects, thus providing more powerful tests. The “Covariance Parameter Estimates” table in Output 9.6 provides estimates of the three variances that range from 4.1667 to 125.58. The value of AIC from the “Information Criteria” table is 132.9, indicating that this simpler covariance structure model is adequately describing the data as the AIC for the model with six unequal variances, is 138.8. The “Type 3 Tests of Fixed Effects” table provides tests for equal sex means, for equal age means, and for no interaction among the levels of sex and the levels of age using the unequal variance covariance structure. There is a highly significant age effect ($p = 0.0002$), a marginal sex effect ($p = 0.0984$), and no indication of an age \times sex interaction ($p = 0.7962$). These significance levels are a little smaller than those in Output 9.5, indicating the advantage of the simpler covariance structure. The “Least Squares Means” table provides the means for the levels of age and for the levels of sex. Note that the degrees of freedom associated with the age means are integers, an indication that only one variance was used in computing that estimated standard error and no approximation for degrees of freedom is needed. That is not the case for the sex means. The estimated standard errors of the differences of means involve more than one variance component, and thus the number of degrees of freedom associated with each is approximated. If the age \times sex means were obtained, the comparisons of the sex means within a level of age would depend only on one variance, whereas comparisons of the levels of age within a gender would involve a combination of variances; the degrees of freedom would then be approximated. The adjusted p -values are similar to those in Output 9.5, albeit a bit smaller in Output 9.6.

9.3 Example: Simple Linear Regression Model with Unequal Variances

This example is used to demonstrate how to use PROC MIXED and PROC NLMIXED to fit a simple linear regression model that uses a function of the independent variable to describe the variance. The simple linear regression model can be expressed as

$$Y_i = \alpha + x_i \beta + e_i \\ i = 1, 2, \dots, n$$

where Y_i is the response variable, x_i is the independent variable, α is the intercept, β is the slope of the model, and e_i is the random error. The usual assumptions about the errors are that they are independently distributed with common variance, $e_i \sim iid N(0, \sigma^2_e)$. PROC MIXED has two built-in variance functions that can be used to describe the variability in the errors as a function of the independent variables. The Power-of-X (Carroll and Ruppert 1988) dispersion function specifies the variance of an error observed at x as $\text{Var}[e_i] = \sigma^2 \exp\{x_i \gamma\}$, where σ^2 is an unknown parameter and γ is an unknown parameter called the **dispersion effects parameter**. Other Power-of-X variance functions can be specified such as

$$\text{Var}[e_i] = \sigma^2 \exp\{x_i \gamma + x_i^2 \delta + \cos(x_i) \lambda\}$$

providing a lot of flexibility. The LOCAL=EXP(X) option in the REPEATED statement fits the variance function $\text{Var}[e_i] = \sigma^2 \exp\{x_i \gamma\}$; i.e., the errors are assumed to be distributed independently with mean 0 and variance $\sigma^2 \exp\{x_i \gamma\}$. This example consists of measuring y and x on each experimental unit and then fitting the simple linear regression model. If you measured y , x , u , and v , you could fit the simple linear regression model $Y_i = a + \beta x_i + e_i$, $i = 1, 2, \dots, n$, and use a variance function that depends on x or u or v or any combination such as

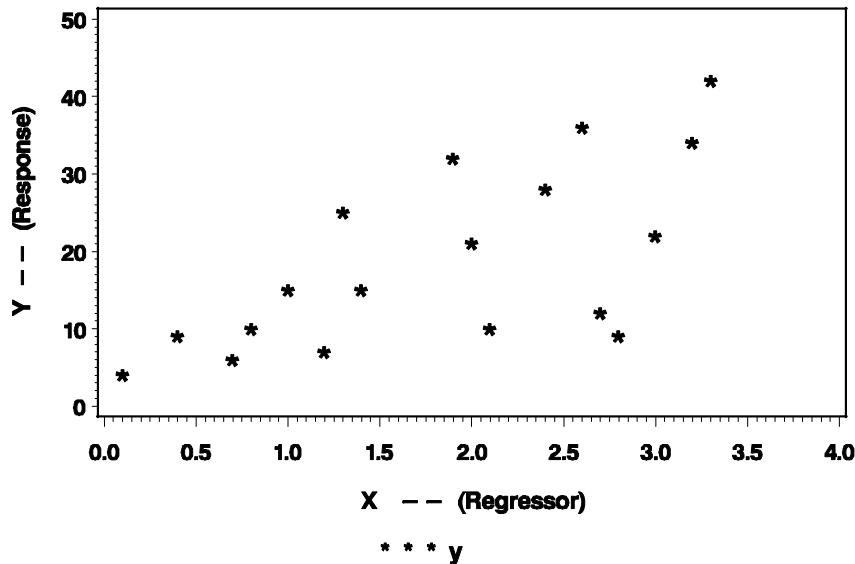
$$\text{Var}[e_i] = \sigma^2 \exp\{x_i \gamma + u_i \phi + v_i \theta + x_i u_i v_i \omega\}$$

The second built-in variance function is called the Power-of-the-Mean function, which specifies the variance of the error as

$$\text{Var}[e_i] = \sigma^2 |\alpha + x_i \beta|^\theta$$

where σ^2 is an unknown parameter and θ is the unknown power parameter.

The data for this example are given as Data Set 9.3, “LR,” in Appendix 2, “Data Sets,” where Y is the response and X is the independent variable. Figure 9.2 shows a scatter plot of the data; the increase in the variance of the responses with x is clearly visible.

Figure 9.2 Scatter Plot of Linear Regression Data

9.3.1 Fit Power-of-X Model Using REML

The following program uses PROC MIXED to fit the simple linear regression model with the Power-of-X dispersion using REML, i.e., $Y_i = a + \beta x_i + e_i$, $i = 1, 2, \dots, n$, with independent errors and $\text{Var}[e_i] = \sigma^2 \exp\{x_i \gamma\}$.

Program

```
proc mixed data=LR;
  model y = x / outp=preml outpm=pm_reml solution;
  repeated / local=exp(x);
  parms (1.0) (4.3);
run;
```

The PARMS (1.0) (4.3) statement is used to provide starting values for γ and σ^2 . The REPEATED statement is used with the LOCAL=EXP(X) option to specify the dispersion part of the model.

Results

The results are given in Output 9.7.

Output 9.7 Results for Fitting Power-of-X Dispersion with the Simple Linear Regression Model Using REML

Covariance Parameter Estimates	
Cov Parm	Estimate
EXP x	1.0907
Residual	7.6071

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.8940	2.1263	16	1.83	0.0857
x	8.2889	1.7642	16	4.70	0.0002

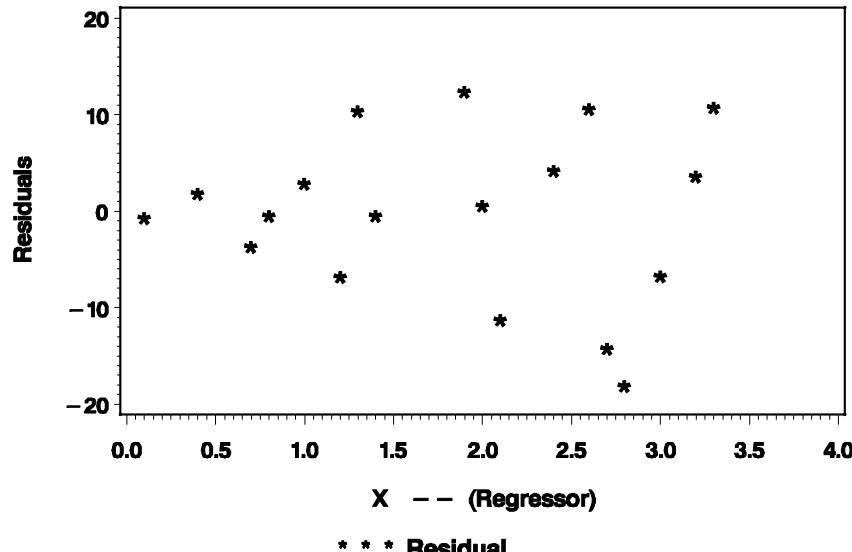
Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
x	1	16	22.07	0.0002

Interpretation

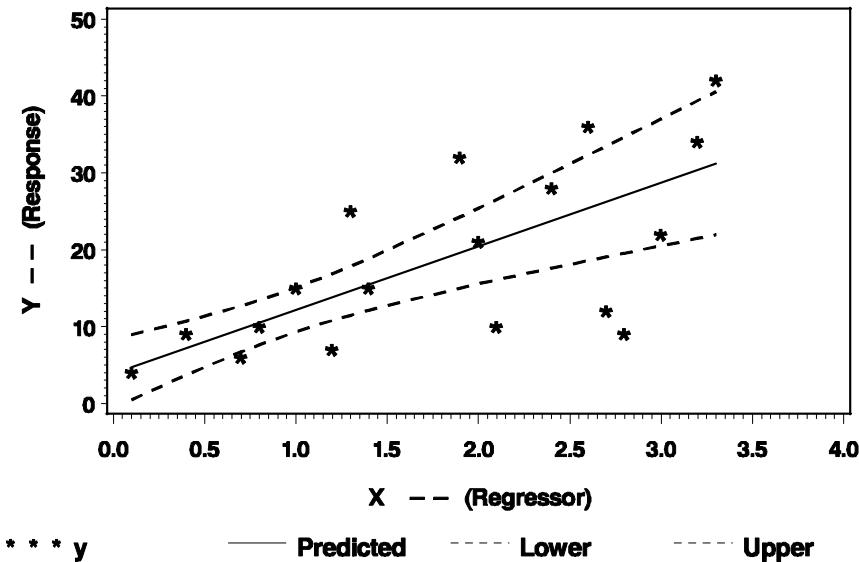
The estimates of the parameters in Output 9.7 from the “Covariance Parameter Estimates” and the “Solution for Fixed Effects” table are $\hat{\alpha} = 3.8940$, $\hat{\beta} = 8.2889$, $\hat{\sigma}^2 = 7.6071$, and $\hat{\gamma} = 1.0907$.

The OUTP=preml option in the MODEL statement was used to provide a data set with predicted values (PRED), residuals (RESID), and lower and upper 95% confidence limits (LOWER and UPPER). A scatter plot of the residuals is shown in Figure 9.3. The dispersion of the residuals increases as the value of x increases.

Figure 9.3 Scatter Plot of Residuals



The data points, the estimated model, and the upper and lower 95% confidence intervals are displayed in Figure 9.4. The widths of the confidence intervals are much larger for large values of x than for small values, showing the effect of the Power-of-X dispersion function.

Figure 9.4 Graph of Data with Estimated Model and 95% Confidence Intervals

9.3.2 Fit Power-of-X Model Using ML

The maximum likelihood (ML) method for estimating the parameters of the model is used to provide a comparison with the results obtained from PROC NLMIXED (Section 9.3.3). The following program requests that the ML method be used to estimate the parameters of the model.

Program

```
proc mixed data=LR method=ml;
  model y = x / outp=pml outpm=pm_ml solution;
  repeated / local=exp(x);
  parms (1.0) (4.3);
run;
```

Results

The results of the maximum likelihood estimation process are given in Output 9.8.

Output 9.8 Results for Fitting Power-of-X Dispersion with the Simple Linear Regression Model Using METHOD=ML

Covariance Parameter Estimates	
Cov Parm	Estimate
EXP x	1.2097
Residual	5.4205

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.8115	1.8659	16	2.04	0.0579
x	8.3809	1.6508	16	5.08	0.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
x	1	16	25.78	0.0001

Interpretation

The maximum likelihood estimates of the parameters in Output 9.8 from the “Covariance Parameter Estimates” and the “Solution for Fixed Effects” are $\hat{\alpha} = 3.8115$, $\hat{\beta} = 8.3809$, $\hat{\sigma}^2 = 5.4205$, and $\hat{\gamma} = 1.2097$. The estimates of the intercept and slope are similar to those obtained using the REML estimates of the variance parameters. The estimate of σ^2 is smaller than that using REML. The estimates of the standard errors of the parameter estimates are smaller when using ML than when using REML. It is generally known that estimates of standard errors of fixed effects based on ML estimates of the variance components are smaller than those based on REML.

9.3.3 Fit Power-of-X Model Using ML from PROC NLMIXED

PROC NLMIXED is a procedure that provides maximum likelihood estimates of a model’s parameters where the model can be nonlinear in the parameters and a variance function can be specified. The PROC NLMIXED code consists of (1) initial values for the parameters in a PARMS statement, (2) providing the expression for the mean of the regression model with the MEAN statement, and (3) providing the distribution of the data with variance function using the MODEL statement. The following program is used to fit the simple linear regression model with the Power-of-X dispersion function. Chapter 15 contains a detailed discussion of PROC NLMIXED.

Program

```
proc nlmixed data=LR;
  parms a=4.5 b=7.5 c=.01 sig2=5;
  mean = a+b*x;
  model y ~ normal(mean,sig2*exp(c*x));
  predict mean out=mean df=16;
run;
```

Results

The results of using NLMIXED are given in Output 9.9.

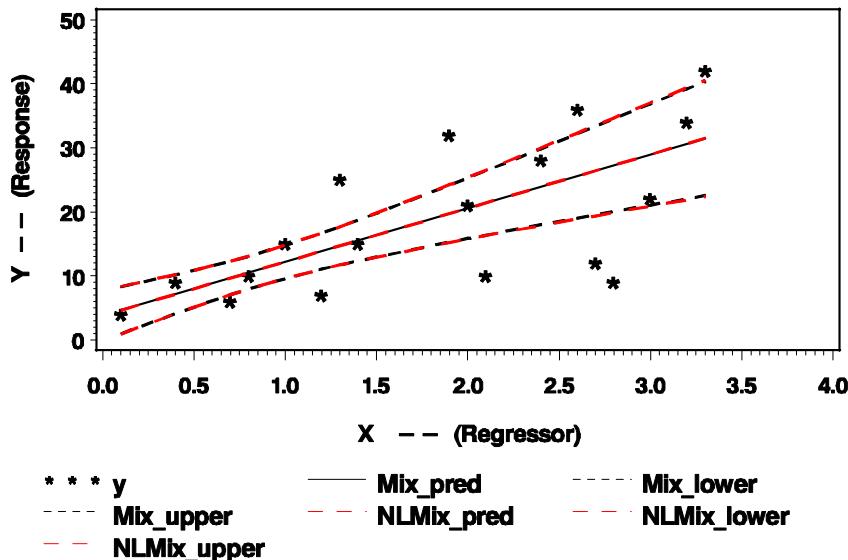
Output 9.9 Results for Fitting Power-of-X Dispersion with the Simple Linear Regression Model Using PROC NLMIXED

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a	3.8056	1.8834	18	2.02	0.0585	0.05	-0.1512	7.7624	-0.00037
b	8.3871	1.6870	18	4.97	<.0001	0.05	4.8428	11.9314	0.000395
c	1.2152	0.4638	18	2.62	0.0173	0.05	0.2408	2.1896	0.001242
sig2	5.3581	4.8670	18	1.10	0.2854	0.05	-4.8671	15.5832	-0.00243

Interpretation

The results in Output 9.9 provide the estimates of the model parameters from PROC NLMIXED, which are $\hat{\alpha} = 3.8056$, $\hat{\beta} = 8.3871$, $\hat{\sigma}^2 = 5.3581$, and $\hat{\gamma} = 1.2152$. The PREDICT statement requests that PROC NLMIXED construct a data set with predicted values of the “mean” as well as the estimated standard errors and 95% confidence intervals. Those confidence intervals are based on 16 degrees of freedom. Figure 9.5 is an overlay plot of the predicted means, confidence intervals, and data for the results from PROC MIXED with METHOD=ML and from PROC NLMIXED. The plots are essentially identical for both methods, thus demonstrating that the two procedures are providing the same results.

Figure 9.5 Overlay of ML Estimates from PROC MIXED and PROC NLMIXED Using EXP



9.3.4 Fit Power-of-the-Mean Model Using REML

The next variance function to be used to model this linear regression data is the Power-of-the-Mean function. The simple linear regression model is $Y_i = \alpha + \beta x_i + e_i$, $i = 1, 2, \dots, n$, where it is assumed the errors are independent with variances $\text{Var}[e_i] = \sigma^2 |\alpha + x_i \beta|^\theta$, where σ^2 is an unknown scale parameter and θ is the unknown power parameter.

In order to fit this Power-of-the-Mean model with PROC MIXED, you must first create a data set that contains estimates of the fixed effects, i.e., estimates of α and β . You can obtain the needed estimates by using ordinary least squares—i.e., by using a model that assumes the errors are identically independently distributed. The following program provides estimates of α and β assuming $e_{ij} \sim iid N(0, \sigma_e^2)$.

Program

```
proc mixed data=LR method=reml;
  model y = x / solution;
  ods output solutionf=est;
run;
```

Results

The results are given in Output 9.10, which consists of the independent errors analysis.

Output 9.10 Results for Fitting Independent Errors for the Simple Linear Regression Model Using PROC NLINMIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	80.5358

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	4.4976	4.4999	16	1.00	0.3324
x	7.7825	2.1730	16	3.58	0.0025

Interpretation

The results in Output 9.10 provide the estimates of the intercept and slope in the “Solution for Fixed Effects” table. These estimates are put into a data set using the ODS OUTPUT SOLUTIONF=EST statement. Starting values of the variance components can also be obtained as follows. The estimate of the variance from Output 9.10 is 80.5358. From the plots of the data and residuals, it is possible that the variance model is $\text{Var}[e_i] = \sigma^2 |\alpha + x_i \beta|^2$. The predicted value of the model toward the middle of the data, $x = 2$, is 20.0626. Set the variance equal to $\sigma^2 = 80.5358/(20.0626)^2$, which on solving provides $\sigma^2 = 0.200$. The starting values for the covariance parameters can be inserted into the PARMS statement through a data set as PARMS/PDATA=<data set name>. The following program generates the data set with the covariance parameter estimates, called COV. The Power-of-the-Mean covariance structure is specified with a REPEATED statement as REPEATED/LOCAL=POM(EST), where EST is a data set with the estimates of the parameters of the mean of the model, in this case, estimates of the intercept and slope.

Program

```

data cov;
  input estimate;
  datalines;
2
.2
;
proc mixed data=LR;
  model y=x/outp=pml outpm=pm_ml solution;
  repeated / local=pom(est);
  parms / pdata=cov;
  ods output solutionf=est1;
run;
proc compare data=est compare=est1;
  var estimate;
run;
data est; set est1; run;
data cov; set cov1; run;
proc mixed data=LR;
  **iteration 2;
  model y = x / solution;
  repeated / local=pom(est);
  parms / pdata=cov;
  ods rtf select covparms solutionf tests3;
  ods output covparms=cov1 solutionf=est1;
run;
proc compare data=est compare=est1;
  var estimate;
run;
data est; set est1; run;
data cov; set cov1; run;
.
.
.
```

Results

The results of the initial step are given in Output 9.11.

Output 9.11 Results for Initial Fitting Power-of-the-Mean Covariance Structure for the Simple Linear Regression Model with PROC MIXED Using METHOD=REML

Covariance Parameter Estimates	
Cov Parm	Estimate
POM	2.6462
Residual	0.02771

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.6133	1.3396	16	2.70	0.0159
x	8.4429	1.4714	16	5.74	<.0001

Value Comparison Results for Variables

Obs	Base	Compare	Diff.	% Diff
	Estimate	Estimate		
1	4.49757	3.61332	-0.8843	-19.6608
2	7.78248	8.44287	0.6604	8.4856

Interpretation

The results in Output 9.11 contain the results for fitting the Power-of-the-Mean covariance structure provide estimates $\hat{\theta} = 2.6462$, $\hat{\sigma}^2 = 0.02771$, $\hat{\alpha} = 3.6133$, and $\hat{\beta} = 8.4429$. These estimates of the intercept and slope are different from those used for the starting values, so PROC COMPARE is used to compare the estimates of the intercept and slope from the initial values to the values in Output 9.10. The results of PROC COMPARE are in the “Value Comparison Results for Variables” table. The estimates changed 19.6% for the intercept and 8.5% for the slope. Now what needs to be done is to carry out several iterations of this process using the ending values for the covariance parameters and the slope and intercept as starting values. The process should continue until the differences between successive estimates of the intercept and slope change very little. The two DATA lines take the current values of the parameter estimates and makes them the initial values for the next iteration. This process was continued for six iterations and the final results are given in Output 9.12.

Results
Output 9.12 Results of Last Iteration from Fitting Power-of-the-Mean Covariance Structure for the Simple Linear Regression Model with PROC MIXED Using METHOD=REML

Covariance Parameter Estimates	
Cov Parm	Estimate
POM	2.4356
Residual	0.04989

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.5717	1.2456	16	2.87	0.0112
x	8.4674	1.4363	16	5.90	<.0001

Value Comparison Results for Variables

Obs	Base	Compare	Diff.	% Diff
	Estimate	Estimate		
1	3.57175	3.57174	-4.069E-6	-0.000114
2	8.46736	8.46736	2.4256E-6	0.0000286

Interpretation

The results in Output 9.12 indicate that estimates of the intercept and slope changed less than 0.0002%. The final estimates of the parameter based on REML estimates of the covariance parameters are $\hat{\theta} = 2.4356$, $\hat{\sigma}^2 = 0.04985$, $\hat{\alpha} = 3.5717$, and $\hat{\beta} = 8.4674$.

9.3.5 Fit Power-of-the-Mean Model Using ML from PROC MIXED

The results for fitting the Power-of-the-Mean model using METHOD=ML for estimating the covariance parameters are needed in order to make comparisons with those obtained with PROC NLMIXED. As in Section 9.3.5, initial values must be provided for the intercept and slope as well as starting values for the covariance parameters. This can be done using the external data set, and then the process needs to go through a few iterations to enable the fixed effects parameters to converge. The following program provides the code to obtain the initial values for the fixed effects parameters, for the covariance parameters, and to fit the Power-of-the-Mean covariance structure using METHOD=ML. This code would be run several times until the fixed effects parameters converge.

Program

```
proc mixed data=LR;
  model y=x/ solution;
  ods select covparms solutionf tests3;
  ods output solutionf=est;
run;
data cov; input estimate;
 datalines;
  .1
  3
  ;
proc mixed data=LR method=ml;
  **2nd interation;
  model y=x/outp=pml outpm=pm_ml solution;
  repeated / local=pom(est);
  parms/pdata=cov;
  ods output covparms=cov1 solutionf=est1;
  ods select covparms solutionf tests3;
run;
proc compare data=est compare=est1;
  var estimate;
run;
```

Results

The results of the last iteration are given in Output 9.13.

Output 9.13 Results of Last Iteration from Fitting Power-of-the-Mean Covariance Structure for the Simple Linear Regression Model with PROC MIXED Using METHOD=ML

Covariance Parameter Estimates	
Cov Parm	Estimate
POM	2.7821
Residual	0.01619

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.4762	0.9411	16	3.69	0.0020
x	8.6061	1.3232	16	6.50	<.0001

Value Comparison Results for Variables

Obs	Base Estimate	Compare		
		Estimate	Diff.	% Diff
1	3.47619	3.47618	-8.629E-6	-0.000248
2	8.60605	8.60606	7.0232E-6	0.0000816

Interpretation

The results in Output 9.13 indicate that estimates of the intercept and slope changed less than 0.0003%. The final estimates of the parameter based on ML estimates of the covariance parameters are $\hat{\theta} = 2.7821$, $\hat{\sigma}^2 = 0.01619$, $\hat{\alpha} = 3.4762$, and $\hat{\beta} = 8.6061$.

9.3.6 Fit Power-of-the-Mean Model Using ML from PROC NLMIXED

PROC NLMIXED can be used to fit the power-of-the-mean covariance structure, but PROC NLMIXED carries out the iteration process for you. The estimates of the variances are based on the method of ML and are used to compare the results from PROC MIXED with METHOD=ML. The following program fits the model using PROC NLMIXED.

Program

```
proc nlmixed data=LR;
parms a=4.5 b=7.5 T=1 sig2=5;
mean = a+b*x;
model y ~ normal(mean,sig2*abs(mean) **T);
predict mean out=mean df=16;
run;
```

Results

The results are given in Output 9.14.

Output 9.14 Results of Fitting Power-of-the-Mean Covariance Structure for the Simple Linear Regression Model with PROC NLMIXED

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a	3.3139	0.9436	18	3.51	0.0025	0.05	1.3315	5.2963	0.000027
b	8.8068	1.3572	18	6.49	<.0001	0.05	5.9556	11.6581	0.000042
T	2.7641	0.7937	18	3.48	0.0027	0.05	1.0967	4.4315	-0.0002
sig2	0.01672	0.03863	18	0.43	0.6703	0.05	-0.06444	0.09789	-0.00382

Interpretation

The parameter estimates in Output 9.14 are $\hat{\theta} = 2.7641$, $\hat{\sigma}^2 = 0.01672$, $\hat{\alpha} = 3.3139$, and $\hat{\beta} = 8.8068$, which are quite similar to those in Output 9.12. The differences in the parameter estimates are likely due to the fact that PROC NLMIXED is an iterative process that provides simultaneous estimates of all of the parameters, whereas PROC MIXED uses the previous estimates of the fixed effects in the current iteration.

9.3.7 Comparison of Power-of-X and Power-of-the-Mean Model Covariance Structures

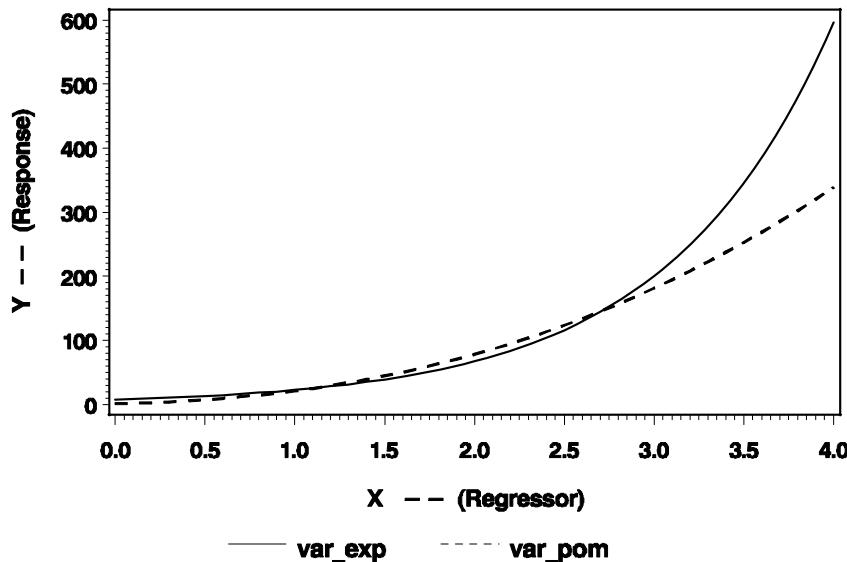
Two covariance structures have been fit to the linear regression data. It is of interest to compare these two structures to determine places where they are similar and where they are different. One way to accomplish this is to compute the two variances for a range of values of x . That was done using the REML estimates from PROC MIXED in the following program. The graph is in Figure 9.6.

Program

```
data varfct;
  do x=0 to 4 by .1;
    var_exp = 7.6071*exp(1.0907*x);
    var_pom = 0.04984*(abs(3.5717+8.4674*x))**2.4359;
    output;
  end;
run;
```

Results

Figure 9.6 Plot of the Two Variance Functions versus X



Interpretation

The graph in Figure 9.6 shows the two variance functions are very close in the range of $x \in [0, 3]$ and the Power-of-X model becomes larger than Power-of-the-Mean for $x > 3$. The range of the data set is from 0 to 3.3, so either of these variance functions is adequate to describe the variability.

9.4 Example: Nested Model with Unequal Variances for a Random Effect

An automobile manufacturer wanted to study the variability in the sales of its auto dealerships in the United States. The United States has been divided into six regions, with each region subdivided into districts with dealerships. Districts were selected at random from the set of districts from each region, and dealerships were selected at random within each of the selected districts. The total amount of new automobile sales (\$100,000) for the previous three months was used as the response variable. The regions are considered as the levels of a fixed effect, the districts nested with a region are considered as the levels of a random effect, and the dealerships are the experimental units. The estimation of the variance components is one of the steps needed to investigate sources of variation in a process such as selling automobiles.

9.4.1 Model with Equal Variances

A model that can be used to describe this data is

$$\begin{aligned} Y_{ijk} &= \mu + \rho_i + d_{j(i)} + e_{ijk} \\ i &= 1, 2, \dots, 6 \\ j &= 1, 2, \dots, n_{dist} \\ k &= 1, 2, \dots, m_{deal} \\ d_{j(i)} &\sim N(0, \sigma_{dist}^2) \\ e_{ijk} &\sim N(0, \sigma_{deal}^2) \end{aligned}$$

The assumptions for this model are that the district effects are identically independently distributed with variance σ_{dist}^2 and the dealerships within a district are identically independently distributed with variance σ_{deal}^2 . The following program can be used to fit the model above to the sales data set, Data Set 9.4, “Sales,” in Appendix 2, “Data Sets.”

Program

```
data sales_univ;
  set sales;
  array sale{8} s1-s8;
  do i=1 to 8;
    sales = sale{i};
    output;
  end;
  drop s1-s8;
run;

title2 'Fit Equal variance model';
proc mixed data=sales_univ ic;
  class region dist;
  model sales = region / ddfm=kr outp=r;
  random dist(region) /solution;
  lsmeans region /diff adjust=tukey;
  ods output solutionr=random;
run;
```

The DATA step converts the sales data from multivariate format (multiple sales numbers per region and district are stored in separate variables) to univariate form.

Results

The results are given in Output 9.15.

Output 9.15 Results for Fitting the Equal Variance-Independent Errors Model to the Sales Data

Covariance Parameter Estimates	
Cov Parm	Estimate
dist(region)	20.8784
Residual	8.3607

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
1018.6	2	1022.6	1022.6	1023.4	1025.3	1027.3

Solution for Random Effects							
Effect	region	dist	Estimate	Std Err Pred	DF	t Value	Pr > t
dist(region)	NC	1	1.2426	2.7732	24.5	0.45	0.6580
dist(region)	NC	2	0.2596	2.7650	24.2	0.09	0.9260
dist(region)	NC	3	-1.5022	2.7650	24.2	-0.54	0.5919
dist(region)	NE	1	-0.4269	1.8746	35.2	-0.23	0.8212
dist(region)	NE	2	-0.2718	1.9372	39	-0.14	0.8892
dist(region)	NE	3	2.8062	1.9372	39	1.45	0.1554
dist(region)	NE	4	-1.3911	1.8746	35.2	-0.74	0.4630
dist(region)	NE	5	2.9777	1.8746	35.2	1.59	0.1211
dist(region)	NE	6	4.6830	2.0528	46	2.28	0.0272
dist(region)	NE	7	-1.7363	1.8746	35.2	-0.93	0.3606
dist(region)	NE	8	-6.6408	1.8746	35.2	-3.54	0.0011
dist(region)	NW	1	2.4895	2.8062	25.5	0.89	0.3833
dist(region)	NW	2	-1.5704	2.8429	26.6	-0.55	0.5853
dist(region)	NW	3	-0.9191	2.7876	24.9	-0.33	0.7444
dist(region)	SC	1	1.3540	2.0886	30.9	0.65	0.5216
dist(region)	SC	2	2.7421	2.1772	35.3	1.26	0.2161
dist(region)	SC	3	2.0610	2.1103	32	0.98	0.3361
dist(region)	SC	4	-1.9797	2.1772	35.3	-0.91	0.3694
dist(region)	SC	5	-1.9482	2.1386	33.4	-0.91	0.3688
dist(region)	SC	6	-2.2292	2.0886	30.9	-1.07	0.2941
dist(region)	SE	1	5.2193	2.0874	30.8	2.50	0.0179

Solution for Random Effects							
Effect	region	dist	Estimate	Std Err Pred	DF	t Value	Pr > t
dist(region)	SE	2	7.8077	2.1760	35.2	3.59	0.0010
dist(region)	SE	3	4.5202	2.1091	31.9	2.14	0.0398
dist(region)	SE	4	-3.7098	2.1760	35.2	-1.70	0.0970
dist(region)	SE	5	-4.2360	2.1091	31.9	-2.01	0.0531
dist(region)	SE	6	-9.6013	2.0874	30.8	-4.60	<.0001
dist(region)	SW	1	-2.1532	2.8187	25.9	-0.76	0.4518
dist(region)	SW	2	7.5868	2.8187	25.9	2.69	0.0123
dist(region)	SW	3	-5.4336	2.7853	24.9	-1.95	0.0624

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
region	5	23	3.99	0.0095

Least Squares Means						
Effect	region	Estimate	Standard Error	DF	t Value	Pr > t
region	NC	42.1149	2.7064	22.5	15.56	<.0001
region	NE	41.7732	1.6635	22.8	25.11	<.0001
region	NW	34.1776	2.7314	23.3	12.51	<.0001
region	SC	32.6783	1.9241	23	16.98	<.0001
region	SE	35.1069	1.9227	22.9	18.26	<.0001
region	SW	40.4056	2.7288	23.3	14.81	<.0001

Differences of Least Squares Means									
Effect	region	_region	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
region	NC	NE	0.3417	3.1767	22.6	0.11	0.9153	Tukey-Kramer	1.0000
region	NC	NW	7.9373	3.8451	22.9	2.06	0.0505	Tukey-Kramer	0.3394
region	NC	SC	9.4366	3.3206	22.7	2.84	0.0093	Tukey-Kramer	0.0859
region	NC	SE	7.0079	3.3198	22.6	2.11	0.0460	Tukey-Kramer	0.3165
region	NC	SW	1.7093	3.8432	22.9	0.44	0.6607	Tukey-Kramer	0.9975
region	NE	NW	7.5956	3.1981	23.2	2.38	0.0262	Tukey-Kramer	0.2062
region	NE	SC	9.0950	2.5435	22.9	3.58	0.0016	Tukey-Kramer	0.0176
region	NE	SE	6.6663	2.5424	22.9	2.62	0.0153	Tukey-Kramer	0.1319
region	NE	SW	1.3676	3.1958	23.1	0.43	0.6727	Tukey-Kramer	0.9979
region	NW	SC	1.4994	3.3410	23.2	0.45	0.6578	Tukey-Kramer	0.9974
region	NW	SE	-0.9293	3.3402	23.2	-0.28	0.7833	Tukey-Kramer	0.9997

Differences of Least Squares Means									
Effect	region	_region	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
region	NW	SW	-6.2280	3.8609	23.3	-1.61	0.1202	Tukey-Kramer	0.5987
region	SC	SE	-2.4287	2.7201	23	-0.89	0.3812	Tukey-Kramer	0.9444
region	SC	SW	-7.7274	3.3389	23.2	-2.31	0.0299	Tukey-Kramer	0.2287
region	SE	SW	-5.2987	3.3381	23.1	-1.59	0.1260	Tukey-Kramer	0.6145

Interpretation

The estimates of the parameters of the model are in Output 9.15, where the estimates of the variance components are in the “Covariance Parameter Estimates” table. The estimates of the district and dealership variance components are 20.88 and 8.36, respectively. The estimates from the “Solution for Random Effects” are the EBLUPs of the district effects within each of the regions. The “Type 3 Tests for Fixed Effects” table provides a test of the equal sales means for the regions. There is evidence that the means are not equal ($p = 0.0095$). The “Least Squares Means” table provides the estimates of the region sales means, and the “Differences of Least Squares Means” table provides pairwise comparisons among the region sales means, with ADJUST=TUKEY being used to account for carrying out multiple tests. The adjusted p -values indicate that the sales mean for NE is significantly larger than the sales mean for SC. It is interesting to note that the mean for NC is larger than the mean for NE, but the mean for NC is not significantly different from the mean for SC.

9.4.2 Testing for Equal District Variances across Regions

There was a concern that the variances associated with the district effects and the dealership effects could be heterogeneous. First, the variances of the districts within a region were evaluated where EBLUPs of the district effects are requested (by /SOLUTION in the RANDOM statement) and a data set was created (ODS OUTPUT SOLUTIONR=RANDOM). The estimates of the model parameters and the EBLUPs of the random effects are given in Output 9.15. The following program provides box plots of the district EBLUPs for each of the regions using ODS statistical graphics in PROC MIXED.

Program

```

ods html;
ods graphics on;

proc mixed data=random boxplot;
  class region;
  model estimate = region;
run;

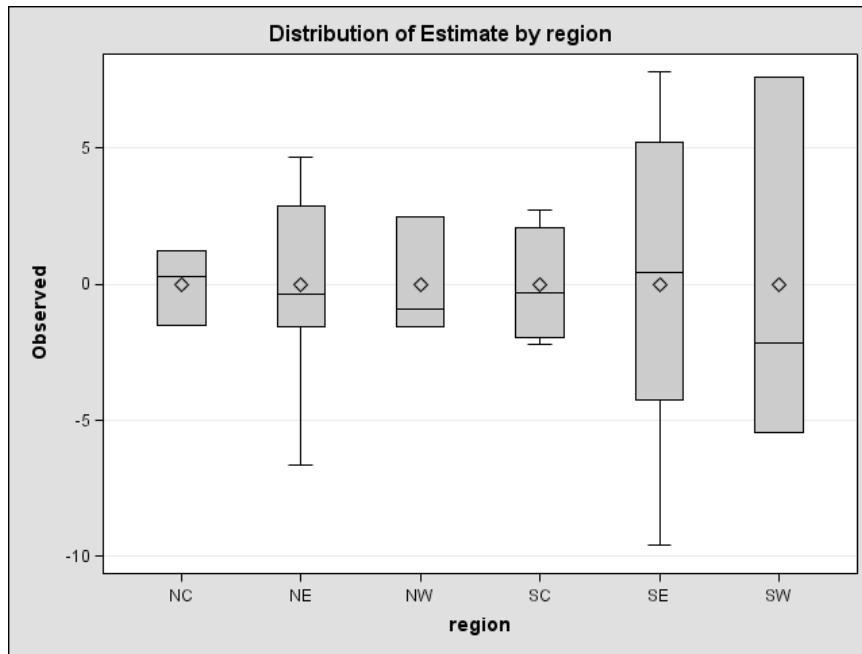
ods graphics off;
ods html close;

```

The BOXPLOT option requests box plots of the response and the residuals for each classification main effect in the model, in this case box plots by region.

Results

Since the response variable in this application corresponds to the EBLUPs, it is sufficient to consider the box plots for the response, rather than the residuals of the fit of the EBLUPs against the region effect (Figure 9.7).

Figure 9.7 Box Plot of District EBLUPs

Interpretation

The box plots in Figure 9.7 indicate that there might be differences among the variances of the district effects across regions, so a model that allows for unequal district variances across the regions is a possibility. The following program can be used to compute a Levene's type test for equality of the district variances across regions.

Program

```
data random1; set random;
  absdist=abs(estimate);
run;
proc mixed data=random1;
  class region;
  model absdist=region;
  lsmeans region/diff;
run;
```

Results

The results are given in Output 9.16.

Output 9.16 Results for the Levene's Test for Equality of District Variances

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	3.4151

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
region	5	23	4.96	0.0032

Interpretation

The DATA step takes the EBLUPs of the districts within a region (in data set random) and computes the absolute values. Within a region the EBLUPs sum to zero, so they are like residuals. The PROC MIXED code fits a one-way analysis of variance model with region as the fixed effect. The “Type 3 Tests of Fixed Effects” table provides the Levene’s type test with an F value of 4.69 and a p -value of 0.0032, indicating there is strong evidence that the district variances are not equal. The LSMEANS statement can be used to construct groups of regions where the district variances are similar within a group and the district variances are unequal between groups. That process was not done here.

9.4.3 Test for Equal Residual Variances

While you are looking at the possibility of heterogeneous variances you might as well investigate the possibility of unequal variances for the dealerships within a district. The following program provides the process for computing Levene’s test for equality of dealership variances within districts across regions.

Program

```
data r1; set r;
absr=abs(resid);
run;
title2 'Test equality of residual means for regions and district
within regions';
proc mixed data=r1;
class region dist;
model absr=region dist(region);
run;
```

Results

The results of Levene’s test are in Output 9.17.

Output 9.17 Results for Using Levene’s Test to Test the Equality of Residual Variances

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	2.3938

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
region	5	165	0.25	0.9393
dist(region)	23	165	1.60	0.0495

Interpretation

The residuals were computed with the code in Section 9.4.1 by using the OUTP=R option in the MODEL statement. One can test the equality of the residual variances across regions and across districts within a region. To carry out these tests, both region and district(region) must be included in the MODEL statement (fixed effects). The region source in the “Type 3 Tests for Fixed Effects” table provides a test of the hypothesis that the residual variances are equal across regions, and the dist(region) source provides a test of the equality of the residual variances across districts within a region. The *p*-value for dist(region) is 0.0495, indicating there is possibly some concern, but because of the small numbers of dealers within a district, that possibility was not considered in the following analyses.

9.4.4 Fitting Model with Unequal Random Effect Variances

A model with different district variances for each of the regions and equal dealership variances within the districts of a region can be fit using the following program. The RANDOM statement using the GROUP=REGION option specifies the unequal district variances within a region.

Program

```
proc mixed data=sales_univ ic;
  title2 'Fit model to Sales data with unequal district
          within region variances';
  class region dist;
  model sales=region/ddfm=kr;
  random dist/group=region;
  lsmeans region/diff adjust=tukey;
run;
```

Results

The results with unequal district variances are given in Output 9.18.

Output 9.18 Results for Fitting the Equal Residual Variances and Unequal District Variances within a Region

Covariance Parameter Estimates		
Cov Parm	Group	Estimate
dist	region NC	1.0309
dist	region NE	12.8191
dist	region NW	3.9243
dist	region SC	4.5475
dist	region SE	51.0994
dist	region SW	51.0666
Residual		8.3639

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
1007.3	7	1021.3	1021.9	1017.6	1021.9	1028.9

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
region	5	5.1	9.62	0.0126

Least Squares Means						
Effect	region	Estimate	Standard Error	DF	t Value	Pr > t
region	NC	42.0879	0.8423	1.98	49.97	0.0004
region	NE	41.7541	1.3266	6.92	31.47	<.0001
region	NW	34.2099	1.3498	1.95	25.34	0.0018
region	SC	32.6760	0.9902	4.91	33.00	<.0001
region	SE	35.1189	2.9553	5.01	11.88	<.0001
region	SW	40.4370	4.1846	2	9.66	0.0106

Differences of Least Squares Means									
Effect	region	_region	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
region	NC	NE	0.3338	1.5714	8.69	0.21	0.8367	Tukey-Kramer	0.9999
region	NC	NW	7.8779	1.5910	3.29	4.95	0.0127	Tukey-Kramer	0.0267
region	NC	SC	9.4118	1.3000	6.34	7.24	0.0003	Tukey-Kramer	0.0050
region	NC	SE	6.9689	3.0730	5.75	2.27	0.0657	Tukey-Kramer	0.3447
region	NC	SW	1.6509	4.2685	2.16	0.39	0.7337	Tukey-Kramer	0.9981
region	NE	NW	7.5441	1.8926	6.01	3.99	0.0072	Tukey-Kramer	0.0627
region	NE	SC	9.0781	1.6554	11.7	5.48	0.0002	Tukey-Kramer	0.0173
region	NE	SE	6.6351	3.2394	7.02	2.05	0.0796	Tukey-Kramer	0.4261
region	NE	SW	1.3171	4.3898	2.41	0.30	0.7881	Tukey-Kramer	0.9994
region	NW	SC	1.5339	1.6741	4.16	0.92	0.4095	Tukey-Kramer	0.9266
region	NW	SE	-0.9090	3.2490	6.57	-0.28	0.7883	Tukey-Kramer	0.9996
region	NW	SW	-6.2270	4.3969	2.41	-1.42	0.2724	Tukey-Kramer	0.7209
region	SC	SE	-2.4429	3.1168	6.11	-0.78	0.4624	Tukey-Kramer	0.9592
region	SC	SW	-7.7609	4.3001	2.23	-1.80	0.1999	Tukey-Kramer	0.5318
region	SE	SW	-5.3180	5.1229	4.08	-1.04	0.3567	Tukey-Kramer	0.8870

Interpretation

The results in the “Covariance Parameter Estimates” table of Output 9.18 show that the district variances within a region are quite different, ranging from 1.0 to 51.1. The AIC for the equal variance model from Output 9.15 is 1022.6, while the AIC for the unequal district within region variance model is 1021.3. Thus this unequal variance model fits the data a little better than the equal variance model according to the AIC. The least-squares means are given as well as the pairwise differences. The estimated standard errors of the least-squares means and of pairwise differences are a reflection of the sample size and the district variance for each region. The adjusted *p*-values indicate that the mean of sales for NC is larger than the means for NW and SC, while the mean of NE is larger than the mean for SC (*p* < 0.05).

Using the unequal variance model when appropriate provides better estimates of the standard errors of differences of fixed effects, and thus the inference should be much more appropriate than when an equal variance model is used in the analyses.

9.5 Example: Within-Subject Variability

This example focuses on within-subject heterogeneity. Vonesh and Carter (1992) describe and analyze data on high-flux hemodialyzers measured to assess their *in vivo* ultrafiltration characteristics. The response is the ultrafiltration rate (UFR in ml/hr) of 20 high-flux membrane dialyzers measured at 7 different transmembrane pressures (TMP in dmHg). The dialyzers are evaluated *in vitro* using bovine blood and blood flow rates (QB) of either 200 or 300 dl/min.

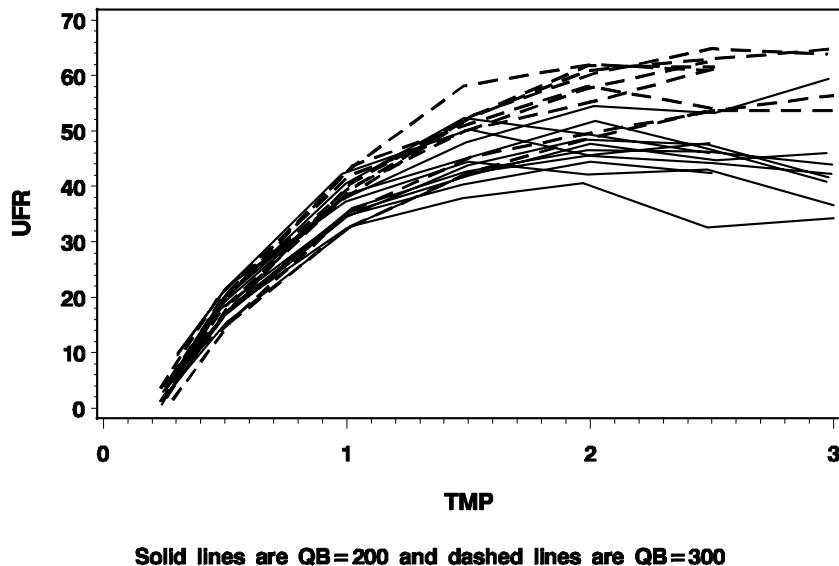
The program to place these data into a SAS data set and plot them as in Figure 9.8 follows. Note that the data are divided by 100 here to add stability to the estimation algorithm in PROC MIXED. See Data Set 9.5, “Dial,” in Appendix 2, “Data Sets,” for the complete data set.

Program

```

data dial;
  input sub qb tmp ufr index type;
  datalines;
  ... more datalines...
run;
axis2 value=( font=swiss height=.2 in)
  width=3
  value=(f=swissb h=.15 in )
  order=0 TO 70 BY 10
  label=(f=swissb H=.16 in a=90 'UFR');
axis1 value=( font=swiss height=.2 in)
  width=3
  label=(f=swissb H=.16 in 'TMP')
  value=(f=swissb h=.15 in )
  order=0 TO 3 By 1 ;
symbol1 v=none i=join r=11 l=1 c=black;
symbol2 v=none i=join r=11 l=3 c=black;
footnote f=swissb h=.15 in
  "Solid lines are QB=200 and dashed lines are QB=300";
proc gplot data=dial;
  plot ufr*tmp=sub/nolegend vaxis=axis2 haxis=axis1;
run;

```

Figure 9.8 High-Flux Hemodialyzer Data

The curves in Figure 9.8 exhibit a definite increase in variability of UFR as TMP increases. It is this heterogeneity that we want to capture with covariance structures.

9.5.1 Basic Unstructured Covariance Model

Following the recommendations of Diggle (1988) and Wolfinger (1993a), a reasonable initial model for these data should involve fairly general specifications for both the mean and the variance-covariance structure. This generality helps avoid misspecification biases that can occur with models that are too simple. Exactly what makes up a general model can be quite problem specific, especially with regard to the mean model. Preliminary plots such as Figure 9.8 are valuable in deciding on a starting model.

Model

For the fixed effects model in this analysis, simple linear or quadratic curves do not appear to be reasonable. Therefore, the mean model is assumed to contain additional cubic and quartic terms, and the parameters for these curves are assumed to be different for each of the two blood flow rates. This is in contrast to the nonlinear mean model of Vonesh and Carter (1992), although the quartic curves track these data fairly well. The fixed effects component of the model is thus

$$Y = \beta_0 + \tau_i + (\beta_1 + \delta_{1i})X + (\beta_2 + \delta_{2i})X^2 + (\beta_3 + \delta_{3i})X^3 + (\beta_4 + \delta_{4i})X^4 + e$$

where

β_0 is the intercept over both QB levels

τ_i is the i^{th} QB level on the intercept

$\beta_1, \beta_2, \beta_3$, and β_4 are the linear, quadratic, cubic, and quartic regression coefficients averaged over QB levels

δ_{1i} , δ_{2i} , δ_{3i} , and δ_{4i} are the effects of the i^{th} QB level on the linear, quadratic, cubic, and quartic regression coefficients

X is the value of TMP for a given observation

For the variance-covariance model, the basic repeated measures model from Chapter 5 is used here. In particular, it is assumed that the data from different dialyzers are independent, and that the data within a dialyzer are correlated in some fashion. More importantly, it appears sensible to allow the variances of the data to increase with higher transmembrane pressures.

The most general covariance structure possible is a 7×7 unstructured matrix, and because this is reasonable to compute, it is taken as the initial model. An advantage of considering this most general model is that the estimate of the covariance can be inspected for heterogeneous patterns in both the variances and correlations.

Program

The appropriate PROC MIXED program is as follows:

```
proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp|tmp;
  repeated / type=un subject=sub r rcorr;
run;
```

The MODEL statement sets up a common quartic curve using the bar (|) operator, and then different curves for the two levels of QB. Although the second set of effects contains the first set, the first set is included in order to carry out tests for whether the different curves are necessary.

The REPEATED statement requests a block-diagonal **R** matrix with blocks defined by the SUBJECT= option and the structure of each block defined by the TYPE= option. For these data, **R** is 140×140 with twenty 7×7 blocks, each one corresponding to a dialyzer. Each block has the same unstructured form with 28 unknown parameters (7 variances and 21 covariances). The R and RCORR options print the estimate of the first block and its corresponding correlation matrix.

Results

The results for the unstructured covariance structure are given in Output 9.19.

Output 9.19 Basic Unstructured Covariance Model

Estimated R Matrix for sub 1							
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	2.7562	2.9042	3.5731	3.0405	0.3594	0.4551	0.6414
2	2.9042	5.1024	6.3987	6.3753	4.1338	3.3188	1.1641
3	3.5731	6.3987	11.1529	12.4563	8.3256	5.4425	4.0180
4	3.0405	6.3753	12.4563	18.5429	13.3760	10.8999	7.6782
5	0.3594	4.1338	8.3256	13.3760	17.7128	13.8347	12.0450
6	0.4551	3.3188	5.4425	10.8999	13.8347	20.3069	11.3283
7	0.6414	1.1641	4.0180	7.6782	12.0450	11.3283	19.6740

Estimated R Correlation Matrix for sub 1							
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.7744	0.6445	0.4253	0.05143	0.06083	0.08710
2	0.7744	1.0000	0.8482	0.6554	0.4348	0.3260	0.1162
3	0.6445	0.8482	1.0000	0.8662	0.5924	0.3616	0.2713
4	0.4253	0.6554	0.8662	1.0000	0.7381	0.5617	0.4020
5	0.05143	0.4348	0.5924	0.7381	1.0000	0.7295	0.6452
6	0.06083	0.3260	0.3616	0.5617	0.7295	1.0000	0.5668
7	0.08710	0.1162	0.2713	0.4020	0.6452	0.5668	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	sub	2.7562
UN(2,1)	sub	2.9042
UN(2,2)	sub	5.1024
UN(3,1)	sub	3.5731
UN(3,2)	sub	6.3987
UN(3,3)	sub	11.1529
UN(4,1)	sub	3.0405
UN(4,2)	sub	6.3753
UN(4,3)	sub	12.4563
UN(4,4)	sub	18.5429
UN(5,1)	sub	0.3594
UN(5,2)	sub	4.1338
UN(5,3)	sub	8.3256
UN(5,4)	sub	13.3760
UN(5,5)	sub	17.7128
UN(6,1)	sub	0.4551
UN(6,2)	sub	3.3188
UN(6,3)	sub	5.4425
UN(6,4)	sub	10.8999
UN(6,5)	sub	13.8347
UN(6,6)	sub	20.3069
UN(7,1)	sub	0.6414
UN(7,2)	sub	1.1641
UN(7,3)	sub	4.0180
UN(7,4)	sub	7.6782
UN(7,5)	sub	12.0450
UN(7,6)	sub	11.3283
UN(7,7)	sub	19.6740

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
27	153.08	<.0001

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
586.1	28	642.1	658.2	649.3	672.6	700.6

Type 3 Tests of Fixed Effects					
Effect	Num DF	Den DF	F Value	Pr > F	
tmp	1	18	577.35	<.0001	
tmp*tmp	1	18	72.60	<.0001	
tmp*tmp*tmp	1	18	13.94	0.0015	
tmp*tmp*tmp*tmp	1	18	4.09	0.0582	
qb	1	18	1.19	0.2902	
tmp*qb	1	18	0.04	0.8378	
tmp*tmp*qb	1	18	0.04	0.8355	
tmp*tmp*tmp*qb	1	18	0.01	0.9363	
tmp*tmp*tmp*tmp*qb	1	18	0.05	0.8193	

Interpretation

The REML algorithm converges in two iterations. Even though these data are perfectly balanced, these iterations are necessary because of the time-varying covariate TMP.

The “Estimated R Matrix for sub 1” table prints the first block of **R**. Its principal diagonal with elements (2.8, 5.1, 11.2, 18.5, 17.7, 20.3, 19.7) reveals that the variances increase steadily with TMP and then level off for the last four measurements.

Even more interesting is the correlation pattern, which appears to decrease with increasing distance between TMP values. For example, the correlations with the first measurement are estimated to be (0.77, 0.64, 0.43, 0.05, 0.06, 0.09). Thus, even though TMP does not measure time, the data appear to have an autoregressive correlation structure as if TMP did measure time. Some more parsimonious covariance structures exhibiting this kind of pattern may therefore fit these data well.

Note that the REML analysis has already numerically confirmed and quantified the heterogeneity of variance evident in Figure 9.8. It has also revealed a definite correlation pattern that is not so obvious in a plot of the dialyzer profiles.

The results in the “Covariance Parameter Estimates” table repeat the values from the “Estimated R Matrix for sub 1” along with their asymptotic standard errors obtained from the inverse of the second derivative matrix from the REML algorithm.

The “Null Model Likelihood Ratio Test” table indicates that this basic unstructured model fits exceedingly better than the null model with independent homogeneous errors. Note that the degrees of freedom for this test are equal to the difference in the number of parameters between the two models, which in this case is $28 - 1 = 27$.

The “Information Criteria” table lists the values of five information criteria as well as the value of the likelihood function.

The “Tests of Fixed Effects” table provides good evidence for dropping some of the higher-order model terms involving QB. Type I (sequential) tests may also be useful in this regard and could be printed using the HTYPE=1 option in the MODEL statement. However, following Diggle (1988) and Wolfinger (1993a), you should not do that until you have selected a covariance structure.

9.5.2 Other Covariance Structures

The strategy here for selecting a covariance structure is to fit a few possible candidate structures and then compare them using various information criteria. In addition to the previously fitted unstructured (UN) model, the structures investigated in this fashion are denoted AR(1), ARH(1), CS, CSH, HF, FA(1), FA1(1), RC, RCQ, and I-I. Brief descriptions of each of these now follow, and further details can be found in the PROC MIXED documentation and in Wolfinger (1996).

- The AR(1) structure is the first-order autoregressive structure. It has homogeneous variances and correlations that decline exponentially with distance. Although this structure is used almost exclusively with equally spaced time series data, it can still be used for these data by making TMP a proxy for time and exploiting the fact that the data are approximately equally spaced in TMP. The AR(1) covariance structure has two unknown parameters: the variance and the lag-one correlation.
- The ARH(1) is a direct generalization of AR(1). It has the same correlation structure, but it has heterogeneous variances instead of homogeneous ones. For these data with seven repeated measurements, it has eight unknown parameters (seven variances and one correlation parameter).
- CS is the well-known compound symmetry structure. Like AR(1), it has two unknown parameters, one modeling a homogeneous variance and the other modeling a correlation. But unlike AR(1), the correlation is assumed to remain constant.
- CSH is to CS what ARH(1) is to AR(1). That is, CSH has constant correlations but heterogeneous variances. Like ARH(1), it has eight unknown parameters for these data.
- HF is the “spherical contrast” structure discussed by Huynh and Feldt (1970). It is similar to CSH in that it has a different variance parameter at each repeated measurement, but the covariances are constructed by taking arithmetic rather than geometric means. Therefore, the correlations are not constant although there is still only one parameter modeling them. HF thus has eight unknown parameters in all.
- FA(1) and FA1(1) are first-order factor analytic structures. They are constructed by taking the outer product of a vector of unknown factor loadings and adding a diagonal matrix of specific variances to it. The specific variances are all different in the standard FA(1) structure, producing a total of fourteen parameters. In the FA1(1) structure, they are the same, producing eight parameters.

The preceding seven structures are fit with PROC MIXED by making a simple change to the program used for the basic UN model in the previous subsection. The MODEL statement remains the same and the TYPE= option of the REPEATED statement is changed to indicate the different covariance structure. For example, the ARH(1) model is fit with the following program:

```
proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp|tmp;
  repeated / type=arh(1) subject=sub;
run;
```

The RC and RCQ models are random coefficient models as discussed in Chapter 7, where RC denotes linear random coefficients and RQ denotes quadratic random coefficients. The two covariance structures are fit with the following programs:

```
proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp|tmp;
  random int tmp / subject=sub type=un;
run;

proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp|tmp;
  random int tmp tmp*tmp / subject=sub type=un;
run;
```

In the first model (RC), the INT and TMP terms in the RANDOM statement model a random intercept and slope, respectively. The SUBJECT=SUB option is required to inform PROC MIXED of when new realizations of the random intercept and slopes are assumed to occur. The TYPE=UN option models an unstructured 2×2 covariance matrix for the random intercept and slope. This results in a **G** matrix that is 44×44 and block diagonal with twenty-two 2×2 blocks. **R** is assumed to equal a constant variance times the identity matrix; thus, this structure has four parameters in all.

The second model (RCQ) adds a third random coefficient, which is a quadratic term in TMP. The resulting **G** matrix from this model is 66×66 and block diagonal with twenty-two 3×3 blocks. **R** is again assumed to equal a constant variance times the identity matrix; thus, this structure has seven parameters in all. Among all of the linear and nonlinear covariance structures considered by Vonesh and Carter (1992), RCQ is the best fitting for a set of nonlinear least-squares residuals, according to an adjusted R^2 criterion.

The final model is the independent increments (I-I) model (Louis 1988), which assumes the observations form a random walk in time. Again, TMP is not a time variable, but the structure is included to see how well it fits. There is no I-I structure directly available in PROC MIXED, but because all of its parameters enter the structure linearly, you can use the TYPE=LIN and LDATA= options as follows:

```
data ii;
  input parm row col1-col7;
  datalines;
  1 1 1 1 1 1 1 1 1
  1 2 1 1 1 1 1 1 1
  1 3 1 1 1 1 1 1 1
```

```

1 4 1 1 1 1 1 1 1
1 5 1 1 1 1 1 1 1
1 6 1 1 1 1 1 1 1
1 7 1 1 1 1 1 1 1
2 2 0 1 1 1 1 1 1
2 3 0 1 1 1 1 1 1
2 4 0 1 1 1 1 1 1
2 5 0 1 1 1 1 1 1
2 6 0 1 1 1 1 1 1
2 7 0 1 1 1 1 1 1
3 3 0 0 1 1 1 1 1
3 4 0 0 1 1 1 1 1
3 5 0 0 1 1 1 1 1
3 6 0 0 1 1 1 1 1
3 7 0 0 1 1 1 1 1
4 4 0 0 0 1 1 1 1
4 5 0 0 0 1 1 1 1
4 6 0 0 0 1 1 1 1
4 7 0 0 0 1 1 1 1
5 5 0 0 0 0 1 1 1
5 6 0 0 0 0 1 1 1
5 7 0 0 0 0 1 1 1
6 6 0 0 0 0 0 1 1
6 7 0 0 0 0 0 1 1
7 7 0 0 0 0 0 0 1
run;

proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp;
    repeated / type=lin(7) ldata=ii sub=sub r rcorr;
run;

```

The TYPE=LIN(7) option specifies a general linear structure that is of the form

$$\sum_{i=1}^7 \theta_i A_i$$

where the A_i are known matrices and the θ are unknown parameters. The LDATA= option specifies an auxiliary SAS data set containing the A_i matrices. This data set contains the PARM, ROW, and COL1-COL7 variables, which contain the indices, rows, and columns of all of the A_i .

The I-I structure is of the general linear form, and is constructed by adding together the A_i , which have their lower $(8-i) \times (8-i)$ elements equal to 1 and the remaining elements equal to zero, with $i = 1, \dots, 7$. These A_i matrices are specified in the preceding data set.

9.5.3 Selecting a Covariance Structure

All of the preceding models are fit with separate runs of PROC MIXED. After each run, the *AIC*, *AICC*, *HQIC*, *BIC*, and *CAIC* values based on the REML likelihood, along with the number of parameters, were saved from each PROC MIXED fit using the ODS feature. The results are given in Output 9.20.

Output 9.20 Information Criteria for Each of the Covariance Structures

Obs	Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC	type	model
1	586.1	28	642.1	658.2	649.3	672.6	700.6	UN	1
2	657.5	2	661.5	661.6	662.0	663.7	665.7	AR(1)	2
3	613.7	9	631.7	633.2	634.0	641.5	650.5	ARH(1)	3
4	694.3	2	698.3	698.4	698.8	700.5	702.5	CS	4
5	662.2	8	678.2	679.4	680.2	686.9	694.9	CSH	5
6	667.5	8	683.5	684.7	685.5	692.2	700.2	HF	6
7	639.6	14	667.6	671.2	671.2	682.8	696.8	FA(1)	7
8	671.6	8	687.6	688.8	689.7	696.4	704.4	FA1(1)	8
9	670.1	4	678.1	678.5	679.2	682.5	686.5	RC	9
10	645.8	7	659.8	660.8	661.6	667.5	674.5	RCQ	10
11	620.4	7	634.4	635.4	636.2	642.1	649.1	I-I	11

The simplest covariance structure is specified by TYPE=CS, with $-2 \text{ Res Log Likelihood} = 694.3$. All of the other structures considered here fit dramatically better than the TYPE=CS structure, illustrating the strong need to account for some kind of heterogeneity and/or correlation in these data.

Program

A convenient way of investigating the information criteria for the models under consideration is to construct plots versus the number of parameters or each information criteria versus the other. The following program provides the plots and produces Figures 9.9 through 9.17:

```

symbol11 color=black font=swiss value='UN'      ' repeat=1;
symbol12 color=black font=swiss value='AR(1)'   ' repeat=1;
symbol13 color=black font=swiss value='ARH(1)'  ' repeat=1;
symbol14 color=black font=swiss value='CS'       ' repeat=1;
symbol15 color=black font=swiss value='CSH'      ' repeat=1;
symbol16 color=black font=swiss value='HF'       ' repeat=1;
symbol17 color=black font=swiss value='FA(1)'   ' repeat=1;
symbol18 color=black font=swiss value='FA1(1)'  ' repeat=1;
symbol19 color=black font=swiss value='RC'       ' repeat=1;
symbol110 color=black font=swiss value='RCQ'     ' repeat=1;
symbol111 color=black font=swiss value='I-I'     ' repeat=1;

proc gplot data=models;
  plot aic*parms = model / nolegend;
  plot bic*parms = model / nolegend;
  plot aic*bic   = model / nolegend;
run;

```

The SYMBOL statements define features of the plots, and there is one for each model. The key specification is for VALUE=, which defines the plotting symbol to be the same as the model type. These symbol definitions are automatically incorporated into the PLOT statements from PROC GPLOT. Refer to *SAS/GRAPH Software: Reference* for further details about SYMBOL and PROC GPLOT.

Results

The plots of the information criteria are shown in Figures 9.9 to 9.17.

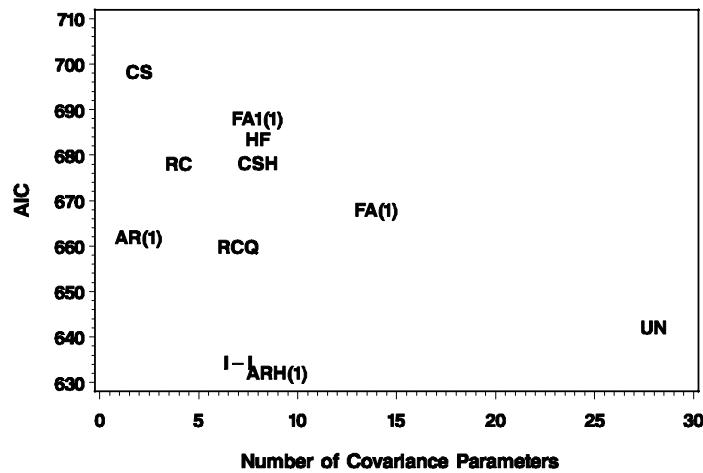
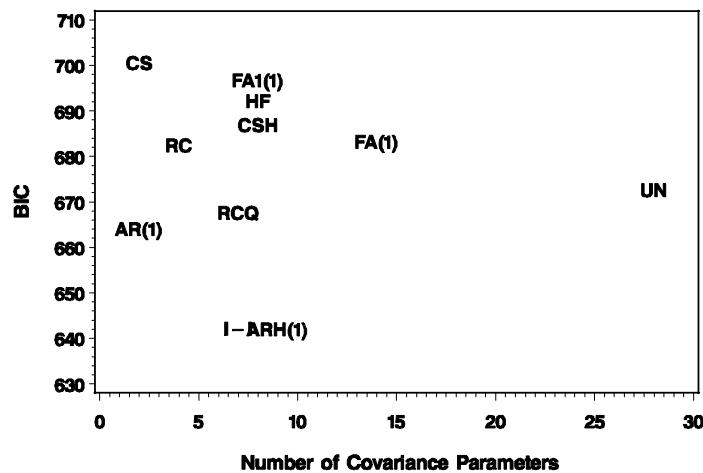
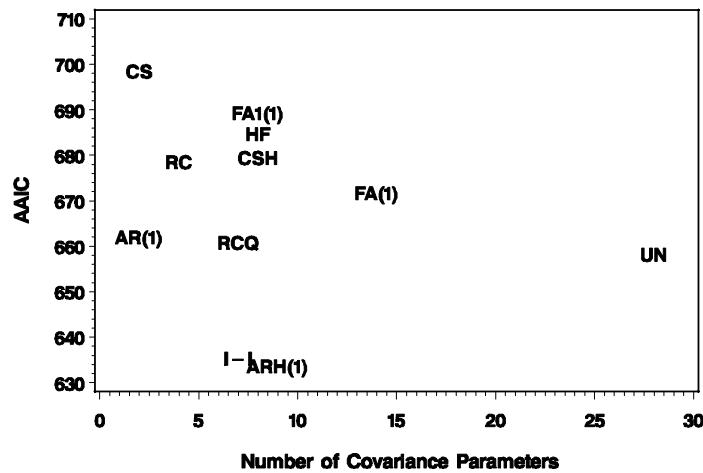
Figure 9.9 A/C for High-Flux Hemodialyzer Models**Figure 9.10** B/C for High-Flux Hemodialyzer Model**Figure 9.11** A/CC for High-Flux Hemodialyzer Models

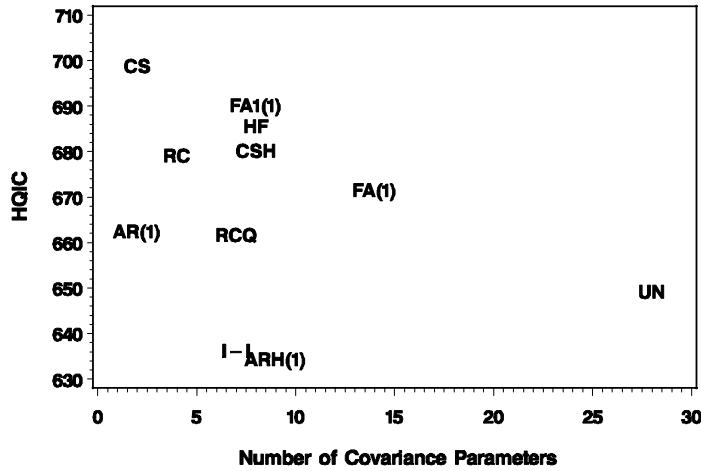
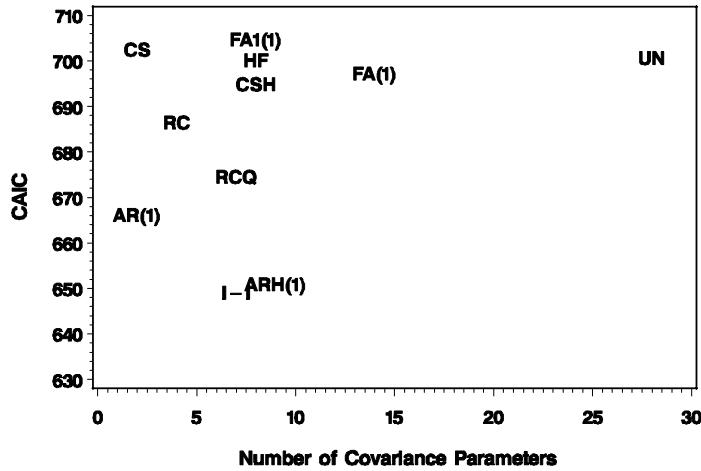
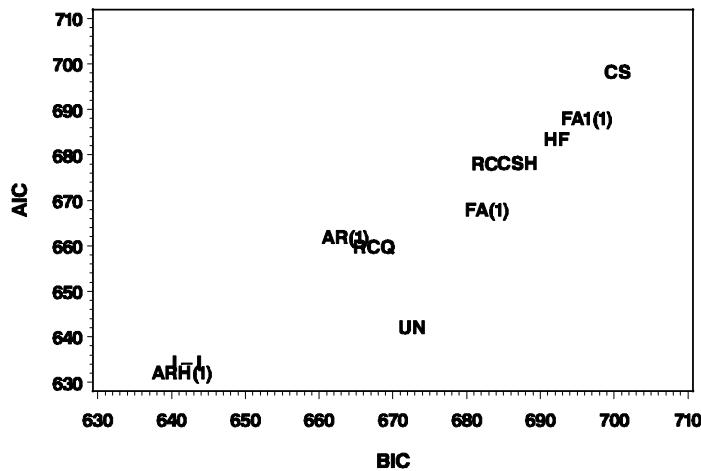
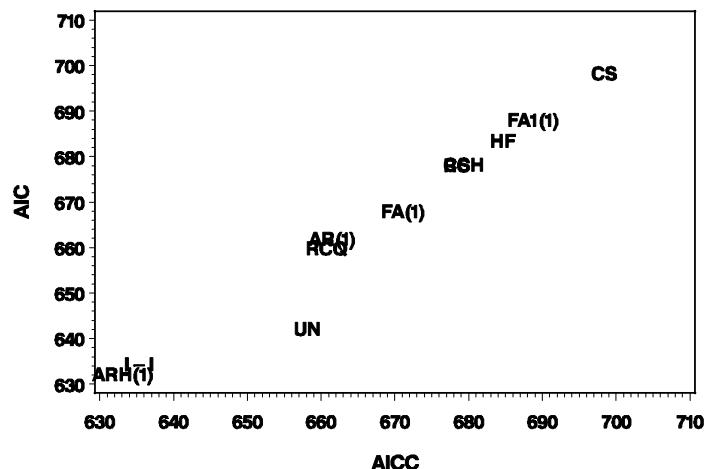
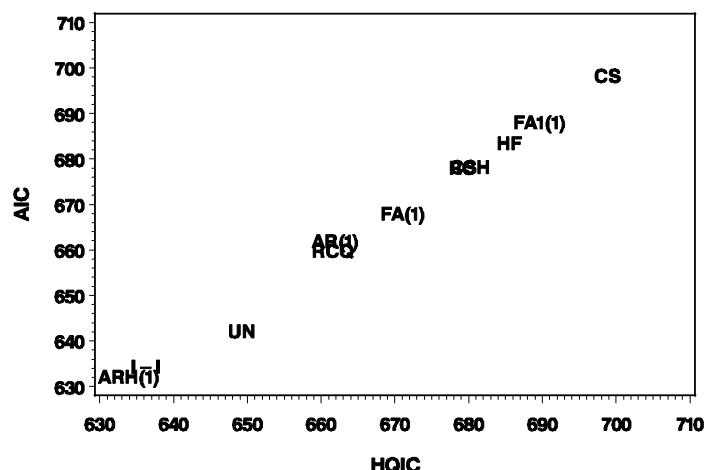
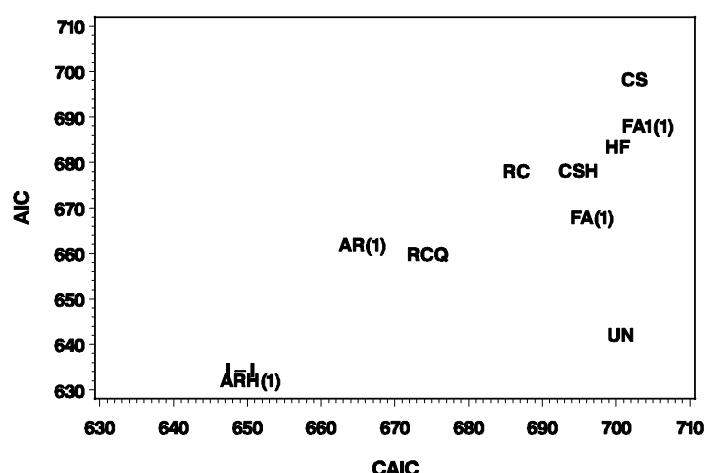
Figure 9.12 HQIC for High-Flux Hemodialyzer Models**Figure 9.13** CAIC for High-Flux Hemodialyzer Models**Figure 9.14** AIC versus BIC for High-Flux Hemodialyzer Models

Figure 9.15 A/C versus A/CC for High-Flux Hemodialyzer Models**Figure 9.16** A/C versus HQ/C for High-Flux Hemodialyzer Models**Figure 9.17** A/C versus CA/C for High-Flux Hemodialyzer Models

Interpretation

Figure 9.9 plots AIC versus model complexity. As suspected from inspection of the unstructured estimate, the heterogeneous AR(1) fits very well and is the best fitting according to AIC. The I-I and UN models also fit well, and CS fits the worst. The other information criteria are displayed in Figures 9.10–9.13. The ARH(1) and I-I covariance structures are consistently better than the others based on any of the criteria.

You can further compare ARH(1) and UN by carrying out a restricted likelihood ratio test. This test can compare two covariance models, one of which is nested within the other. It is constructed by subtracting values of -2 times the maximized restricted likelihoods and comparing this statistic with a chi-square distribution with degrees of freedom equal to the difference in the number of covariance parameters in the two models. The restricted likelihood version of the test is valid provided the fixed effects component of the model remains constant. If nested mean models are also considered, you should switch to a regular likelihood ratio test by using the METHOD=ML option in the PROC MIXED statement to fit both models.

Here the mean models are the same and ARH(1) is nested within UN. The restricted likelihood ratio test has a χ^2 value of $613.7 - 586.1 = 27.6$ with $28 - 8 = 20$ degrees of freedom and p -value > 0.1192 . Thus, one cannot reject the null hypothesis that the UN structure is no better than the ARH(1) structure. Because the variances are mostly monotone increasing, similar results can be obtained for the independent increments structure.

Figures 9.14 through 9.17 are plots of the AIC values versus the other information criteria, respectively. These graphs show ARH(1) and I-I to be the best-fitting models under all of the information criteria.

A final step in this particular analysis is to reduce the mean model under an ARH(1) assumption for the covariance structure. For this example, this entails dropping higher-order terms from the quartic part of the model. You can then carry out various inferences with the final model by using the CONTRAST, ESTIMATE, and LSMEANS statements.

9.5.4 Power-of-the-Mean Model

Another technique for handling within-subject heterogeneity like the one present in the hemodialyzer data is to model it as a power of the mean.

Model

For this model the variance matrix \mathbf{R} is assumed to be of the form

$$\mathbf{R} = \text{diag}\left(\sigma^2 | \mathbf{x}'_i \boldsymbol{\beta}^* |^\theta\right)$$

where

\mathbf{x}'_i is the i^{th} row of the fixed effects design matrix \mathbf{X}

$\boldsymbol{\beta}^*$ is a specified vector of fixed effects parameters

θ is the power to be estimated

Note that by itself this structure sets all covariances equal to zero. For the hemodialyzer data, there is strong evidence of within-subject correlation, and so something should be added to the model to account for it. Although it is possible to add a Power-of-the-Mean model to one of the

REPEATED structures considered in the previous section, the approach for this section is to incorporate it into a random coefficients model.

Because the quadratic random coefficients model (RCQ) fits better than the linear one (RC) in Figure 9.7, it is taken as the base model. Recall that this model has a \mathbf{G} matrix, which is 66×66 and a block diagonal with twenty-two 3×3 blocks. But instead of assuming that \mathbf{R} is equal to a constant variance times the identity matrix, \mathbf{R} is here assumed to have the Power-of-the-Mean structure.

Program

In order to fit this combination of correlated random coefficients and Power-of-the-Mean models with PROC MIXED, you must first create a data set that contains estimates of the fixed effects parameters. This is most easily accomplished by outputting them from an initial fit to the data as follows:

```
proc mixed data=dial;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp / s;
  random int tmp tmp*tmp / type=un sub=sub;
  ods output solutionf=sf;
run;

data cp;
  input est;
  datalines;
  2.24606822
  -3.73117853
  24.08025511
  0.68705846
  -6.82952740
  2.17225966
  1
  0.0663507
run;
```

This program also creates a SAS data set CP that is used to provide a starting covariance parameter values for the full model. Its values are the covariance parameter estimates from the initial model along with a guess of 1 for θ . The estimated residual variance from the initial model fit is divided by 50 to provide an approximation to the new σ^2 . You can now fit the combined model as follows:

```
proc mixed data=dial;
  class qb sub;
  model ufr = tmp|tmp|tmp|tmp qb|tmp|tmp|tmp / s;
  random int tmp tmp*tmp / type=un sub=sub;
  repeated / local=pom(sf);
  parms / pdata=cp;
  ods output solutionf=sf1 covparms=cp1;
run;

proc compare brief data=sf compare=sf1; var est;
data sf; set sf1;
data cp; set cp1;
run;
```

The quadratic random coefficients are specified in the RANDOM statement and the Power-of-the-Mean model is specified using the SF data set in the REPEATED statement. The PARMS statement loads in the initial values for all of the covariance parameters as specified in the CP

data set created earlier. The ODS statement convert tables printed by PROC MIXED into SAS data sets that are used in subsequent processing.

This program is designed to be submitted again and again until β^* in the SF data set equals $\hat{\beta}$ in the SF1 data set. The differences between the two data sets are printed by the call to PROC COMPARE. This iteration algorithm is not guaranteed to converge in general, but when it does, the resulting estimates maximize a restricted version of the pseudo-likelihood described in Chapter 3 of Carroll and Ruppert (1988). The PROC COMPARE call allows you to monitor differences, and the subsequent DATA steps update both the fixed effects and covariance parameter estimates. The latter are saved to improve the stability of the algorithm and to speed convergence of PROC MIXED at each call through the PARMs statement.

Results

The fifth run of the preceding program produces the results given in Output 9.21.

Output 9.21 Power-of-the-Mean Results

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	sub	3.8360
UN(2,1)	sub	-5.8353
UN(2,2)	sub	28.2501
UN(3,1)	sub	1.3778
UN(3,2)	sub	-8.3312
UN(3,3)	sub	2.6970
POM		1.9785
Residual		0.001974

Fit Statistics	
-2 Res Log Likelihood	614.6
AIC (smaller is better)	630.6
AICC (smaller is better)	631.8
BIC (smaller is better)	639.4

PARMS Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	0.00	1.0000

Solution for Fixed Effects						
Effect	qb	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-18.6736	1.0652	18	-17.53	<.0001
tmp		95.2603	4.8072	18	19.82	<.0001
tmp*tmp		-48.6802	6.4162	18	-7.59	<.0001

Solution for Fixed Effects						
Effect	qb	Estimate	Standard Error	DF	t Value	Pr > t
tmp*tmp*tmp		12.6143	3.2717	18	3.86	0.0012
tmp*tmp*tmp*tmp		-1.3567	0.5409	18	-2.51	0.0219
qb	200	1.9929	1.4170	76	1.41	0.1637
qb	300	0
tmp*qb	200	-3.4597	6.4030	76	-0.54	0.5906
tmp*qb	300	0
tmp*tmp*qb	200	0.1371	8.4819	76	0.02	0.9871
tmp*tmp*qb	300	0
tmp*tmp*tmp*qb	200	-1.5487	4.3041	76	-0.36	0.7200
tmp*tmp*tmp*qb	300	0
tmp*tmp*tmp*tmp*qb	200	0.3856	0.7081	76	0.54	0.5877
tmp*tmp*tmp*tmp*qb	300	0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
tmp	1	18	853.48	<.0001
tmp*tmp	1	18	131.39	<.0001
tmp*tmp*tmp	1	18	30.27	<.0001
tmp*tmp*tmp*tmp	1	18	10.81	0.0041
qb	1	76	1.98	0.1637
tmp*qb	1	76	0.29	0.5906
tmp*tmp*qb	1	76	0.00	0.9871
tmp*tmp*tmp*qb	1	76	0.13	0.7200
tmp*tmp*tmp*tmp*qb	1	76	0.30	0.5877

Value Comparison Results for Variables

Obs		Base	Compare	Diff.	% Diff
		Estimate	Estimate		
1		-18.67361	-18.67361	-4.403E-9	2.3579E-8
2		95.26030	95.26030	2.3074E-8	2.4222E-8
3		-48.68024	-48.68024	-3.405E-8	6.9949E-8
4		12.61427	12.61427	1.7395E-8	1.379E-7
5		-1.35668	-1.35668	-2.813E-9	2.0735E-7
6		1.99286	1.99286	3.4927E-9	1.7526E-7
8		-3.45969	-3.45969	-1.763E-8	5.0954E-7
10		0.13706	0.13706	2.4967E-8	0.0000182
12		-1.54872	-1.54872	-1.244E-8	8.0318E-7
14		0.38555	0.38555	1.981E-9	5.1381E-7

Interpretation

The PROC COMPARE results indicate that the estimates have converged. The final estimated power is 1.98, which is very close to a constant coefficient of variation model. Note also that the new value of AIC is 630.6, which is an improvement over the 678.1 value from the standard random coefficients model and is close to the value of 631.7 attained by the ARH(1) structure.

The %NLINMIX macro discussed in Chapter 15 also allows you to specify a Power-of-the-Mean model, and it performs the iterations automatically. In fact, you can use it to specify an arbitrary positive function of the mean as a weight variable in conjunction with any of the aforementioned covariance structures. The final estimates are no longer maximum likelihood, but they do solve a set of pseudo-likelihood generalized estimating equations. %NLINMIX also allows you to specify a nonlinear mean function like the one used by Vonesh and Carter (1992).

The final step in the analysis is to simplify the model by deleting terms that are not significant. Since the ARH(1) covariance structure is nearly as good as the POM covariance structure, the modeling is carried out using ARH(1). The modeling building process deleted the following terms (in order): qb*tmp*tmp*tmp*tmp qb*tmp*tmp*tmp qb*tmp*tmp. The following program fits the reduced fixed effects model with the ARH(1) covariance structure and provides estimates of the means for the levels of qb at several values of tmp.

Program

```
**now build model removed terms as qb*tmp*tmp*tmp*tmp
qb*tmp*tmp*tmp qb*tmp*tmp;
proc mixed data=dial ic;
  class qb sub;
  model ufr = tmp tmp*tmp tmp*tmp*tmp
    tmp*tmp*tmp*tmp qb qb*tmp /outp=pred;
  random sub;
  repeated / type=arh(1) subject=sub;
  lsmeans qb / diff at tmp=.5;
  lsmeans qb / diff at tmp=.6;
  lsmeans qb / diff at tmp=.65;
  lsmeans qb / diff at tmp=.7;
  lsmeans qb / diff at tmp=.75;
  lsmeans qb / diff at tmp=1;
  lsmeans qb / diff at tmp=2;
  lsmeans qb / diff at tmp=3;
  ods output lsmeans=lsm diff=dif;
run;
```

Results

This program produces the results given in Output 9.22.

Output 9.22 ARH(1) Covariance Structure with Simplified Fixed Effects

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
sub		0.6337
Var(1)	sub	1.9986
Var(2)	sub	3.7200
Var(3)	sub	8.2452

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Var(4)	sub	15.7278
Var(5)	sub	18.4853
Var(6)	sub	26.5235
Var(7)	sub	25.6765
ARH(1)	sub	0.7542

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
tmp	1	115	670.81	<.0001
tmp*tmp	1	115	100.96	<.0001
tmp*tmp*tmp	1	115	24.49	<.0001
tmp*tmp*tmp*tmp	1	115	9.17	0.0030
qb	1	115	13.46	0.0004
tmp*qb	1	115	74.16	<.0001

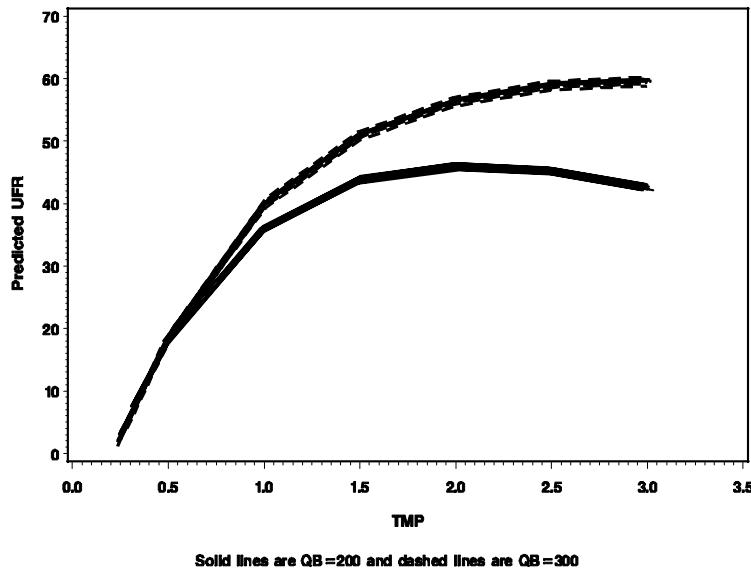
Least Squares Means							
Effect	qb	tmp	Estimate	Standard Error	DF	t Value	Pr > t
qb	200	0.50	18.1085	0.5907	115	30.65	<.0001
qb	300	0.50	18.6393	0.5722	115	32.58	<.0001
qb	200	0.60	22.7858	0.6314	115	36.09	<.0001
qb	300	0.60	23.9806	0.6116	115	39.21	<.0001
qb	200	0.65	24.8966	0.6505	115	38.27	<.0001
qb	300	0.65	26.4233	0.6309	115	41.88	<.0001
qb	200	0.70	26.8646	0.6692	115	40.15	<.0001
qb	300	0.70	28.7232	0.6500	115	44.19	<.0001
qb	200	0.75	28.6964	0.6876	115	41.73	<.0001
qb	300	0.75	30.8870	0.6692	115	46.16	<.0001
qb	200	1.00	36.0322	0.7859	115	45.85	<.0001
qb	300	1.00	39.8826	0.7721	115	51.66	<.0001
qb	200	2.00	45.9189	1.2025	115	38.19	<.0001
qb	300	2.00	56.4083	1.1970	115	47.13	<.0001
qb	200	3.00	42.4829	1.5813	115	26.87	<.0001
qb	300	3.00	59.6114	1.5777	115	37.78	<.0001

Differences of Least Squares Means									
Effect	qb	_qb	tmp	Estimate	Standard Error	DF	t Value	Pr > t	
qb	200	300	0.50	-0.5308	0.7312	115	-0.73	0.4693	
qb	200	300	0.60	-1.1947	0.7495	115	-1.59	0.1137	
qb	200	300	0.65	-1.5267	0.7614	115	-2.01	0.0473	
qb	200	300	0.70	-1.8586	0.7750	115	-2.40	0.0181	
qb	200	300	0.75	-2.1906	0.7903	115	-2.77	0.0065	
qb	200	300	1.00	-3.8504	0.8881	115	-4.34	<.0001	
qb	200	300	2.00	-10.4894	1.4793	115	-7.09	<.0001	
qb	200	300	3.00	-17.1285	2.1855	115	-7.84	<.0001	

Interpretation

The “Differences of Least Squares Means” table compares the levels of qb for a selected set of values of tmp. The *p*-values indicate that the mean response from qb = 300 is significantly larger than the mean response of qb = 200 for tmp > 0.65 and the means are not significantly different for tmp < 0.60. The estimates means are displayed in Figure 9.18.

Figure 9.18 Predicted Models for QB=200 and QB=300



9.6 Example: Combining Between- and Within-Subject Heterogeneity

In this example both within- and between-subject heterogeneity play a role. The data are from Patel (1991) and represent the grip strengths (in mm Hg) of patients with rheumatoid arthritis undergoing two different treatments.

Model

The assumed fixed effects model consists of 12 cell means (2 treatments by 2 genders by 3 measurement occasions), a baseline covariate, and its interaction with treatment, gender, and time. The data set is shown as Data Set 9.6, “Grip,” in Appendix 2, “Data Sets.”

In contrast with the preceding example, some data are missing; that is, some patients are measured on fewer than three occasions. As long as the missingness is ignorable in the sense that the nonresponse mechanism is independent of the subject’s unobserved response (Rubin 1976), a restricted likelihood-based analysis is still appropriate. The following analyses are therefore based on an assumption of ignorable missingness.

Figures 9.19 and 9.20 plot the profiles of the females and males, respectively. No heterogeneity of variance is immediately apparent within either gender, and the amount of variability appears to be roughly the same for both genders.

Figure 9.19 Grip Strength Data for Females

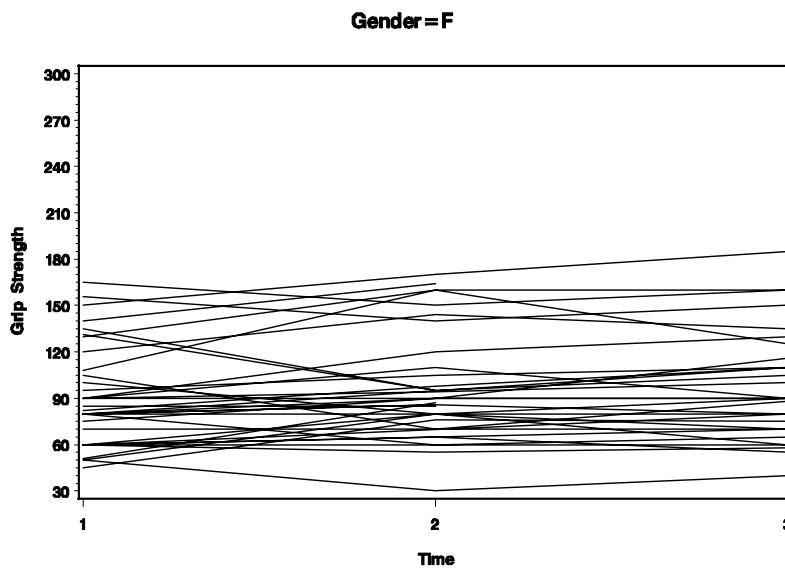
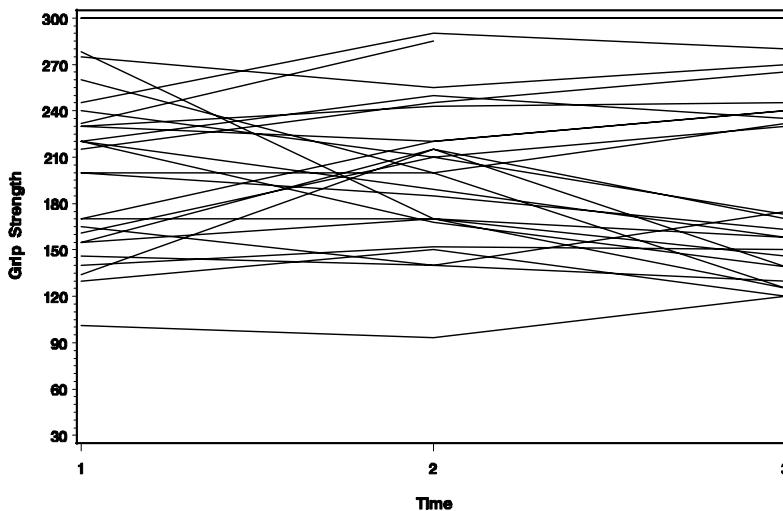


Figure 9.20 Grip Strength Data for Males

Gender=M



9.6.1 Basic Unstructured Covariance Model

The first model to consider is the same as the one used with the previous example: blocks of **R** that have an unstructured form and therefore model completely general within-subject heterogeneity. The PROC MIXED program is as follows:

Program

```
proc mixed data=grip;
  class subject trt Gender time;
  model y = trt|Gender|time x time*x Gender*x trt*x;
  repeated / type=un subject=subject r rcorr;
run;
```

The MODEL statement specifies the 12 cell means involving TRT, GENDER, and TIME. The vertical bar operator (|) is used as a shorthand for all main effects and interactions between TRT, GENDER, and TIME. The mean model also includes different slopes of the baseline covariate *X* for each of the three main effects.

The REPEATED statement requests the block-diagonal unstructured form for **R** with blocks corresponding to levels of the SUBJECT variable. The R and RCORR options request printouts of the first block of **R** and its corresponding correlation form.

Results

The results using the unstructured covariance for observations from the same subject are given in Output 9.23.

Output 9.23 Grip Strength Results Using a Basic Unstructured Model

Estimated R Matrix for subject 1			
Row	Col1	Col2	Col3
1	638.87	337.34	311.94
2	337.34	1003.92	939.04
3	311.94	939.04	1382.69

Estimated R Correlation Matrix for subject 1			
Row	Col1	Col2	Col3
1	1.0000	0.4212	0.3319
2	0.4212	1.0000	0.7970
3	0.3319	0.7970	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	subject	638.87
UN(2,1)	subject	337.34
UN(2,2)	subject	1003.92
UN(3,1)	subject	311.94
UN(3,2)	subject	939.04
UN(3,3)	subject	1382.69

Fit Statistics	
-2 Res Log Likelihood	1691.3
AIC (smaller is better)	1703.3
AICC (smaller is better)	1703.8
BIC (smaller is better)	1716.5

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Trt	1	60	0.13	0.7186
Gender	1	60	2.54	0.1164
trt*Gender	1	60	0.78	0.3820
Time	2	60	1.53	0.2247
trt*time	2	60	0.59	0.5596
Gender*time	2	60	1.94	0.1531

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt*Gender*time	2	60	0.14	0.8691
X	1	60	97.73	<.0001
x*time	2	60	0.84	0.4363
x*Gender	1	60	0.04	0.8415
x*trt	1	60	0.09	0.7716

Interpretation

The blocks of **R** are 3×3 for this example. The estimated variances definitely appear to be increasing with time, but the covariances and correlations display no obvious pattern.

9.6.2 Incorporating Between-Subject Heterogeneity

As in the previous example, you can fit other structures by changing the TYPE= option in the REPEATED statement. Instead of fitting other structures, however, consider the two classification variables (TRT and GENDER) that may be used to model between-subject heterogeneity. The following program specifies separate unstructured estimates for each of the two levels of TRT.

Program

```
proc mixed data=grip;
  class subject trt gender time;
  model y = trt|gender|time x time*x gender*x trt*x;
  repeated / type=un sub=subject group=trt
    r=1,2 rcorr=1,2;
run;
```

The GROUP= option in the REPEATED statement sets up the between-subject heterogeneity. Note also that the first two blocks of **R** and their corresponding correlations are requested with the R and RCORR options.

Results

The results are given in Output 9.24.

This model produces AIC=1705.6, which is worse than the previous homogeneous model. However, using GROUP=GENDER instead of GROUP=TRT in the REPEATED statement produces the results in Output 9.24.

Output 9.24 Grip Strength Results Using Between-Subject Heterogeneity

Estimated R Matrix for subject 1			
Row	Col1	Col2	Col3
1	1065.73	668.92	598.46
2	668.92	1957.79	1909.08
3	598.46	1909.08	2782.19

Estimated R Correlation Matrix for subject 1			
Row	Col1	Col2	Col3
1	1.0000	0.4631	0.3476
2	0.4631	1.0000	0.8180
3	0.3476	0.8180	1.0000

Estimated R Matrix for subject 2			
Row	Col1	Col2	Col3
1	319.92	83.3498	105.33
2	83.3498	294.20	210.96
3	105.33	210.96	275.76

Estimated R Correlation Matrix for subject 2			
Row	Col1	Col2	Col3
1	1.0000	0.2717	0.3546
2	0.2717	1.0000	0.7406
3	0.3546	0.7406	1.0000

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	subject	Gender F	319.92
UN(2,1)	subject	Gender F	83.3498
UN(2,2)	subject	Gender F	294.20
UN(3,1)	subject	Gender F	105.33
UN(3,2)	subject	Gender F	210.96
UN(3,3)	subject	Gender F	275.76
UN(1,1)	subject	Gender M	1065.73
UN(2,1)	subject	Gender M	668.92
UN(2,2)	subject	Gender M	1957.79
UN(3,1)	subject	Gender M	598.46
UN(3,2)	subject	Gender M	1909.08
UN(3,3)	subject	Gender M	2782.19

Fit Statistics	
-2 Res Log Likelihood	1632.9
AIC (smaller is better)	1656.9
AICC (smaller is better)	1658.8
BIC (smaller is better)	1683.3

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	60	0.10	0.7562
Gender	1	60	1.40	0.2412
trt*Gender	1	60	0.64	0.4253
time	2	112	0.83	0.4382
trt*time	2	112	0.50	0.6089
Gender*time	2	112	2.19	0.1165
trt*Gender*time	2	112	0.11	0.8943
x	1	60	147.73	<.0001
x*time	2	112	0.38	0.6862
x*Gender	1	60	0.02	0.9022
x*trt	1	60	0.04	0.8403

Interpretation

Note that AIC = 1656.9 indicates a considerable improvement over the homogeneous unstructured model. The estimates for the two different blocks of **R** reveal that the males are much more variable than the females, with the male variances being (1066, 1958, 2782) and the female variances (320, 294, 276). Furthermore, the male variances are increasing sharply over time, while those for the females are slowly decreasing. The correlation patterns for the two genders are roughly the same.

The preceding statements are in apparent conflict with the observations made in Figure 9.19 and Figure 9.20—namely, that they display no heterogeneity either within or between genders. The resolution of this conflict lies in the fact that the particular mean model used for these data, and primarily the covariate **X**, accounts for a significant portion of the variability, and it is the remaining variability that is modeled in the **R** matrix.

Figures 9.21 and 9.22 correspond to Figures 9.19 and 9.20, except that they plot the ordinary least-squares residuals after fitting the preceding mean model. They reflect the numerical results obtained from the REML analysis. These results show that both the mean and variance models account for variability in the data, and the general mixed model allows you the flexibility to choose which way to model it.

The programs to generate Figures 9.21 and 9.22 are as follows:

```

proc mixed data=grip;
  class subject trt gender time;
  model y = trt|gender|time x time*x gender*x trt*x / outp=p;
  id time subject gender;
run;
symbol i=join c=black r=38;
proc gplot data=p;
  where gender = 'F';
  plot resid*time=subject / nolegend hminor=0;
run;

```

```

proc gplot data=p;
  where gender = 'M';
  plot resid*time=subject / nolegend hminor=0;
run;

```

The OUTP=P option requests that the table of predicted values and residuals be converted to a data set named P. The ID statement is necessary to include extra variables in the P data set. The SYMBOL statement requests that the profiles be interpolated with a straight line and that their color be black. The two PROC GPLOT calls construct Figures 9.21 and 9.22, respectively, with the WHERE statements performing the appropriate subsetting of the P data set. Refer to *SAS/GRAPH Software: Reference* for further details about the SYMBOL statement and PROC GPLOT.

Figure 9.21 Residuals from Grip Strength Data for Females

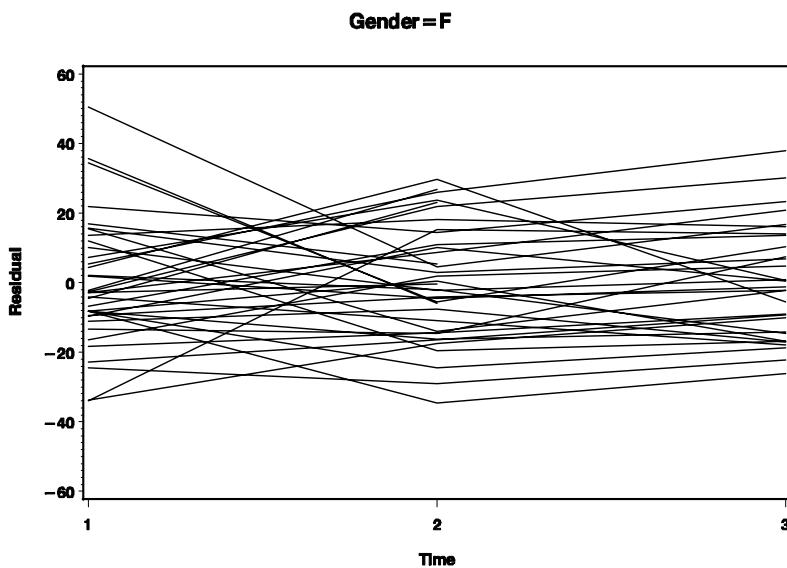
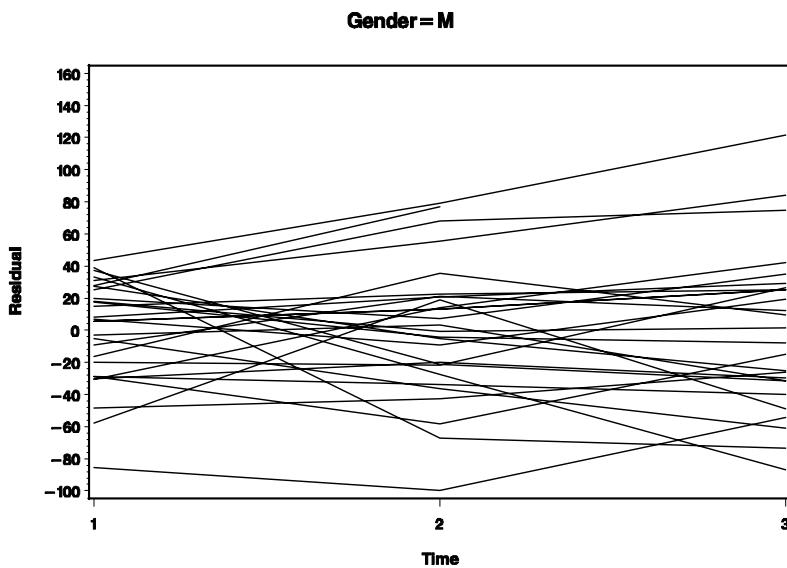


Figure 9.22 Residuals from Grip Strength Data for Males



9.6.3 Heterogeneous Random Coefficients

As a more parsimonious alternative to the previous heterogeneous unstructured model, you can account for both between- and within-subject heterogeneity using a random coefficients model. First consider a random coefficient model with common residual variances.

Program

```
proc mixed data=grip;
  class subject trt gender time;
  model y = trt|gender|time x time*x gender*x trt*x;
  random int t / type=un sub=subject group=gender;
run;
```

Results

The results for the random coefficient model are given in Output 9.25.

Output 9.25 Grip Strength Results Using Between-Subject Heterogeneity Random Coefficients Model

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	subject	Gender F	43.0317
UN(2,1)	subject	Gender F	16.2634
UN(2,2)	subject	Gender F	4.06E-17
UN(1,1)	subject	Gender M	1496.45
UN(2,1)	subject	Gender M	-624.59
UN(2,2)	subject	Gender M	575.20
Residual			237.33

Fit Statistics	
-2 Res Log Likelihood	1653.9
AIC (smaller is better)	1665.9
AICC (smaller is better)	1666.4
BIC (smaller is better)	1679.2

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	53	0.10	0.7509
Gender	1	60	1.57	0.2147
trt*Gender	1	53	0.63	0.4322
time	2	53	1.83	0.1699
trt*time	2	53	0.32	0.7268
Gender*time	2	53	2.15	0.1268
trt*Gender*time	2	53	0.17	0.8458

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
x	1	53	143.59	<.0001
x*time	2	53	1.11	0.3369
x*Gender	1	53	0.00	0.9630
x*trt	1	53	0.04	0.8513

Next consider a random coefficient model with heterogeneous residual variances for the two genders.

Program

```
proc mixed data=grip;
  class subject trt gender time;
  model y = trt|gender|time x time*x gender*x trt*x;
  random int t / type=un sub=subject group=gender;
  repeated / sub=subject group=gender;
run;
```

Results

The results for the random coefficient model with unequal gender variances are given in Output 9.26.

Output 9.26 Grip Strength Results Using Between-Subject Heterogeneity with Heterogeneous Random Coefficients and Heterogeneous Residual Variances

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	subject	Gender F	176.36
UN(2,1)	subject	Gender F	-32.1699
UN(2,2)	subject	Gender F	22.4088
UN(1,1)	subject	Gender M	1072.96
UN(2,1)	subject	Gender M	-433.89
UN(2,2)	subject	Gender M	469.29
Residual	subject	Gender F	146.22
Residual	subject	Gender M	424.67

Fit Statistics	
-2 Res Log Likelihood	1645.6
AIC (smaller is better)	1661.6
AICC (smaller is better)	1662.5
BIC (smaller is better)	1679.2

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	53	0.09	0.7656
Gender	1	60	1.57	0.2157
trt*Gender	1	53	0.60	0.4414
time	2	53	0.68	0.5126
trt*time	2	53	0.35	0.7072
Gender*time	2	53	1.54	0.2245
trt*Gender*time	2	53	0.13	0.8815
x	1	53	147.01	<.0001
x*time	2	53	0.39	0.6803
x*Gender	1	53	0.00	0.9643
x*trt	1	53	0.03	0.8595

Interpretation

Both models fit separate unstructured parameters in **G** for the two genders via the GROUP= option in the RANDOM statements. However, they differ in that the first code models a common residual variance for the two genders, while the latter models separate estimates. The AIC values for the two models are 1665.9 and 1661.6, respectively, favoring the second specification for these data. But the model with unstructured covariance matrices for each level of gender has an AIC value of 1656.9, which is the smallest of all of the covariance structures considered.

9.7 Example: Log-Linear Variance Models

The pre-etch integrated circuit data from Phadke et al. (1983) provide an example. The widths of lines made by a photoresist-nanoline tool are measured in five different locations on silicon wafers with measurements being taken before an etching process. There are eight experimental factors for the experiment: mask dimension, viscosity, spin speed, bake temperature, bake time, aperture, exposure, and developing time. These eight factors are applied to silicon wafers and then the line widths are determined at five locations on each wafer. Refer to Wolfinger and Tobias (1998) for further details.

The experimental design for the wafer experimental unit is derived from an L_{18} orthogonal array with 18 combinations of the eight factors where 15 of the combinations were replicated, for a total of 33 wafers. Each wafer was measured at five locations for a total of 165 observations. Such designs are used to investigate many experimental factors with relatively few experimental runs. The data are given as Data Set 9.7, “Preetch,” in Appendix 2, “Data Sets,” and you can construct the corresponding SAS data set as follows.

```

data preetch;
  input expt wafer mask viscos spin baketemp
         baketime aperture expos develop etch y1-y5;
  y = y1; loc = 'top'; output;
  y = y2; loc = 'cen'; output;
  y = y3; loc = 'bot'; output;
  y = y4; loc = 'lef'; output;
  y = y5; loc = 'rig'; output;
  drop y1-y5;
  datalines;
...datalines...
run;

```

9.7.1 Initial Model

The structure of the data set is that there are two sizes of experimental units, the wafer and the site on a wafer. Thus the basic model has the form of a multi-level model or split-plot model with two error terms. The RESIDUAL measures the variability of sites within a wafer. The wafer error term is specified using a RANDOM statement as in the following program. With only 18 combinations of the 8 factors, a main-effects-only model is fit. But the location is measured on each of the wafers, so LOC and LOC interacting with each of the 8 factors applied to the wafers are included as fixed effects in the model.

Program

```

title "Initial Model with Equal Variances";
proc mixed data=preetch ic;
  class expt wafer mask viscos spin baketemp
             baketime aperture expos develop loc;
  model y = mask viscos spin baketemp
             baketime aperture expos develop
             loc loc*mask loc*viscos loc*spin loc*baketemp
             loc*baketime loc*aperture
             loc*expos loc*develop / outp=pred ddfm=kr;
  random int / sub=wafer(expt);
run;

```

Results

The results of the split-plot model are given in Output 9.27.

Output 9.27 Results for Initial Model with Independent Errors

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	wafer(expt)	0.01755
Residual		0.005795

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
-29.3	2	-25.3	-25.2	-24.3	-22.3	-20.3

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
mask	1	19	61.83	<.0001
viscos	1	19	84.07	<.0001
spin	2	19	36.69	<.0001
baketemp	1	19	1.52	0.2333
baketime	2	19	0.24	0.7872
aperture	2	19	1.77	0.1977
expos	2	19	27.09	<.0001
develop	2	19	12.98	0.0003
loc	4	76	21.43	<.0001
mask*loc	4	76	5.98	0.0003
viscos*loc	4	76	2.07	0.0932
spin*loc	8	76	2.76	0.0099
baketemp*loc	4	76	0.61	0.6551
baketime*loc	8	76	0.37	0.9344
aperture*loc	8	76	2.80	0.0090
expos*loc	8	76	1.98	0.0607
develop*loc	8	76	0.92	0.5065

Interpretation

The wafer error term is specified by RANDOM INT/SUBJECT=WAFER(EXPT). The main effects for the 8 wafer factors account for 13 degrees of freedom, leaving 19 degrees of freedom to be associated with the wafer error term. In the “Type 3 Tests of Fixed Effects” table, there are 19 degrees of freedom for the denominator degrees of freedom for the eight main effects and 76 degrees of freedom for error for all effects involving LOC. The estimates of the variance components are in the “Covariance Parameter Estimates” table, $\hat{\sigma}_{\text{wafer}}^2 = 0.01755$ and $\hat{\sigma}_{\text{loc}}^2 = 0.005795$, and the value of AIC is -25.3.

9.7.2 Full Model with Power-of-X Dispersion Effects

Because any of the experimental factors can be influencing the line-width response, the next model fit includes all of the factors as both location (fixed) effects and dispersion effects. For this example, we use the log-linear variance model described by Harvey (1976) and Aitkin (1987), where the variances are of the form

$$\mathbf{R} = \text{diag}[\sigma^2 \exp(\mathbf{u}_i' \mathbf{d})]$$

where

σ^2 is an unknown scale parameter

\mathbf{u}_i is the i^{th} row of some known design matrix

\mathbf{d} is an unknown vector of parameters, sometimes called **dispersion effects parameters**

The \mathbf{u}_i 's provide flexibility in modeling the variances as a function of known experimental conditions. These models appear to be applicable to many situations, including industrial experiments, where modeling and controlling variability are important objectives.

You can specify dispersion effects in PROC MIXED with the LOCAL=EXP option in the REPEATED statement as follows:

Program

```
title "Model with Local=exp(all terms)";
proc mixed data=preetch covtest ic;
  class expt wafer mask viscos spin baketemp
    baketime aperture expos develop loc;
  model y = mask viscos spin baketemp
    baketime aperture expos develop loc loc*mask
    loc*viscos loc*spin loc*baketemp loc*baketime
    loc*aperture loc*expos loc*develop /
    outp=pred ddfm=kr;
  random int / sub=wafer(expt);
  repeated / subject=wafer(expt)
    local=exp(mask viscos spin baketemp baketime
    aperture expos develop);
run;
```

The LOCAL=EXP option enables you to specify effects after it, just as you would in the MODEL or RANDOM statements. PROC MIXED parses the effects in the same fashion and uses them to construct the \mathbf{u}_i vectors and the corresponding \mathbf{d} parameters to be estimated. This model thus lets you actually model the variance of the data as a function of experimental effects that may also be influencing the mean of the data. Note that the same effects are included both in the MODEL statement and as the argument to the LOCAL=EXP option. The output from this program is shown in Output 9.28.

Results

The results are given in Output 9.28.

Output 9.28 Log-Linear Variance Model: Results for Full Model

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	wafer(expt)	0.01737	0.006032	2.88	0.0020
EXP mask		-0.1326	0.2210	-0.60	0.5485
EXP viscos		0.05895	0.2554	0.23	0.8175
EXP spin 1		0.01202	0.2653	0.05	0.9639
EXP spin 2		-0.2087	0.3266	-0.64	0.5228
EXP baketemp		-0.2005	0.2125	-0.94	0.3453
EXP baketime 1		-0.3548	0.2741	-1.29	0.1956
EXP baketime 2		0.006765	0.2732	0.02	0.9802
EXP aperture 1		-0.2451	0.3039	-0.81	0.4199
EXP aperture 2		0.2148	0.2720	0.79	0.4296

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
EXP expos 1		-0.01568	0.2709	-0.06	0.9538
EXP expos 2		0.5448	0.2899	1.88	0.0602
EXP develop 1		-0.5953	0.3213	-1.85	0.0639
EXP develop 2		0.4381	0.2875	1.52	0.1276
Residual		0.004604	0.001117	4.12	<.0001

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
-44.2	15	-14.2	-8.1	-6.6	8.2	23.2

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
mask	1	18.8	63.25	<.0001
viscos	1	18.7	85.78	<.0001
spin	2	18.8	37.53	<.0001
baketemp	1	19.2	1.57	0.2255
baketime	2	18.8	0.24	0.7879
aperture	2	18.7	1.81	0.1915
expos	2	18.8	27.97	<.0001
develop	2	18.8	13.33	0.0002
loc	4	45.2	27.70	<.0001
mask*loc	4	42.1	6.26	0.0005
viscos*loc	4	25.6	3.09	0.0335
spin*loc	8	44.6	2.87	0.0116
baketemp*loc	4	41.7	0.67	0.6181
baketime*loc	8	32.7	0.99	0.4649
aperture*loc	8	49.4	2.90	0.0098
expos*loc	8	37.2	1.83	0.1032
develop*loc	8	39.9	1.08	0.3935

Interpretation

The key statistics to examine in such analyses are the *p*-values associated with each of the location and dispersion effects. Significant *p*-values are those that are likely to affect process mean and variability, respectively.

For the location effects, the “Tests of Fixed Effects” table reveals that MASK, VISCOS, SPIN, BAKETIME, EXPOS, and DEVELOP all appear to be significant. Also, LOC and the interaction of LOC with MASK, VISCOS, SPIN, and APERTURE appear to be significant. The analysis of dispersion effects in the “Covariance Parameter Estimates” table are more difficult

because some of the effects are modeled by two parameters in **d**, because they have three levels in the experiment. A conservative approach is to retain effects for which any of their associated parameters are significant. This approach declares EXPOS and DEVELOP as possible significant dispersion effects.

Alternatively, you can compute likelihood ratio tests of the location and dispersion effects by fitting a series of reduced models, leaving out one effect in each. For these fits you should use standard maximum likelihood (with the METHOD=ML option in the PROC MIXED statement) for the location effect likelihood ratio tests because the restricted likelihood adjustment depends on the location effects design matrix. Restricted likelihood ratio tests are appropriate for the dispersion effects. The AIC for this model is -14.2, a value that is larger than the -25.3 associated with the first or basic model, indicating that this dispersion model is not describing the variability in the data as well as the simple model. An improvement in the AIC is possible if unimportant terms in the dispersion part of the model are removed. As a final step, the dispersion effects part of the model is simplified using a stepwise deletion process.

9.7.3 Reduced Model with Power-of-X Dispersion Effects

A stepwise deletion process was used to delete terms from the dispersion effects part of the model, starting with the terms that had the largest significance levels. The DEVELOP and EXPOS terms were retained as indicated by the following program.

Program

```
title "Model with Local=exp(reduced terms) ";
proc mixed data=preetch ic;
  class expt wafer mask viscos spin baketemp
    baketime aperture expos develop loc;
  model y = mask viscos spin baketemp
    baketime aperture expos develop loc loc*mask
    loc*viscos loc*spin loc*baketemp loc*baketime
    loc*aperture loc*expos loc*develop/outp=pred ddfm=kr;
  random int / sub=wafer(expt);
  repeated/subject=wafer(expt) local=exp( develop expos );
run;
```

Results

The results are given in Output 9.29.

Output 9.29 Results for Reduced Model with EXP Dispersion Effects

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	wafer(expt)	0.01735
EXP develop 1		-0.4693
EXP develop 2		0.4369
EXP expos 1		0.07291
EXP expos 2		0.4475
Residual		0.004895

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
-38.1	6	-26.1	-25.2	-23.1	-17.2	-11.2

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
mask	1	18.6	63.21	<.0001
viscos	1	18.8	84.69	<.0001
spin	2	18.8	37.12	<.0001
baketemp	1	18.8	1.63	0.2168
baketime	2	18.8	0.25	0.7827
aperture	2	18.7	1.80	0.1926
expos	2	18.7	27.74	<.0001
develop	2	18.7	13.23	0.0003
loc	4	74.3	24.54	<.0001
mask*loc	4	69.1	6.67	0.0001
viscos*loc	4	64.1	2.56	0.0468
spin*loc	8	71.8	3.06	0.0051
baketemp*loc	4	64.9	0.73	0.5756
baketime*loc	8	66.8	0.95	0.4809
aperture*loc	8	68.5	2.92	0.0074
expos*loc	8	58.9	2.35	0.0292
develop*loc	8	57.5	1.06	0.4045

Interpretation

The estimates of the variance components are given in the “Covariance Parameter Estimates” table of Output 9.28, where $\hat{\sigma}_{\text{wafer}}^2 = 0.01735$ and $\hat{\sigma}_{\text{loc}}^2 = 0.004895$, compared to 0.01755 and 0.005795 for the initial model (Output 9.27). Thus, modeling the dispersion with the REPEATED statement reduced the estimate of σ_{loc}^2 but did not have much effect on σ_{wafer}^2 . The AIC of this model with the reduced dispersion part is -26.1, a value that is a little smaller than the initial model. From the “Type 3 Tests of Fixed Effects” table you see that one additional term is significant, EXPOS × LOC ($p < 0.05$).

9.7.4 Model with UN Repeated Measures Error Structure

As a final look at this data set, consider it to be a repeated measures problem where the subject is the wafer and the locations on a wafer are the repeated measurements. The following program carries out the repeated measures analysis using the UN covariance structure.

Program

```

title "UN cov matrix for locs";
proc mixed data=preetch covtest ic;
  class expt wafer mask viscos spin baketemp
    baketime aperture expos develop loc;
  model y = mask viscos spin baketemp baketime aperture
    expos develop loc loc loc*mask loc*viscos
    loc*spin loc*baketemp loc*baketime
    loc*aperture loc*expos loc*develop/ ddfm=bw;
  repeated loc/sub=wafer(expt) type=un;
run;

```

Results

The results of the repeated measures analysis are given in Output 9.30.

Output 9.30 Results for Model with UN Repeated Measures

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	wafer(expt)	0.02128
UN(2,1)	wafer(expt)	0.01752
UN(2,2)	wafer(expt)	0.02414
UN(3,1)	wafer(expt)	0.01481
UN(3,2)	wafer(expt)	0.01549
UN(3,3)	wafer(expt)	0.01676
UN(4,1)	wafer(expt)	0.02034
UN(4,2)	wafer(expt)	0.02309
UN(4,3)	wafer(expt)	0.02032
UN(4,4)	wafer(expt)	0.03388
UN(5,1)	wafer(expt)	0.01468
UN(5,2)	wafer(expt)	0.01593
UN(5,3)	wafer(expt)	0.01447
UN(5,4)	wafer(expt)	0.01881
UN(5,5)	wafer(expt)	0.02064

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
-40.3	15	-10.3	-4.2	-2.8	12.1	27.1

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
mask	1	19	61.83	<.0001
viscos	1	19	84.07	<.0001
spin	2	19	36.69	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
baketemp	1	19	1.52	0.2333
baketime	2	19	0.24	0.7872
aperture	2	19	1.77	0.1977
expos	2	19	27.09	<.0001
develop	2	19	12.98	0.0003
loc	4	19	25.94	<.0001
mask*loc	4	19	6.13	0.0024
viscos*loc	4	19	2.27	0.0994
spin*loc	8	19	2.75	0.0337
baketemp*loc	4	19	0.78	0.5525
baketime*loc	8	19	0.30	0.9584
aperture*loc	8	19	2.52	0.0467
expos*loc	8	19	2.20	0.0755
develop*loc	8	19	0.97	0.4892

Interpretation

The value of AIC for the repeated measures analysis is -10.3 , indicating that it is not as good as the models with the EXP dispersion model. The significant fixed effects are identical to those in the analyses of the initial model and the full dispersion model.

The information criteria for each of the four models are listed in Output 9.31. The reduced dispersion model has the smallest AIC and AICC values, while the initial model has the smallest HQIC, BIC, and CAIC values. Thus, the model you select for the final analysis depends on your choice of information criteria.

Output 9.31 Information Criteria for the Four Covariance Structures

Obs	Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC	model
1	-29.3	2	-25.3	-25.2	-24.3	-22.3	-20.3	Initial
2	-44.2	15	-14.2	-8.1	-6.6	8.2	23.2	Full
3	-38.1	6	-26.1	-25.2	-23.1	-17.2	-11.2	Reduced
4	-40.3	15	-10.3	-4.2	-2.8	12.1	27.1	UN

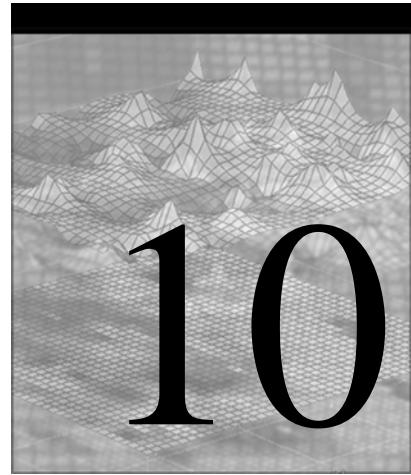
Output 9.32 contains the significance levels for the fixed effects part of the model for the four models fit to the data. All of the models have significant MASK, VISCOS, SPIN, EXPOS, DEVELOP, LOC, LOC \times MASK, LOC \times SPIN, and LOC \times APERTURE effects ($p < 0.05$). The full and reduced models have a significant VISCOS \times LOC interaction, and the reduced model has a significant EXPOS \times LOC interaction. Thus, the significant factors in the model depend on which form of the dispersion function you select. In this case, the important terms in the model depend on your choices of information criteria.

Output 9.32 Significance Levels of Type 3 Fixed Effect F -tests for the Four Covariance Structures

Obs	Effect	NumDF	Initial	Full	Reduced	UN
1	mask	1	0.00000	0.00000	0.00000	0.00000
2	viscos	1	0.00000	0.00000	0.00000	0.00000
3	spin	2	0.00000	0.00000	0.00000	0.00000
4	baketemp	1	0.23326	0.22554	0.21679	0.23326
5	baketime	2	0.78722	0.78792	0.78267	0.78722
6	aperture	2	0.19770	0.19152	0.19263	0.19770
7	expos	2	0.00000	0.00000	0.00000	0.00000
8	develop	2	0.00028	0.00025	0.00026	0.00028
9	loc	4	0.00000	0.00000	0.00000	0.00000
10	mask*loc	4	0.00031	0.00048	0.00013	0.00242
11	viscos*loc	4	0.09320	0.03351	0.04681	0.09943
12	spin*loc	8	0.00989	0.01157	0.00511	0.03368
13	baketemp*loc	4	0.65507	0.61806	0.57557	0.55254
14	baketime*loc	8	0.93436	0.46494	0.48086	0.95841
15	aperture*loc	8	0.00903	0.00979	0.00737	0.04671
16	expos*loc	8	0.06070	0.10316	0.02917	0.07552
17	develop*loc	8	0.50651	0.39347	0.40455	0.48922

9.8 Summary

PROC MIXED has the capability of fitting several types of unequal variance models to data. The classical method of analyzing data with unequal variances was to find a transformation that eliminated the problem and then to carry out the analysis on the transformed data. With PROC MIXED you can model the unequal variances and then estimate and make comparisons among the fixed effects with the selected unequal variance model. This chapter considered heterogeneous variance models—that is, models for which the variability of the data is assumed to change according to one or more factors. The first two examples involved differences in the variances for a two-way analysis of variance and a simple linear regression model. The third example provided a case where the unequal variances occur for a factor specified by a RANDOM statement. For data that consist of repeated observations on a number of subjects, heterogeneity can be classified as either within-subject or between-subject. The fourth example, on high-flux hemodialyzers, illustrated the wide variety of possible within-subject covariance structures and methods of selecting between them. The fifth example, on grip strengths, showed how between-subject heterogeneity can also play an important role in data modeling. The sixth and final example, on integrated circuits, highlighted the log-linear variance model and how it allows you to directly model heterogeneity.



Mixed Model Diagnostics

10.1 Introduction	413
10.2 From Linear to Linear Mixed Models	415
10.2.1 Residuals in the Linear Model	415
10.2.2 Influence Measures in the Linear Model	417
10.2.3 Random Effects and the Consequences of Generalized Least Squares	418
10.3 The Influence Diagnostics	424
10.3.1 Overall Influence	424
10.3.2 Influence on the Parameter Estimates	424
10.3.3 Influence on the Precision of Estimates	425
10.3.4 Iterative or Non-iterative Analysis	426
10.4 Example: Unequally Spaced Repeated Measures.....	426
10.4.1 No Temporal Effects in Mean Structure.....	427
10.4.2 With Temporal Effects in Mean Structure.....	433
10.5 Summary.....	435

10.1 Introduction

A statistical model is the mathematical manifestation of a data generating mechanism. A model that fits the data well represents a mechanism that could have produced the observed data. Naturally, we raise questions as to what extent the model and data agree, whether the model needs to be adjusted, whether the data need to be adjusted (e.g., outlier removal), or both. We think of the **model** in very general terms in this chapter. The specification of a model includes the choice of response, transformations, fixed effects, random effects, distributional assumptions, covariance structure, etc. In the classical linear model, where diagnostic tools have a long history, specific methods have been developed to inquire about specific model

assumptions and breakdowns. For example, the graphical examination of **fitted residuals** is used to assess distributional assumptions such as variance homogeneity and lack of serial correlation, and to examine the need to add or transform fixed model effects. Summary measures and statistics based on residuals are used to test a model's **goodness of fit**. The interrelationship between fixed effects and their impact on the analysis is studied with **collinearity** diagnostics. The importance and weight of individual observations on the analysis are gauged with **influence** diagnostics.

It is often helpful to think of these diagnostic techniques as **perturbation** analysis. You are interested in gauging the stability of the model output under perturbations of its inputs. For example, influence analysis perturbs the data by removing observations from the analysis; residual plots are often compared after adding or removing regressor variables, a perturbation of the fixed effects structure. The challenge is to determine when changes in model output are substantive enough to warrant reformulation of the model or data without developing a model that is overly rigid or crafted too closely to the data at hand.

Maybe surprisingly, comparatively little work in this area has been done for mixed models. The reasons for this are explored in detail in Section 10.2. Briefly, mixed models have these characteristics:

- They are considerably more difficult to fit than linear models, typically requiring iterative optimization.
- They have more model components.
- They can have many more residuals.
- They have conditional and marginal distributions.
- They are often applied to data with clustered structure (independent subjects).

This chapter presents residual and influence diagnostics for linear mixed models and their connection to similar or identical measures for the linear model. Special emphasis is placed on understanding the connection to the linear model as well as the important differences in residual and influence analysis. Throughout, it is important to remember that the results of diagnostic analysis depend on the model. For example, an observation can be highly influential and/or an outlier because the model is not correct. The appropriate action may be to change the model, not to remove the data point. Outliers can be the most important and noteworthy data points, since they can point to a model breakdown. The task is to develop a model that fits the data, not to develop a set of data that fits a particular model.

Section 10.2 explores the important differences between the linear and the linear mixed model. Sections 10.3 and 10.4 examine the statistics for residual and perturbation analysis, respectively. The examples rely on the GLIMMIX¹ procedure and the residual and influence diagnostics in the MIXED² procedure.

¹ The GLIMMIX procedure is an add-on in SAS 9.1 to SAS/STAT for the (32-bit) Windows platform. It does not ship with SAS 9.1. You can obtain the GLIMMIX procedure for SAS 9.1 as a download from www.sas.com/statistics. This site also contains the documentation for the GLIMMIX procedure.

² Residual and influence analyses and graphical displays in the MIXED procedure are experimental in SAS 9.1, meaning that tabular and graphical output and the syntax for requesting them are subject to change in a future release.

10.2 From Linear to Linear Mixed Models

10.2.1 Residuals in the Linear Model

In the standard linear model with uncorrelated errors

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \quad \mathbf{e} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$$

residual analysis starts with the fitted residuals (the **raw** residuals)

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\beta}$$

where $\hat{\beta}$ is the ordinary least-squares estimate

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The properties of the OLS residuals are easily studied. For example, $\hat{\mathbf{e}}$ has zero mean and, if the model contains an intercept, the OLS residuals also sum to zero. It is thus customary to examine the residuals in order to learn about the properties of the unobservable model errors \mathbf{e} .

Unfortunately, $\hat{\mathbf{e}}$ does not share many properties with \mathbf{e} beyond a zero mean. For example, the OLS residuals are not homoscedastic and uncorrelated; their variance is

$$\text{Var}[\hat{\mathbf{e}}] = \sigma^2 \mathbf{M} = \sigma^2 (\mathbf{I} - \mathbf{H}) = \sigma^2 (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

a non-diagonal matrix with possibly heterogeneous values on the diagonal. Furthermore, the fitted residuals are rank-deficient; if \mathbf{X} has rank k , then there can be only $n-k$ linearly independent fitted residuals. To overcome these problems with the raw residuals, manipulations and transformations of the vector of fitted residuals are common.

10.2.1.1 Studentization

If h_{ii} is the i th diagonal element of the **leverage** (“**hat**”) matrix \mathbf{H} , then the i th **standardized** residual is obtained by dividing it by its standard deviation,

$$\frac{\hat{e}_i}{\sqrt{\text{Var}[\hat{e}_i]}} = \frac{\hat{e}_i}{\sqrt{\sigma^2(1-h_{ii})}}$$

The standardized residual is of theoretical, not practical interest, because σ^2 is unknown. Replacing the error variance with its estimate yields the **studentized** residual

$$r_i = \frac{\hat{e}_i}{\sqrt{\hat{\text{Var}}[\hat{e}_i]}} = \frac{\hat{e}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}}$$

The studentized residual remedies one problem of the raw residual, variance heterogeneity. If the estimate of the variance is replaced with an estimate that does not use the i^{th} observation, a **leave-one-out** estimate of σ^2 , the **externally studentized** residual t_i results:

$$t_i = \frac{\hat{e}_i}{\sqrt{\hat{\sigma}_{-i}^2(1-h_{ii})}}$$

The subscript notation $-i$ is intended to signify that the estimate was obtained without the participation of the i^{th} observation. External studentization also stabilizes the variance of the residual. In addition, if the data are normally distributed, the t_i follow a t -distribution—a fact sometimes used to compare t_i 's to cutoffs from t -distributions.

10.2.1.2 Error Recovery

The standardized and studentized residuals combat a single shortcoming of the raw residuals. A different route is taken by methods that simultaneously rectify more than one issue. Schabenberger and Gotway (2005, Ch. 6.1.2) collect them under the term **error recovery**. An error recovery method generates $n-k$ variance-covariance stabilized residuals that have zero mean and constant variance, and are uncorrelated. Residuals in this class are the recursive or sequential residuals (Brown, Durbin, and Evans 1975, Kianifard and Swallow 1996) and the linearly unbiased scaled (LUS) estimates of Theil (1971). Residual recursion is the process of fitting the model initially to k data points and then adding the remaining data points in turn. The j^{th} recursive residual is the scaled difference between y_j and \hat{y}_{-j} , the predicted values based on previous observations. Recursive residuals are useful to detect outliers, serial correlation, and departures from the constant variance assumptions. A drawback is their dependence on which k observations are initially used to fit the model.

LUS estimates apply error recovery based on projections. We are seeking an $(n \times n)$ matrix \mathbf{Q} such that the first $n-k$ elements of $\mathbf{Q}'\hat{\mathbf{e}}$ have unit variance and are uncorrelated. This process is also known as **whitening** the residuals. The matrix \mathbf{Q} can be constructed from a spectral decomposition of the variance matrix $\sigma^2\mathbf{M}$ of $\hat{\mathbf{e}}$. You can obtain recursive residuals and LUS estimates in fixed effects models with the AUTOREG procedure in SAS/ETS.

In general, if the random vector \mathbf{u} has variance \mathbf{A} , and \mathbf{A} is positive-definite, then you can use the lower triangular Cholesky root \mathbf{C} of \mathbf{A} to construct a vector of uncorrelated random variables. Since $\mathbf{CC}' = \mathbf{A}$, $\text{Var}[\mathbf{C}'\mathbf{u}] = \mathbf{I}$. This idea can not readily be applied to a vector of fitted residuals, since \mathbf{M} is a rank-deficient matrix. But you can compute the Cholesky root of the estimate of $\text{Var}[\mathbf{Y}]$. In the linear model this leads to the simple matrix $\sigma\mathbf{I}$. Applying the inverse Cholesky root of the variance of the data to the fitted residuals leads to the **scaled** or **Cholesky** residuals

$$\mathbf{C}^{-1}\hat{\mathbf{e}}$$

Like the LUS estimates, Cholesky residuals depend on the data order, and it is not possible to associate the j^{th} recovered error with a single observation (let alone the j^{th} observation). These residuals are linear combinations of one or more residuals and cannot be used to identify outliers. LUS estimates produce uncorrelated residuals, but Cholesky residuals are not uncorrelated, because they are scaled not by the inverse Cholesky root of *their* variance, but by the inverse root of the variance of the data. LUS and Cholesky residuals are well suited to

diagnose deviations from the normality assumption (see, e.g., Houseman, Ryan, and Coull 2004 and Fitzmaurice, Laird, and Ware 2004, Ch. 9). In the linear model with uncorrelated errors, for example, the i^{th} Cholesky residual is identical to the **Pearson** residual,

$$\frac{\hat{e}_i}{\sqrt{\hat{\sigma}^2}}$$

10.2.2 Influence Measures in the Linear Model

The goal of influence analysis is to determine how observations impact the analysis. Of course, all observations are influential, in the sense that their removal changes the numerical results of the analysis. The goal is to identify those observations that are *so* influential that their presence or absence from the data changes an important aspect of the analysis, yields qualitatively different inferences, or violates assumptions of the model. The goal is *not* to determine observations for removal from the analysis, but to determine which cases exert undue influence on the analysis and in what form. You thus need to decide prior to a perturbation analysis **what matters**. If removing a data point changes considerably the estimate of the residual variance but not the fixed effect solutions, and the goal of the analysis is prediction, the data point does not present a problem. If, however, you are also interested in confidence intervals for predicted values, the influence on the residual variance estimate matters.

The basic idea of perturbation analysis is simple: Remove one or more observations, recompute the important model output quantities, and quantify the change. In the linear model without random effects this process is made simple by the following three facts:

1. The fixed effects solution does not depend on covariance parameters; $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ does not depend on σ^2 .
2. The change in $\hat{\beta}$ after removing the i^{th} observation can be computed without refitting the model. It is available based upon quantities from the overall fit:

$$\hat{\beta}_{-i} = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \hat{e}_i / (1 - h_{ii}).$$
3. Many important influence diagnostics are closely related; they can be derived from the same basic quantities: the raw residual \hat{e}_i and the leverage h_{ii} .

For example, Cook's D statistic (Cook 1977, 1979), which measures the scaled change in the fixed effects solutions following removal of observation i , can be written in terms of the studentized residual and the leverage,

$$D = \frac{r_i^2 h_{ii}}{\text{rank}(\mathbf{X})(1 - h_{ii})}$$

The DFFITS statistic, which measures the change in fit in terms of standard error units (Belsley, Kuh, and Welsch 1980), can be written in terms of the externally studentized residual and the leverage

$$DFFITS_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}}$$

10.2.3 Random Effects and the Consequences of Generalized Least Squares

10.2.3.1 Marginal and Conditional Residuals

When your model contains random effects, or when the \mathbf{R} matrix has a structure more complicated than $\sigma^2\mathbf{I}$, residual and influence diagnostics are considerably more difficult than in the linear model. In the presence of random effects, there are now two sets of residuals. In the mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, with normal random effects and errors, the conditional distribution of $\mathbf{Y} | \mathbf{u}$ has mean $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ and variance \mathbf{R} . The marginal distribution of \mathbf{Y} has mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\mathbf{ZGZ}' + \mathbf{R}$. Consequently, we can define the **conditional** residuals

$$\mathbf{r}_c = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$$

that measure deviations from the conditional mean and **marginal** residuals

$$r_m = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{r}_c + \mathbf{Z}\hat{\mathbf{u}}$$

that measure deviations from the overall mean. Marginal and conditional residuals have different standardization, and studentization, since the variances of \mathbf{r}_m and \mathbf{r}_c differ. As a consequence there are two versions of each of the basic residuals in Section 10.2.1—studentized marginal, studentized conditional, externally studentized marginal, externally studentized conditional residuals, and so forth. More importantly, the properties of the marginal and conditional residuals differ greatly, and the marginal residuals in the mixed model can behave differently from *the* residuals in the linear model. The latter is a consequence of a non-diagonal variance matrix $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$. The fixed effects solution in the linear mixed model are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

Because \mathbf{V} is estimated, this is an **estimated generalized least squares** (EGLS) estimator. The marginal residuals have zero mean in the mixed model, as they do in the standard linear model,

$$E[\mathbf{r}_m] = E[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}] = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

But even if the model contains an intercept, the residuals do not necessarily sum to zero for your particular data. The residuals will sum to zero in certain balanced situations, as in the following example.

10.2.3.2 Example: A Split-Plot Design

To highlight the difference between conditional and marginal residuals in the mixed model and the corresponding quantities in a fixed effects model, we consider a split-plot experiment discussed by Littell, Freund, and Spector (1991). The experiment was conducted using four blocks. Each block was divided into halves, and cultivar (CULT, 2 levels) A or B was randomly assigned to each half (or whole-plot unit). Each whole-plot unit consisted of three plot, or split-plot, units. Three inoculation treatments (INOC) were randomly assigned to sub-plot units

within each whole plot. Hence, the block halves represent the whole-plot experimental units and the CULT factor is the whole plot factor. The inoculation treatments compose the sub-plot factor. The data for this experiment are given as Data Set 10.2, “Cultivar-Inoculation Trial,” in Appendix 2, “Data Sets.”

Program

We fit this model with the GLIMMIX procedure because of the ease with which you can produce residual graphics. The same CLASS, MODEL, and RANDOM statements can be used to fit the model with PROC MIXED.

```
ods html;
ods graphics on;
proc glimmix data=cultspd
    plots=(studentpanel(type=noblup)
           studentpanel(type=blup));
    class block cult inoc;
    model drywt = cult inoc cult*inoc / ddfm=satterth;
    random block block*cult;
run;
ods graphics off;
ods html close;
```

The ODS GRAPHICS statement requests that statistical graphics be produced by the following procedure invocations. The ODS HTML statement selects a destination for the display of the listing and graphical output.

The PLOTS= option in the PROC GLIMMIX statement requests two panel plots of studentized residuals. You can use the TYPE= option of the STUDENTPANEL plot to determine what type of residual to construct. The NOBLUP type results in the marginal residuals \mathbf{r}_m . The BLUP type produces the conditional residuals \mathbf{r}_c .

Results

Output 10.1 displays partial output from the GLIMMIX procedure for the split-plot analysis (these results agree with those from PROC MIXED). The panel of graphics of the studentized residuals appear in Figures 10.1 and 10.2.

Output 10.1 Results for Split-Plot Analysis

Fit Statistics	
-2 Res Log Likelihood	65.06
AIC (smaller is better)	71.06
AICC (smaller is better)	72.78
BIC (smaller is better)	69.22
CAIC (smaller is better)	72.22
HQIC (smaller is better)	67.02
Generalized Chi-Square	12.70
Gener. Chi-Square / DF	0.71

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
block	0.8800	1.2264
block*cult	0.8182	0.8654
Residual	0.7054	0.2880

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
cult	1	3	0.76	0.4471
inoc	2	12	83.76	<.0001
cult*inoc	2	12	1.29	0.3098

Figure 10.1 Marginal Studentized Residuals in Split-Plot Analysis

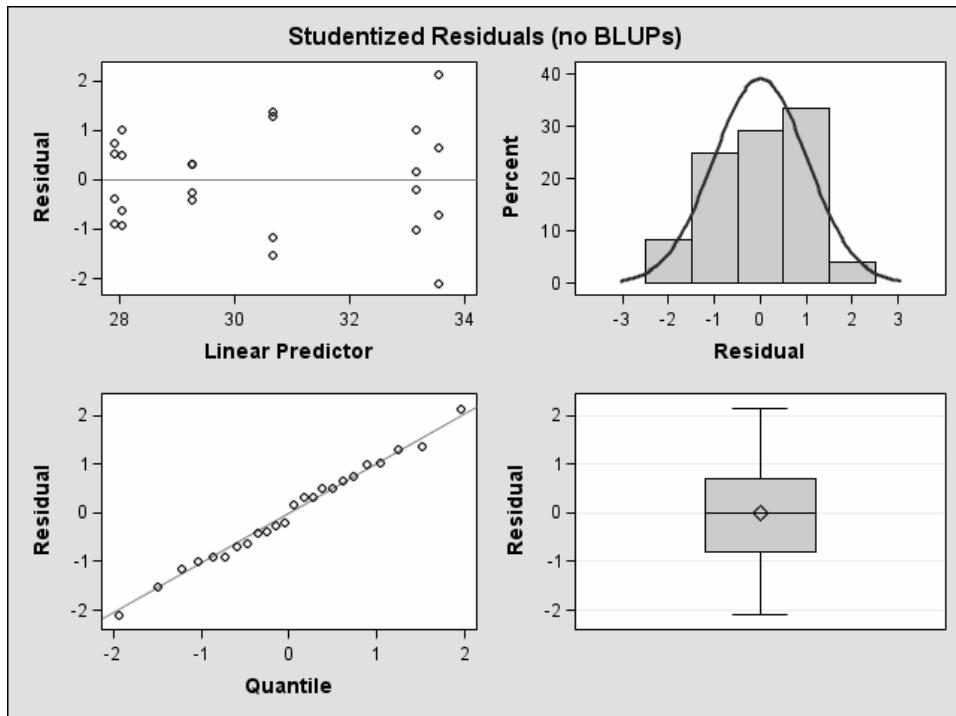
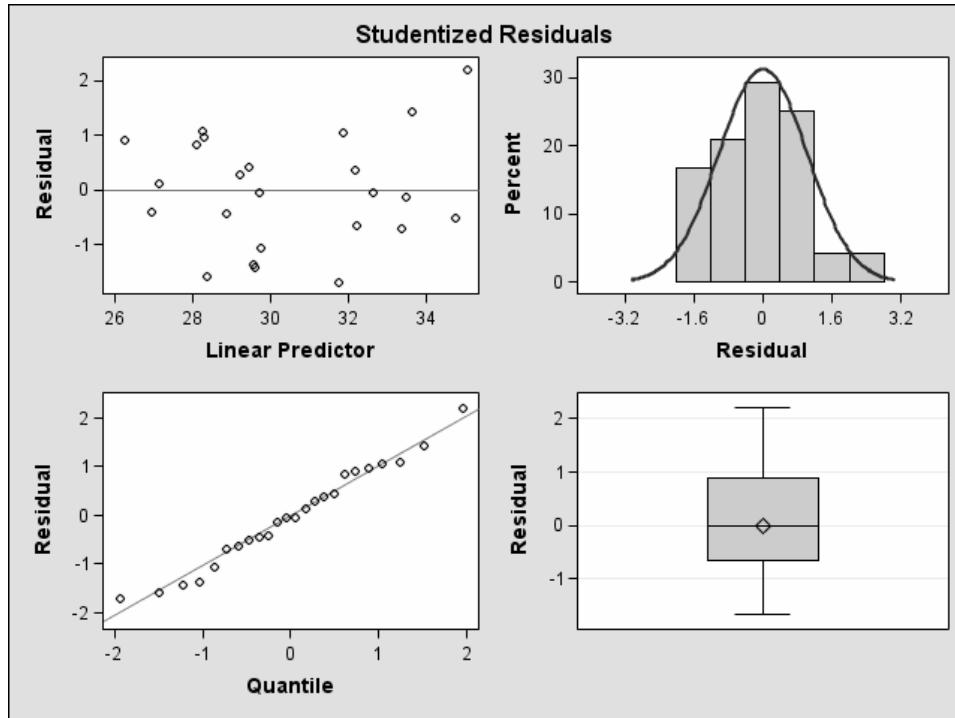


Figure 10.2 Conditional Studentized Residuals in Split-Plot Analysis

Interpretation

In the upper-left corner of the panel display you can see a scatter plot of the residual versus the linear predictor. The linear predictor is simply $\mathbf{x}_i'\hat{\beta}$ for a marginal residual (Figure 10.1) and $\mathbf{x}_i'\hat{\beta} + \mathbf{z}_i'\hat{\mathbf{u}}$ for a conditional residual (Figure 10.2). Because the BLOCK and BLOCK \times CULT effects are random effects, there are only six distinct values for the linear predictor in Figure 10.1, corresponding to the six combinations of the INOC and CULT factors. The marginal residuals do not contain direct information about the BLOCK and BLOCK \times CULT effects in this model. Their influence on the marginal predicted values is *indirect*, through the variance component estimates that determine the V matrix. In Figure 10.2, the BLOCK and BLOCK \times CULT effects affect the predicted values directly, through the BLUPs of the random effects.

The box plots in the lower-right corner of Figures 10.1 and 10.2 show that there are no outlying marginal or conditional residuals and that the residuals have a mean of zero. You can readily verify that the residuals in this model have a mean of zero with the following statements:

```

proc glimmix data=cultspd;
  class block cult inoc;
  model drywt = cult inoc cult*inoc / ddfm=satterth;
  random block block*cult;
  output out=gmxout student(blup) = resid_cond
                    student(noblup)= resid_marg;
run;
proc means data=gmxout min max mean;
  var resid_cond resid_marg;
run;

```

The OUTPUT statement saves the conditional and marginal residuals to a data set, and PROC MEANS computes the sample statistics (Output 10.2). The residuals do not sum to exactly 0.0 because of the limits of finite precision computing.

Output 10.2 Sample Statistics for Studentized Residuals

Variable	Label	Minimum	Maximum	Mean
resid_cond	Studentized Residual	-1.6812934	2.2118656	6.467049E-15
resid_marg	Studentized Residual (no blups)	-2.1040484	2.1412882	4.303965E-14

If you ignore the split-plot nature of the experiment for the moment and resort to a fixed effects model, you can fit the model as a two-way factorial.

Program

```

ods html;
ods graphics on;
proc glimmix data=cultspd plots=studentpanel(type=noblup);
  class block cult inoc;
  model drywt = block cult block*cult
                inoc cult*inoc / ddfm=satterth;
run;
ods graphics off;
ods html close;

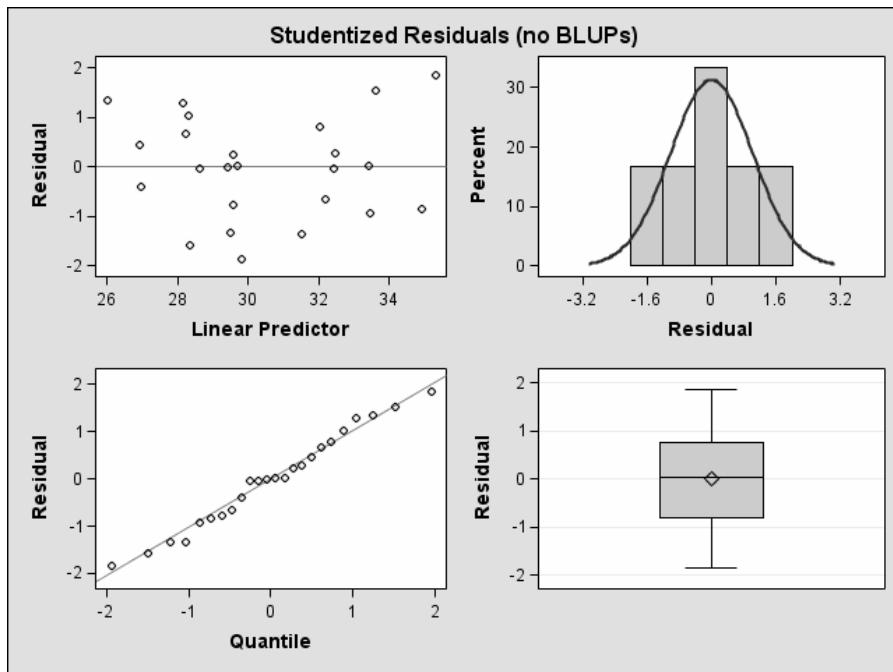
```

Results

The marginal studentized residuals—the only studentized residuals in this case—are shown in Figure 10.3.

Interpretation

The scatter plot and the residual histogram are quite different from the marginal mixed model residuals in Figure 10.1. The linear predictor is now sensitive to the levels of the BLOCK effect.

Figure 10.3 Studentized Residuals in Factorial Analysis

10.2.3.3 Generalized Least Squares and Influence Diagnostics

The modifications brought about by EGLS estimation are much more severe for influence diagnostics than for residual diagnostics. When your model contains random effects or the **R** matrix is not just $\sigma^2\mathbf{I}$, the estimate of β depends on the estimates of the covariance parameters. An update formula to compute $\hat{\beta}_{-i}$ from $\hat{\beta}$ is then only available if the covariance parameters are known (Cook and Weisberg 1982), in which case they would not change with the removal or addition of points. The practical consequence is that you can avoid refitting the model following the removal of observations only if you are willing to assume that the removal of an observation does not alter the covariance parameters.

A further complication of the linkage between fixed effects solutions and covariance parameter estimates is that it takes considerably more work to determine how the influence of an observation manifests itself. For example, consider a repeated measures study in which the correlations over time are modeled with an autoregressive structure (AR(1)). An observation that has small leverage in the regressor space would not exert much influence on the fixed effects in a standard model for independent data. But in the repeated measures model the observation can have considerable **indirect influence**. If it affects the AR(1) correlation parameter, then it can affect $\hat{\beta}$ because it alters $\hat{\mathbf{V}}$. But the reverse is also possible. Even if an observation affects the correlation parameter considerably, it may exert little influence on the fixed effects. For the purpose of prediction this influence is inconsequential, but for the accuracy of prediction intervals the influence can be critical.

Finally, an important difference between influence diagnostics in classical linear models and in mixed models is the clustered nature of the data. In many applications, notably longitudinal data and repeated measures, the data consist of independent groups or clusters, the SUBJECTs in your analysis. You are then more interested in assessing the influence of entire subjects than the influence of individual observations. The questions of how a particular patient affects the results

in a clinical trial or the impact of a center in a multi-center study arise naturally. This calls for **set-deletion** diagnostics, the removal of more than one observation at a time.

10.3 The Influence Diagnostics

The procedure we recommend for perturbation analysis in mixed models has a “drill-down” character. Because influence can manifest itself in so many ways, it is best to start with a summary measure of the overall influence for the observation or sets of observations of interest. This usually points out the observations that are particularly influential on *some* aspect of the analysis. If no observations are identified that would warrant further investigation at this stage, the process stops. Otherwise, we ask more specific questions about the nature of the influence. This can be done by quantifying impact on **vectors** of parameter estimates and impact on the precision of those vectors. For example, we can inquire about the overall change in the vector of fixed effects or in the vector of covariance parameters and their precision. Finally, a plot of the parameter estimates after deletion compared to the full-data estimates is often insightful.

10.3.1 Overall Influence

An overall influence statistic measures the global impact on the model, by quantifying a change in the objective function. In linear mixed models the objective function is tied to the maximum likelihood and the residual maximum likelihood in ML and REML estimation, respectively. An overall influence measure is thus the **likelihood distance** of Cook and Weisberg (1982), also termed the **likelihood displacement** by Beckman, Nachtsheim, and Cook (1987). The basic concept is to do the following:

1. remove an observation or group of observations
2. compute the parameter estimates for the reduced data set
3. assess the height of the original (restricted) likelihood surface at the delete-data parameter estimates

Notice that distance or displacement measure is not the difference of the log likelihoods or restricted log likelihoods in two models, one with n , the other with $n-d$ data points (where d denotes the number of observations deleted). It is the distance between the (restricted) log likelihoods based on all n observations, computed at two points in the parameter space. You *cannot* calculate this measure by running PROC MIXED twice, once with the full data and once with the reduced data set, and then taking differences of the log likelihoods.

10.3.2 Influence on the Parameter Estimates

A common way to measure the impact on a vector of parameter estimates is to compute a quadratic form in the difference between the full-data and reduced-data estimates. Cook’s D statistic is of this nature. In general it can be expressed as

$$D(\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_U)' \text{Var}[\hat{\boldsymbol{\beta}}]^{-1} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_U) / \text{rank}(\mathbf{X})$$

where the subscript U denotes the estimates after deleting observations in the set U (to allow for multiple point deletion). A closely related statistic is the multivariate DFFITS statistic of Belsley, Kuh, and Welsch (1980, p. 32):

$$MDFFITS(\beta) = (\hat{\beta} - \hat{\beta}_U)' \text{Var}[\hat{\beta}_U]^{-1} (\hat{\beta} - \hat{\beta}_U) / \text{rank}(\mathbf{X})$$

The primary difference between the two statistics is that the variance matrix of the fixed effects is based on an externalized estimate that does not involve the observations in the set U .

The idea of D and $MDFFITS$ can be applied to the covariance parameters as well. For either statistic, influence increases with the magnitude of the statistics.

10.3.3 Influence on the Precision of Estimates

D and $MDFFITS$ are quadratic forms in the difference of the full-data and reduced-data parameter estimates. To contrast the change in the precision, we need to engage not the vectors themselves, but their covariance matrices. The two common ways to do this are through trace and determinant operations. This leads to the COVTRACE and COVRATIO statistics.

$$\text{COVTRACE}(\beta) = \left| \text{trace} \left(\text{Var}[\hat{\beta}]^{-1} \text{Var}[\hat{\beta}_U] \right) - \text{rank}(\mathbf{X}) \right|$$

$$\text{COVRATIO}(\beta) = \frac{\left| \text{Var}[\hat{\beta}_U] \right|}{\left| \text{Var}[\hat{\beta}] \right|}$$

The reasoning behind the COVTRACE statistic is as follows. If \mathbf{A} is a positive semi-definite matrix, then $\text{trace}(\mathbf{A}^{-1}\mathbf{A})$ equals the rank of \mathbf{A} . If the variance of parameter estimates is not affected by the removal of observations, then the trace in the COVTRACE statistic should equal the rank of \mathbf{X} . The yardstick for lack of influence is thus a value of 0. As the value of the COVTRACE statistic increases, so does the influence on the precision of the parameter estimates and linear functions thereof, such as tests of fixed effects.

The COVRATIO statistic relates the determinants of the covariance matrices of the full-data and reduced-data estimates. The yardstick of no influence is a value of 1.0. Values larger than 1.0 indicate increased precision in the full-data case, and values smaller than 1.0 indicate higher precision for the reduced-data estimates. Such an interpretation in terms of increase or decrease in precision is not possible with the COVTRACE statistic. The disadvantage of the COVRATIO statistic is primarily computational. When the \mathbf{X} matrix is of less than full rank, special manipulations are required to compute the determinants of the non-singular components of the variance matrices.

As for the Cook's D and $MDFFITS$ statistics, the covariance trace and ratios can be computed for the fixed effects or the covariance parameters. Obtaining deletion statistics for the covariance parameters is not possible, however, unless the covariance parameters are updated as well. This requires refitting the linear mixed model, also known as **iterative** influence analysis.

10.3.4 Iterative or Non-iterative Analysis

A **non-iterative influence analysis** relies on the full-data estimates and update formulas to compute estimates for the model in the reduced data set and to compute influence statistics from the delete-data estimates. As mentioned earlier, this is possible in the linear mixed model only if the covariance parameters are known. In practice, this implies holding the covariance parameters fixed at the full-data estimates, unless a closed-form update is possible. You can obtain such a closed-form update of the profiled residual variance, but not of other covariance parameters. This has an effect on the values of the influence statistics. A likelihood distance in non-iterative analysis, for example, reflects only changes in the estimates of β and σ^2 . By holding other covariance parameters fixed, the estimate $\hat{\beta}_v$ tends not to move as far from the full-data estimate compared to an analysis in which all parameters are allowed to vary. If the likelihood surface is unimodal, the resultant likelihood distance is smaller. In other words, by not allowing all covariance parameters to be updated, one tends to underestimate the full importance of an observation for the analysis.

On the other hand, a non-iterative analysis consumes considerably fewer computing resources. In general, however, **iterative influence analysis**, which involves refitting of the model for the reduced data set, is recommended. It provides influence statistics for the covariance parameters and measures the full importance of observations on all aspects of the analysis. It can be computationally expensive, however. Consider, for example, a longitudinal study with N subjects where deletion sets correspond to the observations for a particular subject. An iterative influence analysis involves the fitting of $N+1$ mixed models, each an iterative process. The initial model is fit to the full data, followed by N fits, each of which burdens the observations with the corresponding subject removed. Fortunately, the computational burden can be reduced in some ways:

- You can limit the number of iterations for each deletion set. Allowing for only a single iteration leads to so-called **one-step** updates, which are arguably better than updating only the fixed effects. Limiting the number of iterations has the further advantage of yielding additional qualitative information about the deletion points or sets. If, for example, you allow for up to five additional iterations, and some sets do not converge within that limit, the influence measures calculated from the delete estimates at the fifth iteration are probably underestimates. The parameter estimates were still changing.
- In contrast to the initial fit of the model, you have good starting values for fitting the model to the delete-data: the full-data parameter estimates. If the observations removed are *not* influential, then convergence will be obtained quickly, typically in one or two steps. This is true for most of the observations.

The experimental INFLUENCE option in the MIXED procedure of SAS 9.1 implements both iterative and non-iterative influence analyses. The non-iterative analysis is the default. The following section presents an example.

10.4 Example: Unequally Spaced Repeated Measures

We return in this example to Data Set 5.4, “Heart Rates.” Repeated measurements on the heart rates of patients were taken at five unequally spaced repeated time intervals: at 1 minute, 5 minutes, 15 minutes, 30 minutes, and 1 hour. Each patient is subjected to one of three possible drug treatment levels, a , b , and p . The treatment level p represents a placebo.

We first fit these data with a repeated measures model that adjusts the mean for treatment and baseline effects. A model that includes temporal effects in the mean structure is considered in Section 10.4.2.

10.4.1 No Temporal Effects in Mean Structure

Program

```
ods html;
ods graphics on;
proc mixed data=hr order=data;
  class drug hours patient;
  model hr = drug basehr / noint s residual
    influence(iter=5 effect=patient est);
  repeated hours / type=sp(exp)(hours) sub=patient;
run;
ods graphics off;
ods html close;
```

The RESIDUAL option produces statistical graphics of various mixed model residuals if the ODS GRAPHICS is ON and an output destination has been chosen that supports graphical displays (e.g., with ODS HTML). The INFLUENCE option in the MODEL statement requests influence diagnostics. When ODS GRAPHICS are ON, this produces graphical displays of influence diagnostics in addition to the tabular output in the listing. The ITER=5 sub-option of the INFLUENCE option specifies that the influence analysis is iterative with a maximum of five iterations per observation or deletion set. The EFFECT=PATIENT sub-option specifies that observations that share the same level of the PATIENT variable comprise a deletion set. PROC MIXED computes influence diagnostics at the level of the patient, rather than for the individual observation. The EST sub-option requests that graphics of the delete estimates are produced in addition to the graphics of the influence diagnostics.

Results

The results of the influence analysis are reported in a table with one row per deletion set. The table is reproduced in Output 10.3.

Output 10.3 Influence Diagnostics for Repeated Measures Model

Influence Diagnostics for Levels of patient							
patient	Number of Observations in Level	Iterations	PRESS Statistic	Cook's D	MDFFITS	COVRATIO	COVTRACE
201	5	2	280.23	0.00692	0.00578	1.3067	0.2860
202	5	2	408.45	0.15659	0.09979	1.5690	0.5691
203	5	1	263.81	0.03905	0.03374	1.2008	0.1944
204	5	2	67.20	0.00058	0.00047	1.4555	0.4041
205	5	2	477.33	0.00380	0.00328	1.1951	0.1898
206	5	1	103.38	0.01702	0.01317	1.4143	0.3840
207	5	2	854.76	0.00037	0.00032	0.9498	0.0253
208	5	3	775.13	0.21872	0.17803	0.9822	0.0129

Influence Diagnostics for Levels of patient							
patient	Number of Observations in Level	Iterations	PRESS Statistic	Cook's D	MDFFITS	COVRATIO	COVTRACE
209	5	3	790.12	0.08113	0.06889	1.2268	0.2189
210	5	2	350.21	0.05812	0.04994	1.1412	0.1443
211	5	1	170.31	0.03380	0.02885	1.2350	0.2248
212	5	2	77.05	0.00004	0.00004	1.6180	0.5352
214	5	1	989.60	0.07652	0.06918	0.9926	0.0002
215	5	1	201.56	0.02360	0.01747	1.4343	0.4113
216	5	3	1421.35	0.25594	0.26536	0.5132	0.6045
217	5	2	119.56	0.00902	0.00732	1.4729	0.4162
218	5	2	459.17	0.07585	0.06576	1.1018	0.1080
219	5	2	1335.53	0.19162	0.17917	0.7750	0.2359
220	5	1	88.39	0.01266	0.01059	1.3395	0.3114
221	5	2	45.21	0.00323	0.00252	1.5581	0.4840
222	5	1	558.05	0.04327	0.03735	1.1806	0.1774
223	5	2	174.31	0.00049	0.00040	1.4622	0.4069
224	5	1	821.01	0.11293	0.10519	0.8810	0.1178
232	5	2	105.62	0.00571	0.00477	1.3598	0.3271

Influence Diagnostics for Levels of patient						
patient	Cook's D CovParms	MDFFITS CovParms	COVRATIO CovParms	COVTRACE CovParms	RMSE without deleted level	Restricted Likelihood Distance
201	0.01886	0.01660	1.2220	0.2118	9.13972	0.0437
202	0.07213	0.07021	1.0627	0.0648	9.14579	0.6946
203	0.04527	0.04256	1.0997	0.0995	9.16496	0.1999
204	0.05584	0.04822	1.2697	0.2545	9.26450	0.0548
205	0.01437	0.01316	1.1696	0.1643	9.07817	0.0283
206	0.05781	0.05212	1.1600	0.1562	9.21659	0.1216
207	0.12247	0.13595	0.8630	0.1406	8.80300	0.1359
208	0.11759	0.13323	0.8638	0.1386	8.92787	1.0894
209	0.84822	0.74545	1.2360	0.2571	8.89571	1.1744
210	0.04671	0.04592	1.0621	0.0628	9.11372	0.2751
211	0.04838	0.04486	1.1255	0.1238	9.18051	0.1785
212	0.05521	0.04715	1.3121	0.2915	9.26857	0.0514
214	0.04301	0.04412	1.0313	0.0329	8.93370	0.3615
215	0.01272	0.01169	1.1481	0.1434	9.15591	0.1055
216	0.49491	0.66349	0.6157	0.4287	8.49945	1.9113
217	0.04807	0.04119	1.3065	0.2866	9.24438	0.0766

Influence Diagnostics for Levels of patient						
patient	Cook's D CovParms	MDFFITS CovParms	COVRATIO CovParms	COVTRACE CovParms	RMSE without deleted level	Restricted Likelihood Distance
218	0.02239	0.02256	1.0251	0.0260	9.06948	0.3275
219	0.16452	0.19196	0.7866	0.2248	8.73670	1.0575
220	0.09604	0.08523	1.1778	0.1741	9.25525	0.1401
221	0.05713	0.04866	1.3203	0.2987	9.26869	0.0657
222	0.00701	0.00650	1.1533	0.1487	9.07493	0.1766
223	0.05379	0.04539	1.3248	0.3030	9.23268	0.0509
224	0.23626	0.25793	0.9762	0.0145	8.78276	0.7404
232	0.04977	0.04378	1.2153	0.2061	9.23950	0.0686

Interpretation

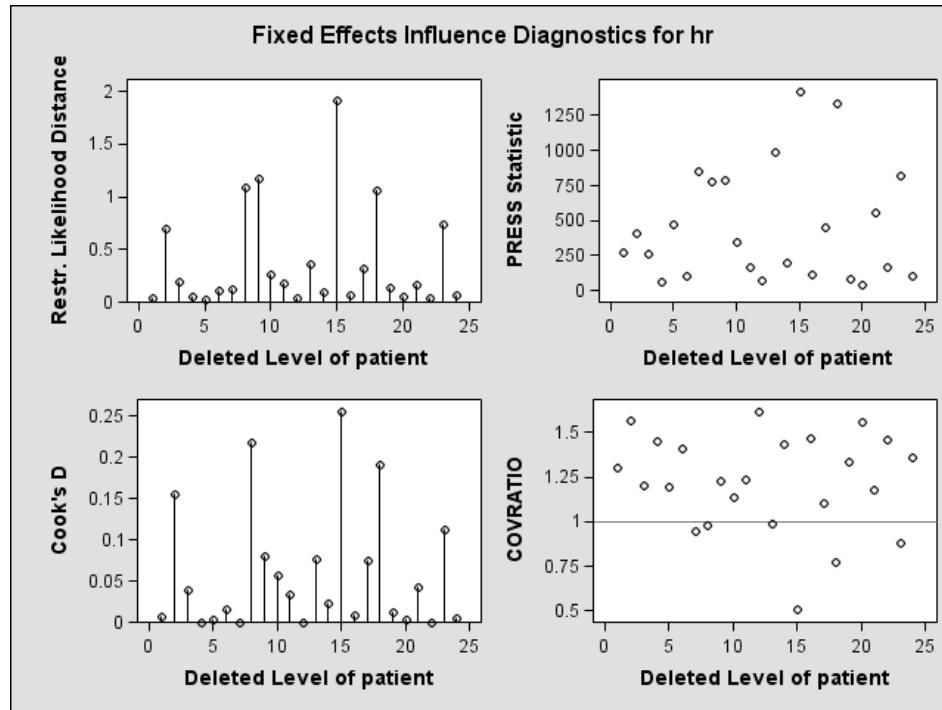
The listing results of the influence analysis are presented here in two tables. The first contains information about the number of iterations, and statistics related to the fixed effects. The second table contains information about the covariance parameters and the overall influence (residual likelihood ratio distance).

The first key result is that the number of iterations required to obtain convergence in the reduced data models is less than five in all cases. For most patients, convergence was achieved in a single step or with only two updates.

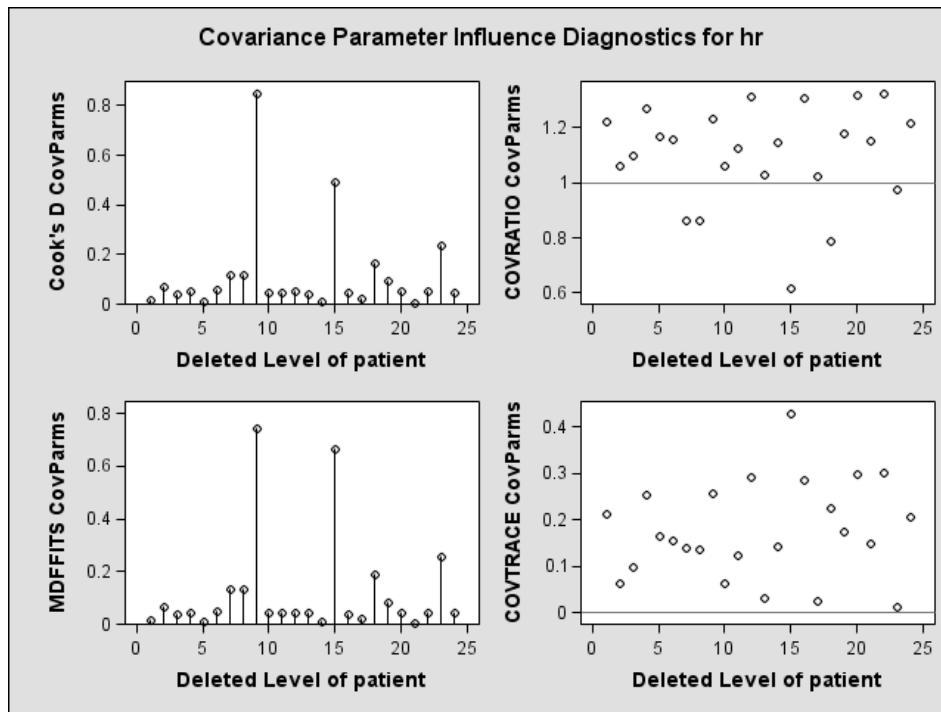
We have printed in bold the observations in Output 10.3 that appear worthy of a closer look. The last column in Output 10.3 suggests that the overall influence of patients 202, 208, 209, 216, 219, and 224 stands out compared to that of the rest of the patients. In order to determine what model components these study subjects influence, we examine the individual statistics. For example, the covariance parameter deletion diagnostics for patient 202 do not appear large compared to other patients. The Cook's D statistic is larger than that of most patients, however. If patient 202 is eliminated from the study, the primary change in model output is in the fixed effects estimates.

Patient 209, on the other hand, exhibits a small Cook's D and *MDFFITS* statistic, and his or her influence is primarily on the point estimates of the covariance parameters, as seen from a large statistic for "Cook's D CovParms" (0.84822). Finally, some patients exert influence throughout the model (for example, patient 216). Removing the data from this individual changes greatly the fixed effects solutions as well as the covariance parameter estimates (Cook's D is large in both cases). Furthermore, the data from this individual exert considerable influence on the precision of estimates. The COVRATIO statistic is considerably less than 1.0 for the fixed effects and the covariance parameters (0.5132, 0.6157). The presence of this individual's data reduces the precision of the estimates. Removal increases precision.

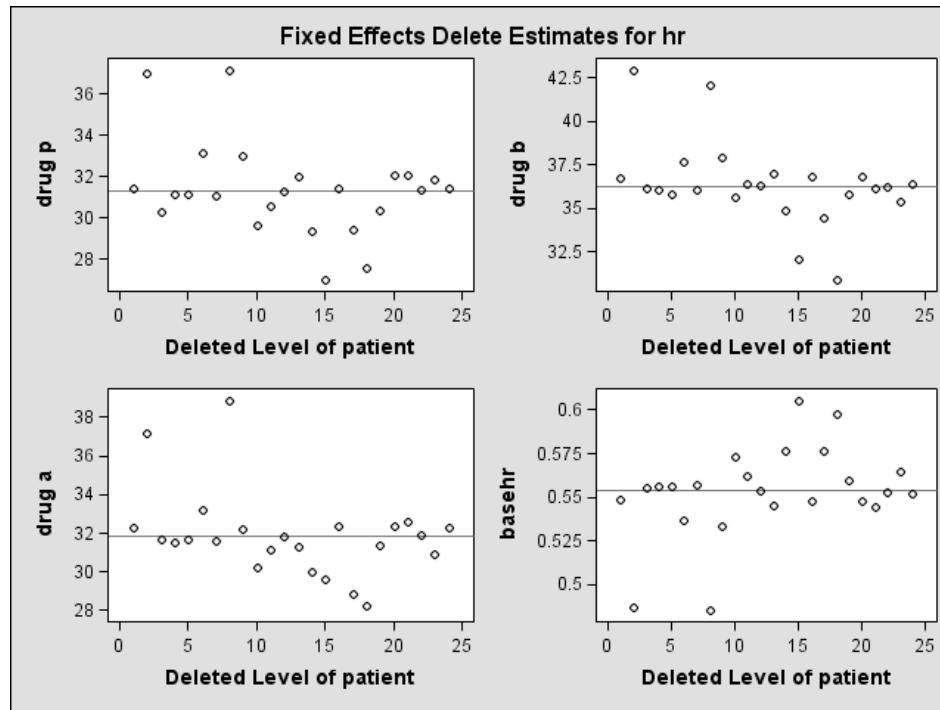
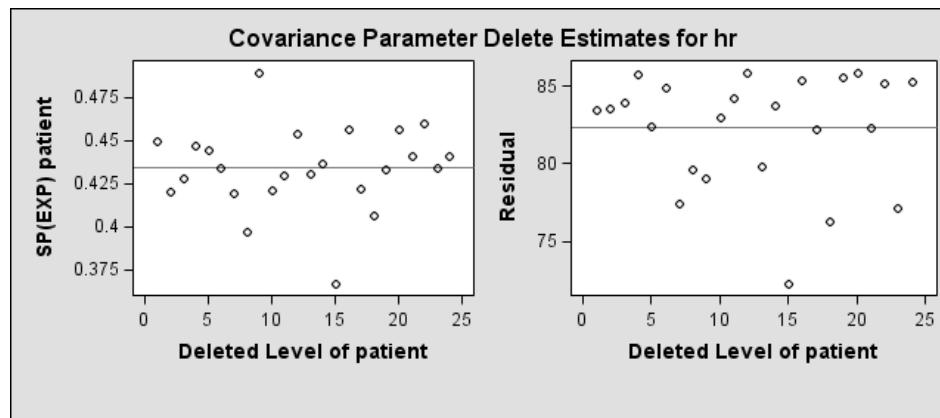
It can be tedious to go through a large table of influence diagnostics to examine in detail which observation or set of observations is influential and how. Graphical displays are better suited for this task. Figures 10.4 and 10.5 display the deletion statistics for each patient. Figures 10.6 and 10.7 compare the individual parameter estimates obtained after deleting a patient's data from the full-data estimates.

Figure 10.4 Overall Influence and Influence on Fixed Effects

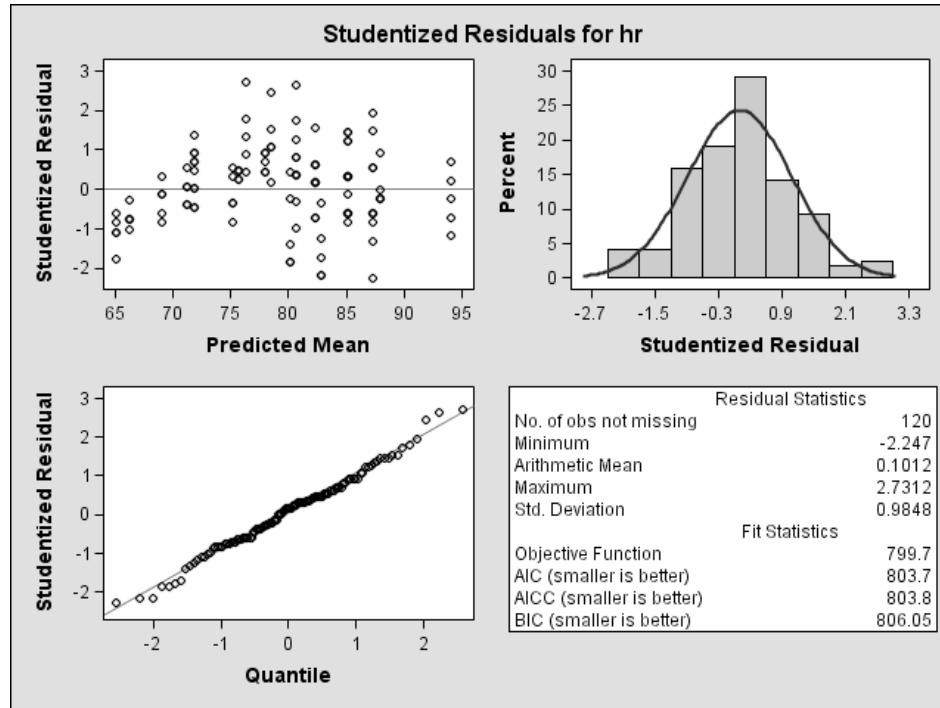
The upper-left corner of Figure 10.4 shows a needle plot of the REML likelihood distance. The patients identified in Output 10.3 clearly stand out. The plot in the lower-left corner gives Cook's D for the fixed effects. The same patients stand out, with the exception of the ninth level of the patient effect (patient 209). The COVRATIO plot in the lower-right corner clearly identifies the patients whose presence in the study reduces the precision of the fixed effects (primarily levels 15 and 18, corresponding to patients 216 and 219).

Figure 10.5 Influence on Covariance Parameter Estimates

The graphs on the left side of Figure 10.5 provide information about the effect on the covariance parameter estimates; the graphs on the right side provide information about their precision. Note how level 9 (patient 209) stands out as being influential on the covariance parameter estimates. The impact of this patient's data on the fixed effects is minor (Figure 10.4). If the goal of the analysis is to predict the heart rate of patients, the data from patient 209 is inconsequential. However, if the analytic goal depends on quantities involving the covariance parameter estimates, the five data points associated with patient 209 are important. Their presence in the data reduces the estimate of the SP(EXP) parameter (Figure 10.7), essentially reducing the degree of correlation among data points from any patient. Observations from level 15 (patient 216), on the other hand, have the opposite effect. The temporal correlation drops when the data points are removed.

Figure 10.6 Comparison of Full-Data and Delete-Data Fixed Effect Estimates**Figure 10.7** Comparison of Full-Data and Delete-Data Covariance Parameter Estimates

Based on this influence analysis, you can reconsider the model and/or data. For example, you could eliminate certain patients from the data and base the final analysis on only the remainder. But this is usually not the right action to take. As mentioned earlier, the results of a perturbation analysis can be seen only in light of the model you work with. Observations may appear as unusual because the model is wrong. The plot of studentized residuals for this model in Figure 10.8 certainly looks suspicious. There is a strong trend in the residuals. An important effect may be missing from the model.

Figure 10.8 Studentized Residuals

10.4.2 With Temporal Effects in Mean Structure

The shortcoming of the previous model is that it does not include systematic effects of time. The following program allows for HOURS effects and their interaction with the treatments. Graphical results are presented in Figures 10.9 and 10.11 for comparison with those in the earlier section.

Program

```

ods html;
ods graphics on;
proc mixed data=hr order=data;
  class drug hours patient;
  model hr = drug hours drug*hours basehr / s
    residual
    influence(iter=5 effect=patient);
  repeated hours / type=sp(exp)(hours) sub=patient;
run;
ods graphics off;
ods html close;

```

Program

Figure 10.9 Studentized Residuals

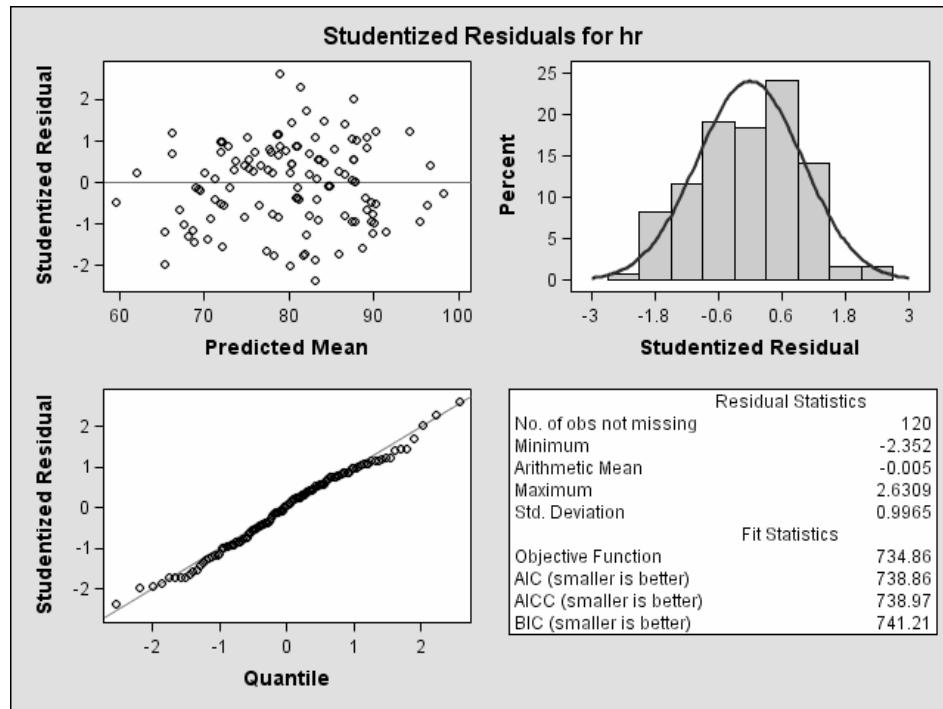


Figure 10.10 Overall and Fixed Effects Influence

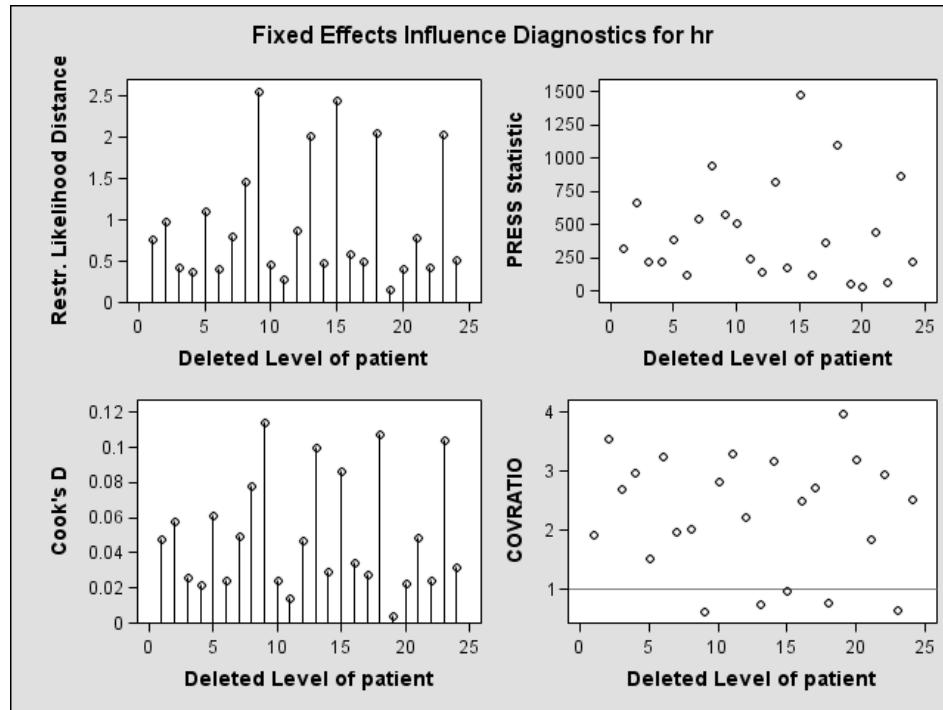
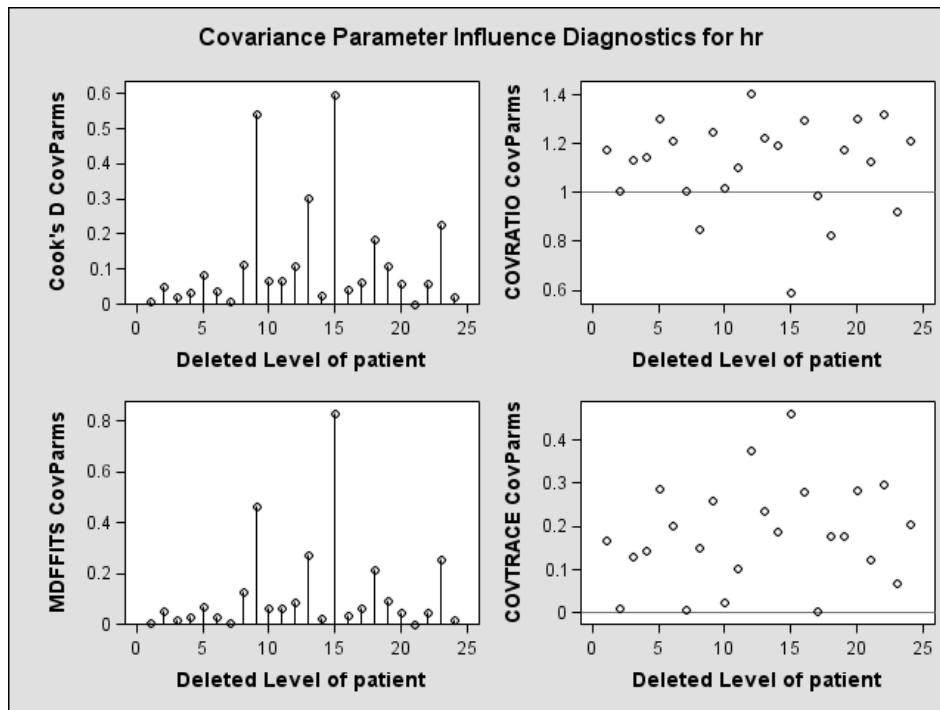


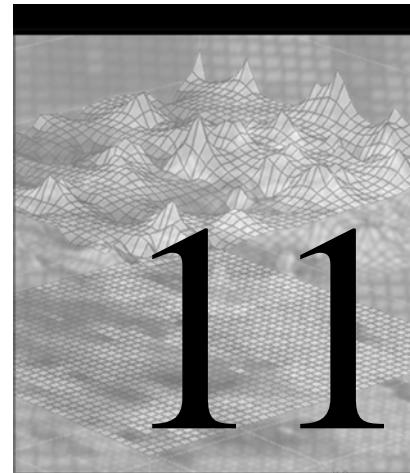
Figure 10.11 Influence on Covariance Parameters

Interpretation

The studentized residuals show much less trend than before (Figure 10.9); also notice that their mean is now much closer to zero. The likelihood distance and Cook's D statistics for the fixed effects are much more evenly distributed compared to the analysis without temporal adjustment in Section 10.4.1. The “sore thumbs” that continue to stick out are levels 9 and 15 of the PATIENT effect (corresponding to patients 209 and 216). The residuals for patient 209 have a “zigzag” appearance over time, which runs counter to the notion of positive serial correlation. The residuals for patient 216 are most extreme in negative value (see PRESS Statistic in Figure 10.10).

10.5 Summary

This chapter discussed residual and influence analysis in mixed models. Section 10.2 made the transition from diagnostic work in standard linear models without random effects or correlated errors to mixed models. Section 10.3 introduced the diagnostics available for mixed models in SAS 9.1. Section 10.4 analyzed a repeated measures study.



Spatial Variability

11.1	Introduction	438
11.2	Description	438
11.3	Spatial Correlation Models	440
11.3.1	Spatial Correlation Models Used in PROC MIXED	440
11.3.2	Models with a Nugget Effect	442
11.4	Spatial Variability and Mixed Models.....	442
11.4.1	Integrating Spatial Variability into Mixed Models	442
11.4.2	Geostatistics Related to PROC MIXED.....	444
11.5	Example: Estimating Spatial Covariance	447
11.5.1	Estimating the No-Nugget Model.....	447
11.5.2	Likelihood Ratio Test of Spatial Covariance.....	450
11.5.3	Comparing Spatial Covariance Models	450
11.5.4	Comparing Spatial Models to Nonspatial Models.....	451
11.5.5	Estimating the Model with a Nugget Effect	453
11.6	Using Spatial Covariance for Adjustment: Part 1, Regression	457
11.6.1	A Regression Example.....	458
11.6.2	Comparison with Independent Errors Model.....	459
11.7	Using Spatial Covariance for Adjustment: Part 2, Analysis of Variance	460
11.7.1	Possible Models for Wheat Yield Trial	460
11.7.2	Analysis Using REML Estimation of Covariance Components	462
11.7.3	Alternative Estimate of Covariance Parameters Based on Semivariogram	466
11.7.4	Impact of Spatial Adjustment on Treatment Effect Inference.....	468
11.8	Example: Spatial Prediction—Kriging	471
11.8.1	Ordinary Kriging	471
11.8.2	Universal Kriging	476
11.9	Summary.....	478

11.1 Introduction

Spatial variability, as the name implies, refers to variation among observations in space. This variation can result from changes in the mean of the attribute of interest or from stochastic sources. Typically, changes in mean value are associated with more large-scale sources of variation, and stochastic components are assumed to contribute to variation at the small- and micro-scale level. One of the goals of spatial statistics is to model and estimate the patterns that manifest themselves in spatial variability through these sources. There are two main sets of methods to study spatial data: (1) those applicable to random locations, such as trees in a forest whose locations are random occurrences, and (2) those appropriate for fixed locations, such as a field study using a sampling design or a designed experiment. Spatial data with fixed locations are commonly further divided into geostatistical and lattice data. In the former, the spatial domain where observations are collected is continuous. In the latter, the domain consists of discrete units. An example of geostatistical data is locations at which a company drills for oil. Whether the drilling sites are selected at random or systematically has no bearing on the nature of the data. An example of lattice data is the analysis of U.S. Census data at the county level; the counties of a state are discrete areal units. Designed field studies or experiments also give rise to lattice data; each experimental field unit represents one cell of the lattice.

PROC MIXED has several features that enable you to work with spatial data with fixed locations. The procedure requires that observations be uniquely identified with a spatial coordinate. This is always the case for geostatistical data. In the case of lattice data, you can select a representative coordinate to represent the areal unit, such as the seat of the county government or the center of an experimental unit. However, many of the spatial covariance models provided by PROC MIXED require that the spatial (error) process that generated the data satisfy certain stationarity conditions. For example, the models usually assume equal variance among the observations. Lattice data with irregularly shaped areal units are typically nonstationary. In summary, PROC MIXED enables you to model geostatistical data; you can model lattice data to the extent that the assumptions of the models are met by data collected on discrete units. Lattice data from field experiments with homogeneously shaped experimental units usually fall into this category. Models specifically designed for lattice data, such as conditional or simultaneous autoregressive models, are based on neighborhood structures. These models cannot be fit with the MIXED procedure.

11.2 Description

In a designed field study or experiment, small-scale dependence—that is, the relationship among observations in proximity to each other—is often of particular interest. Small-scale dependence can be positive or negative. PROC MIXED has several features that enable you to work with *positive* dependence.

Positive, small-scale spatial dependence, or **positive spatial correlation**, refers to the tendency of observations to exhibit positive association. That is, relatively high (or low) values tend to be surrounded by other relatively high (or low) values. With spatial data, correlation is typically positive. Negative correlations occur, for example, with cyclical or seasonal data when the overall trend is not properly removed. The covariance structures for spatial data in PROC MIXED assume that the spatial dependence is positive. The tendency for observations to be positively correlated is one important feature of spatial data. In addition, the correlations often

are a function of the distance between observations. The correlations decline with increasing spatial separation. For example, in a field survey to assess ground water contamination, several wells are drilled in the survey area and water quality measurements taken at each well.

Typically, observations taken at wells in relatively close proximity are more highly correlated than those taken at more distant wells. In agronomic and horticultural experiments, adjacent experimental units have common fertility characteristics, whereas more distant experimental units tend to be less alike.

As the number of treatments increases, the potential importance of accounting for spatial variability in field experiments also increases. Cultivar evaluation trials, for example, often involve dozens or even hundreds of varieties. In many experiments, the most important objective is to find out if new treatments or varieties perform satisfactorily in marginal environments. In order to make realistic conclusions, these experiments must be conducted in fields with difficult growing conditions. Such fields typically exhibit a high degree of spatial heterogeneity. You can base statistical analysis on the linear model that corresponds to a particular experimental design. For example, a randomized block design is analyzed based on a model with block and treatment effects and uncorrelated, homoscedastic errors. The original motivation for blocking was to reduce spatial variability. If the variability occurs at a scale larger or smaller than the size of the blocks, blocking is not an efficient method. As a result, spatial methods in which the mean and covariance structure of experimental data are modeled have been shown to lead to more efficient analyses (Zimmerman and Harville 1991, Brownie, Bowman, and Burton 1993, Stroup, Baenziger, and Mulinze 1994, Brownie and Gumpertz 1997). Failure to account for spatial dependence can then result in erroneous conclusions.

A good starting point for understanding spatial variability is to think of it as a multidimensional extension of repeated measures. In repeated measures, observations taken over time are correlated. In spatial statistics, observations are correlated in two spatial dimensions. As with repeated measures, most spatial procedures assume that the **errors**—i.e., the elements of \mathbf{e} in the mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$ —are correlated. Spatial correlation can be reflected in \mathbf{G} , the covariance matrix of the random model effects, \mathbf{u} , and in \mathbf{R} , the covariance matrix of the model errors.

Statistical methods for spatial correlation can be divided into two basic groups, **characterization** and **adjustment**. Characterization involves estimating covariance parameters. Adjustment involves removing the effects of spatial correlation to obtain more accurate estimates of, for example, treatment means or differences. The capabilities of PROC MIXED come into play for both methods, but the procedure is especially suited for adjustment. A further important step in the analysis of geostatistical data is the prediction of values at unobserved locations. The methods for prediction are usually summarized under the general heading of **kriging**, a term coined by G. Matheron in honor of the South African mining engineer D. G. Krige (Krige 1951, Matheron 1963).

PROC MIXED allows you to work with several different spatial correlation models. The purpose of this chapter is to demonstrate the use of PROC MIXED for characterization, adjustment, and spatial prediction. Section 11.3 describes basic models for spatial variability. Section 11.4 integrates spatial models into the mixed model framework and describes the basic features of inference. Section 11.5 presents a characterization example. Section 11.6 presents two adjustment examples, and Section 11.7 presents examples of ordinary and universal kriging with PROC MIXED. Section 11.8 contains a cautionary note about covariance estimation.

11.3 Spatial Correlation Models

Models for spatial correlation have their origins in statistical methods developed in geology for mining applications. Owing to this history, many important spatial models are called “geostatistical” models. Important references include Journel and Huijbregts (1978), Isaaks and Srivastava (1989), and Cressie (1993). While PROC MIXED uses geostatistical models, its approach differs from geostatistics, primarily for the sake of consistency with mixed model theory. Readers familiar with standard geostatistics will notice substantial differences in notation although the concepts are the same.

This section presents the basic spatial models used by PROC MIXED. Section 11.4 shows how these models are included in the mixed model. Also, Section 11.4 discusses key geostatistical concepts and their relationship to spatial mixed models.

11.3.1 Spatial Correlation Models Used in PROC MIXED

In the simplest spatial statistics problem, you have a set of observations whose physical location and response are known. Your primary objective is to estimate spatial correlation. The model is

$$Y_i = \mu + e_i$$

where Y_i is the i^{th} observation with mean μ , and e_i is the corresponding error. Let s_i denote the physical location of Y_i , where s_i is specified by two coordinates. For example, the coordinates could be latitude and longitude. Alternatively, the coordinates might be indices on a grid, such as north-south and east-west, or row and column dimensions, respectively. For simplicity, we refer to the elements of s_i as the coordinates. The specific meaning depends on the particular data set. For example, the coordinates may represent “row” and “column” position in a field experiment.

In general, you can define spatial correlation models by letting

$$\begin{aligned}\text{Var}[e_i] &= \sigma_i^2 \\ \text{Cov}[e_i, e_j] &= \sigma_{ij}\end{aligned}$$

Typically, the covariance is assumed to be a function of the distance between the locations s_i and s_j . If d_{ij} denotes the distance between s_i and s_j , the covariance models have the general form

$$\text{Cov}[e_i, e_j] = \sigma^2 [f(d_{ij})]$$

Models for which $f(d_{ij})$ is the same for all pairs of equally distant locations in a given direction, such as along the row, along the column, or diagonal, are called (second-order) **stationary models**. If, in addition, $f(d_{ij})$ does not depend on the direction, then the covariance structure is said to be **isotropic**. PROC MIXED can fit isotropic and various forms of anisotropic covariance models. The basic isotropic covariance models in PROC MIXED are the following:

1. Spherical

$$f(d_{ij}) = [1 - 1.5(d_{ij}/\rho) + 0.5(d_{ij}/\rho)^3] \times 1\{d_{ij} < \rho\}$$

2. Exponential

$$f(d_{ij}) = \exp(-d_{ij}/\rho)$$

3. “Gaussian”

$$f(d_{ij}) = \exp(-d_{ij}^2/\rho^2)$$

4. Linear

$$f(d_{ij}) = (1 - \rho d_{ij}) \times 1\{\rho d_{ij} < 1\}$$

5. Linear Log

$$f(d_{ij}) = [1 - \rho \log(d_{ij})] \times 1\{\rho \log(d_{ij}) < 1\}$$

6. Power

$$f(d_{ij}) = \rho^{d_{ij}}$$

7. Matérn

$$f(d_{ij}) = (d_{ij}/2\nu)^{\nu} 2K_{\nu}(d_{ij}/\rho) / \Gamma(\nu)$$

The function $1\{d_{ij} < \rho\}$, used in the spherical model, is an indicator function that equals 1 when $d_{ij} < \rho$ and equals 0 otherwise. Similar $1\{\bullet\}$ functions used for the linear and linear log models equal 1 when the condition within the parentheses holds and equal 0 otherwise.

We used quotes to label the Gaussian covariance structure because it has nothing to do with the Gaussian (=Normal) distribution. The latter is of the greatest importance in applied and mathematical statistics. The opposite is true for the Gaussian covariance model. Its name stems from a resemblance of the normal probability density function and the spectral density function for a process with that covariance function. In most applications the Gaussian covariance model is unrealistic because it implies great continuity of the process.

The power function is a reparameterization of the exponential covariance model (2).

The Matérn covariance model enables you to estimate the smoothness of the spatial process through the additional parameter (ν). The function $K_{\nu}(\bullet)$ is the modified Bessel function of the second kind of real order. Because of the involvement of this Bessel function and the additional smoothness parameter, the Matérn model is computationally very intensive. For $\nu = 1/2$ it reduces to the exponential model. For $\nu = 1$, the model is known as Whittle’s model (Whittle 1954).

For readers familiar with geostatistics, the parameter σ^2 corresponds to the sill and ρ is related to the **range** of the process. The range of a second-order stationary spatial process is that distance at which observations are no longer correlated. For covariance models where the covariances reach zero only asymptotically (Gaussian, exponential, power, Matérn), the **practical range** is defined as the distance where the covariance is reduced to 5% of the sill (see Section 11.4.2). You should be careful in equating the parameter ρ in the parameterizations above with the range or practical range. Only for the spherical model does ρ equal the range. For example, in the exponential model the practical range is 3ρ ; in the Gaussian model it is $\rho\sqrt{3}$. In what follows we refer to the parameter ρ simply as the “range” parameter. You need to make the necessary conversion to the practical range based on the covariance model.

If a process is stationary, but the covariance function is not isotropic, then the covariances are direction dependent. The correlation contours are no longer spherical. For example, the range of a spatial process may be different in the direction of the prevailing wind and in the direction

perpendicular to it. A particular form of anisotropy is geometric anisotropy, where the correlation contours are elliptical. It can be corrected by rotating the coordinate system and by applying a compression or stretching of the coordinates in the direction of one of the main axes of the ellipse. PROC MIXED can model geometrically anisotropic variations of the exponential, spherical, and Gaussian covariance structure.

11.3.2 Models with a Nugget Effect

In some applications, the covariance models given above do not adequately account for abrupt changes over relatively small distances. You can model these cases by adding an additional parameter. The resulting covariance models have the general form

$$\begin{aligned}\text{Var}[e_i] &= \sigma^2 + \sigma_1^2 \\ \text{Cov}[e_i, e_j] &= \sigma^2 f(d_{ij})\end{aligned}$$

where the $f(d_{ij})$ can be any of the models given above. For these models, the parameters σ_1^2 , $\sigma^2 + \sigma_1^2$, and ρ correspond to the geostatistical parameters **nugget**, **sill**, and “**range**,” respectively. Using geostatistics terminology, models with $\text{Var}[e_i] = \sigma^2 + \sigma_1^2$ are called **models with a nugget effect**, whereas models with $\text{Var}[e_i] = \sigma^2$ are called **no-nugget** models (see Section 11.4.2 for additional details). In a nugget model, σ^2 is termed the **partial sill**.

11.4 Spatial Variability and Mixed Models

This section shows how the spatial variability models discussed in Section 11.3 are incorporated into mixed models. The underlying theory was discussed by Zimmerman and Harville (1991).

11.4.1 Integrating Spatial Variability into Mixed Models

You can apply the covariance models described above to mixed models in general. For example, in a field experiment with different treatments, a potentially useful model is

$$Y_{ij} = \mu + \tau_i + e_{ij}$$

where the covariance structure of the e_{ij} 's follows one of the covariance models given above.

Recall the form of the general mixed model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where the random effects **u** have covariance matrix **G** and the model errors have covariance matrix **R**. You can incorporate spatial dependency in this model through **G** or through **R**. The two approaches are called the **conditional** and the **marginal** methods, respectively, because a model that has random effects **u** (and hence **G**) has a conditional distribution, that of $\mathbf{Y}|\mathbf{u}$. Another terminology that is common is to refer to **G-side** random components if you model variation through **G**, and to **R-side** random components if dependency is modeled through **R**. Since (restricted) likelihood estimation in PROC MIXED proceeds by maximizing the marginal log likelihood of **Y**, the two approaches yield the same result, if the marginal distributions are the same.

First consider the case of G-side modeling and let $\mathbf{Z} = \mathbf{I}$, $\text{Var}[\mathbf{u}] = \sigma^2 \mathbf{F}$, and $\text{Var}[\mathbf{e}] = \sigma_e^2 \mathbf{I}$, where \mathbf{F} is an $N \times N$ matrix whose ij^{th} element is $f(d_{ij})$. The covariance matrix of the conditional distribution is $\text{Var}[\mathbf{Y}|\mathbf{u}] = \sigma^2 \mathbf{I}$, and that of the marginal distribution is $\text{Var}[\mathbf{Y}] = \sigma^2 \mathbf{F} + \sigma_e^2 \mathbf{I}$. The G-side formulation automatically leads to a model with a nugget effect. Now consider the case of a model with only R-side components ($\mathbf{G} = \mathbf{0}$). If we define $\text{Var}[\mathbf{e}] = \sigma^2 \mathbf{F} + \sigma_e^2 \mathbf{I}$, a model with a nugget effect, the same distribution results as in the G-side model, since $\text{Var}[\mathbf{e}] = \text{Var}[\mathbf{Y}]$ in a marginal model. However, we can also model the covariance structure of the error as $\text{Var}[\mathbf{e}] = \sigma^2 \mathbf{F}$, a no-nugget model, for which there is no G-side equivalent.

You can obtain the three types of models (G-side, R-side with nugget, R-side without nugget) with PROC MIXED syntax as follows. We use an exponential covariance model here and assume that the variables X and Y are the spatial coordinates.

Model with Nugget (G-side)

```
data MyData;
  set MyData;
  obs = _n_;
run;
proc mixed data=MyData;
  class obs;
  model response = ;
  random obs / type=sp(exp)(x y);
run;
```

The variable OBS is added to the input data set; it contains the observation numbers. It is used as a CLASS variable and as a random effect. This makes the \mathbf{Z} matrix the identity matrix, so that there are N random effects in the vector \mathbf{u} . The covariance structure among the random effects is the spatial exponential structure (TYPE=SP(EXP)). The Euclidean distance between two observations is determined by the values of the variables X and Y. The estimates of the covariance parameters reported by PROC MIXED in the “Covariance Parameter Estimates” table are the estimates of the partial sill, the range parameter, and the nugget effect, respectively.

Model without Nugget (R-side)

```
proc mixed data=MyData;
  model response = ;
  repeated / subject = intercept
             type      = sp(exp)(x y);
run;
```

The R-side covariance structure is specified with the REPEATED statement of PROC MIXED. In contrast to the G-side version, we now need to add a SUBJECT= option in order to instruct the procedure which sets of observations are to be correlated with the structure chosen with the TYPE= option. The INTERCEPT specification creates an effect that is constant across the observations in the data set. This implies that all observations belong to the same subject. The TYPE= option then specifies that the observations from this subject have an exponential covariance structure. This model has two covariance parameters. They are reported in the “Covariance Parameter Estimates” table as the range parameter and the sill, respectively.

Model with Nugget (R-side)

```
proc mixed data=MyData;
  model response = ;
  repeated / subject = intercept
    type      = sp(exp)(x y)
    local;
run;
```

To add a nugget effect in an R-side model, you specify the LOCAL option in the REPEATED statement. This adds the term $\sigma_l^2 \mathbf{I}$ to the covariance structure. The estimates of the covariance parameters reported by PROC MIXED in the “Covariance Parameter Estimates” table are the estimates of the partial sill, the range parameter, and the nugget effect, respectively.

As with other mixed models, PROC MIXED obtains estimates of the variance and covariance components of \mathbf{R} and \mathbf{G} using REML, and obtains estimates of $\boldsymbol{\beta}$ and predictions of \mathbf{u} from solutions to the mixed model equations.

Inference on estimates or contrasts involving predictable functions $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$ uses the same approach as has been described elsewhere in this book. That is, standard errors of $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'(\mathbf{u} - \hat{\mathbf{u}})$ are computed as $\mathbf{L}'\mathbf{C}\mathbf{L}$, where $\mathbf{L}' = [\mathbf{K}' \mathbf{M}']$ and \mathbf{C} is the generalized inverse of the left side of the mixed model equations using estimated \mathbf{G} and \mathbf{R} . If you choose DDFM=KENWARDROGER, the matrix \mathbf{C} is adjusted for the uncertainty in the covariance parameter estimates. Test statistics for hypotheses use the F -approximation computed from the Wald statistic divided by the rank of \mathbf{L} or adjusted tests based on the Kenward-Roger or Satterthwaite method.

While competing covariance models cannot be tested directly, you can compare their model-fitting criteria. In principle, the *best* model is the one whose AIC, AICC, BIC, and $-2(\text{Res})\log\text{Likelihood}$ criteria are smallest. You can use the likelihood ratio test to compare models if one is a subset of the other. For example, a model with no spatial correlation—that is, with $\mathbf{R} = \sigma^2 \mathbf{I}$ —can be compared to any of the spatial models with no nugget effect, whose $\mathbf{R} = \sigma^2 \mathbf{F}$, because \mathbf{F} reduces to \mathbf{I} if the spatial parameter $\rho = 0$. Thus, comparing the $-2(\text{Res})\log\text{Likelihood}$ for each model yields a likelihood ratio test for $H_0: \rho = 0$ with 1 degree of freedom. Similarly, you can compare a model with and without nugget via a likelihood ratio test, since the no-nugget model is nested within the nugget model ($H_0: \sigma_l^2 = 0$).

11.4.2 Geostatistics Related to PROC MIXED

Two important steps in the geostatistical method are (1) to describe and estimate the pattern of spatial variability in the observed data and (2) to use this information to predict the spatial process at unobserved locations. The standard tool for the description and estimation of spatial variability is the **semivariogram**. The methods for spatial prediction are summarized as kriging methods.

The semivariogram is a standard statistical measure of spatial variability as a function of the distance between observations. It is defined as one-half the variance of the difference between two observations a given distance apart. In geostatistics, h is standard notation for distance, also termed the **lag** between two observations.

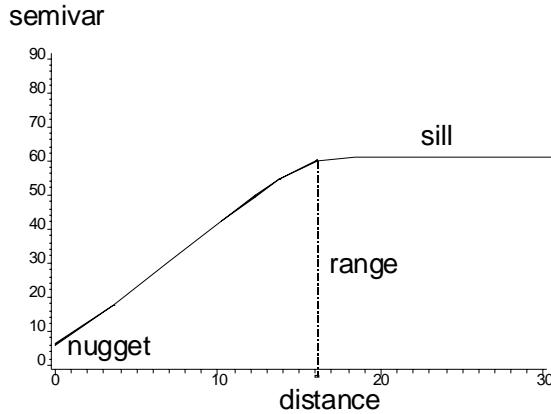
Figure 11.1 Semivariogram

Figure 11.1 is an idealized graph of the semivariogram. The key features of the semivariogram are as follows:

- the nugget, defined as the intercept, i.e., the semivariogram at $h = 0$. Standard geostatistical notation for the nugget is c_0 .
- the sill, defined as the value of the semivariogram at the plateau reached for larger h . The sill corresponds to the variance of an observation. Standard geostatistical notation for the sill is $C(0)$.
- the range, defined as the value of h at which the semivariogram reaches the sill. For distances less than the range, observations are spatially correlated. A semivariogram that stays low near the origin implies high short-range spatial autocorrelation. For distances greater than or equal to the range, spatial correlation is effectively zero. In models where the sill is only reached asymptotically (exponential, Gaussian, Matérn), the practical range is defined as the distance at which the semivariogram reaches 95% of the sill.

The **empirical semivariogram** is a plot of an estimate of the semivariogram against lag distance h . A large number of semivariogram estimators have been developed; see, for example, Chilès and Delfiner (1999) and Schabenberger and Gotway (2005). The **classical estimator** is due to Matheron (1963) and is computed from data using the formula

$$\gamma(h) = \frac{1}{2m} \sum (y_i - y_j)^2$$

where m is the number of pairs of observations a distance, h , apart. $\gamma(h)$ is estimated for all distances at which pairs of observations exist or at a discrete set of lag values within a tolerance to ensure that a sufficient number of observations contribute to each value of $\gamma(h)$.

The semivariogram has a functional relationship to the spatial covariance matrix in the mixed model of the previous subsection. Standard geostatistical notation for the covariance between two observations h units apart is $C(h)$. For no-nugget models,

$$C(h) = C(0) - \gamma(h)$$

$$\gamma(0) = 0$$

Also, the semivariogram can be written in terms of the sill as

$$\gamma(h) = C(0)[1 - f(h)]$$

where $f(\bullet)$ may be one of the functions of distance, such as spherical, given in Section 11.3.1. Thus, for these models, the sill, $C(0)$, corresponds to σ^2 as defined in Section 11.3.1. Also, since the semivariogram is equal to $C(0)[1-f(h)]$, then

$$C(h) = C(0)\{1 - [1 - f(h)]\} = \sigma^2 [f(h)]$$

The parameter ρ in the various functions $f(d_{ij})$ given in Section 11.3 corresponds to the range.

For models with a nugget effect, $C(0)$ is the variance of observations. The partial sill, denoted c_s , is the difference between the variance and the nugget—i.e., $c_s = C(0) - c_0$. The semivariance for such models is

$$\gamma(h) = c_s[1 - f(h)]$$

Thus, the sill corresponds to $\sigma^2 + \sigma_1^2$ and the **partial sill** corresponds to σ^2 . Relating these parameters to the covariance,

$$C(0) = \sigma^2 + \sigma_1^2$$

$$C(h) = \sigma^2 [f(h)], \text{ for } h > 0$$

After estimating the empirical semivariogram, the geostatistical method proceeds to estimate the parameters of a semivariogram model. This is often accomplished by fitting the semivariogram model to the empirical semivariogram by nonlinear least squares. Kriging (Krige 1951) is a method of predicting responses for unobserved locations based on information about the spatial dependence. Standard kriging methods are conceptually *best linear unbiased prediction* methods and can be written equivalently in terms of semivariances or covariances. In other words, you can solve the kriging problem regardless of whether estimates of the spatial variability are obtained by least-squares fitting of the empirical semivariogram or as REML estimates in a mixed model. Although PROC MIXED was not developed with kriging problems in mind, you can use it to solve the kriging problem in spatial models with a moderate number of observations. Examples of using PROC MIXED for ordinary and universal kriging are presented in Section 11.7.

The remaining sections of this chapter look at various PROC MIXED examples for spatial statistics.

11.5 Example: Estimating Spatial Covariance

This section presents a characterization example. Here, the objective is to find the most appropriate spatial model and estimate its covariance parameters.

The data for this example are from an agronomic uniformity trial, conducted on an 8×8 grid of plots. In a uniformity trial, a test crop is grown on a field with no experimental treatments applied. The response variable, in this case yield, is recorded. The idea is to characterize variability in the field in order to plan the design of experiments to be conducted on that field in the future. Traditionally, uniformity trials are used to decide how to block—e.g., whether to use a complete or incomplete block design or how to position the blocks. They may also be used to anticipate spatial variability.

The data are given as Data Set 11.5 “Uniformity Trial,” in Appendix 2, “Data Sets.” The variables ROW and COL locate the plots in the 8×8 grid. The field consisted of 4 quarters—the variable REP identifies each quarter field. Each quarter consisted of 4 “sub-quarters,” each identified by the variable BLOC. The response variable of interest was YIELD.

The researchers’ primary objective was to use the information from these data to design a yield trial with 16 treatments. Specifically, they wanted to decide whether to conduct future experiments on the field as (1) a complete block design with 4 blocks (denoted REP in this example), (2) an incomplete block design with 16 blocks (denoted BLOC to avoid confusion with complete blocks), or (3) a completely random design with spatially correlated errors.

It is worth noting that, historically, blocking has acted as a simple model of spatial variation in field experiments. That is, blocking assumes that experimental units close together behave alike and can thus be grouped into reasonably homogeneous blocks. The existence of spatial correlation is implicit in blocking. While spatial covariance models are not necessarily alternatives to blocking in the design of experiments, they do provide an alternative to the standard ANOVA model for block designs as a way of accounting for field heterogeneity.

11.5.1 Estimating the No-Nugget Model

PROC MIXED allows you to analyze the data for any of the models mentioned above.

Program

For example, you can use the following program to fit a spherical covariance model:

```
title 'no nugget - spherical covariance model';
proc mixed data=spatvar;
  model yield = ;
  repeated / subject = intercept
             type      = sp(sph) (row col);
  parms (0 to 10 by 2.5)
        (1 to 10 by 3 ) ;
run;
```

We strongly recommend a parameter search using the PARMS option to get the REML procedure started in the vicinity of plausible values ρ and σ^2 . These values are determined, for example, from previous experience or from an empirical semivariogram based on the residuals of an independent errors model. Otherwise, the REML procedure can converge to local maxima and provide grossly unreasonable estimates. The SUBJECT=INTERCEPT option treats all the observations in the data set as potentially correlated. You may need to change the SUBJECT=

option depending on your data. For example, if an experiment were conducted at two locations, you would want to specify SUBJECT=LOCATION so that you do not try to compute correlations among observations at different locations.

The results of this program are given in Output 11.1.

Results

Output 11.1 No-Nugget – Spherical Covariance Model

Parameter Search				
CovP1	CovP2	Variance	Res Log Like	-2 Res Log Like
0	1.0000	3.1578	-127.6941	255.3883
0	4.0000	3.1578	-127.6941	255.3883
0	7.0000	3.1578	-127.6941	255.3883
0	10.0000	3.1578	-127.6941	255.3883
2.5000	1.0000	3.0563	-113.9248	227.8496
2.5000	4.0000	3.0563	-113.9248	227.8496
2.5000	7.0000	3.0563	-113.9248	227.8496
2.5000	10.0000	3.0563	-113.9248	227.8496
5.0000	1.0000	6.6501	-118.1688	236.3377
5.0000	4.0000	6.6501	-118.1688	236.3377
5.0000	7.0000	6.6501	-118.1688	236.3377
5.0000	10.0000	6.6501	-118.1688	236.3377
7.5000	1.0000	9.6770	-117.5687	235.1374
7.5000	4.0000	9.6770	-117.5687	235.1374
7.5000	7.0000	9.6770	-117.5687	235.1374
7.5000	10.0000	9.6770	-117.5687	235.1374
10.0000	1.0000	13.0262	-117.9635	235.9270
10.0000	4.0000	13.0262	-117.9635	235.9270
10.0000	7.0000	13.0262	-117.9635	235.9270
10.0000	10.0000	13.0262	-117.9635	235.9270

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
1	2	227.50476920	0.00001863
2	1	227.50373100	0.00000000

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(SPH)	Intercept	2.7120
Residual		3.2607

Fit Statistics	
-2 Res Log Likelihood	227.5
AIC (smaller is better)	231.5
AICC (smaller is better)	231.7
BIC (smaller is better)	235.8

PARMS Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	0.35	0.5564

Interpretation

The smallest -2 Res Log Like was achieved in the “Parameter Search” table for $\rho = 2.5$. Since ρ is the parameter listed first in the “Covariance Parameter Estimates” table, it corresponds to CovP1 in the “Parameter Search” table. Notice that for a given value of ρ the value of the sill parameter (CovP2) does not appear to be of consequence. The same -2 Res Log Like value results. This is a consequence of profiling the residual variance from the likelihood problem. PROC MIXED computes the residual variance after solving the mixed model equations. Regardless of the value of CovP2, we obtain the same estimate of the (profiled) sill for a given value of ρ . This estimate is shown in the column “Variance.”

The estimated spatial covariance parameters are given in the “Covariance Parameter Estimates” table. The estimate of the *range* ρ is reported as SP(SPH); the estimate is 2.712. Note that in the spherical model this is the range in the geostatistical sense. That is, observations separated by more than 2.712 distance units are not spatially correlated. The estimate of the *sill* σ^2 is reported as “Residual.” The estimate of the sill σ^2 is 3.261.

The model-fitting criteria are shown in the “Fit Statistics” table. These can be compared to corresponding criteria for other covariance models, such as exponential or Gaussian, or to alternative models, such as the complete block or incomplete block models mentioned above.

The likelihood ratio test information is associated with the PARMS statement. Treat this output with care. The parameter tested is ρ , but the hypothesis actually tested here is $H_0: \rho = \rho_0$, where ρ_0 is the starting value of ρ from the search procedure initiated by the PARMS option. You can determine ρ_0 from the output: it is the value of ρ that corresponds to the highest value of the REML log likelihood (Res Log Like) in the “Parameter Search” table. In this case, $\rho_0 = 2.5$. The *p*-value of 0.5564 means that there is not a significant difference between the log likelihood evaluated at the final converged estimate and at the starting values from which the optimization commenced. It does *not* mean that there is no significant spatial autocorrelation.

11.5.2 Likelihood Ratio Test of Spatial Covariance

A more reasonable way to evaluate the model is to test $H_0: \rho = 0$. You can do this by computing a likelihood ratio statistic as the difference between the $-2 \text{ Res Log Likelihood}$ given above (227.5) and the $-2 \text{ Res Log Likelihood}$ obtained from the model for which $\rho = 0$ —i.e., the independent errors model. You can obtain the latter using this program:

```
title 'zero covariance model - independent errors';
proc mixed data=spatvar;
    model yield = ;
run;
```

The resulting $-2 \text{ Res Log Likelihood}$ is 255.4. Thus, the likelihood ratio statistic is $255.4 - 227.5 = 27.9$. Because the spherical model has two covariance parameters (ρ and σ^2), whereas the independent errors model has one (σ^2), the resulting likelihood ratio test has 1 degree of freedom. Comparing 27.9 to $\chi^2_{(1)}$, H_0 is clearly rejected for any reasonable α level. We can conclude that these data contain significant spatial variability.

11.5.3 Comparing Spatial Covariance Models

You can compare the spherical model to other models.

Program

For instance, you can estimate the parameters of the exponential model, using this program:

```
title 'no nugget - exponential covariance model';
proc mixed data=spatvar;
    model yield = ;
    repeated / subject = intercept
        type      = sp(exp)(row col);
    parms (0 to 10 by 2.5)
        (1 to 10 by 3 ) ;
run;
```

This program is identical to the program for the spherical model, except TYPE=SP(EXP) replaces TYPE=SP(SPH) in the REPEATED statement. The same program using TYPE=SP(GAU) estimates the parameters of the Gaussian covariance model.

The results for the exponential and Gaussian models are given in Output 11.2.

Results

Output 11.2 Results of Fitting Exponential and Gaussian Covariance Models

a. Exponential

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(EXP)	Intercept	1.7548
Residual		4.0393

Fit Statistics	
-2 Res Log Likelihood	232.6
AIC (smaller is better)	236.6
AICC (smaller is better)	236.8
BIC (smaller is better)	240.9

b. Gaussian

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(GAU)	Intercept	0
Residual		3.1578

Fit Statistics	
-2 Res Log Likelihood	255.4
AIC (smaller is better)	257.4
AICC (smaller is better)	257.5
BIC (smaller is better)	259.5

Interpretation

You can use the model-fitting information in Output 11.2 to decide whether either the exponential or Gaussian model fits the data better than the spherical. Because all three models have two covariance parameters (ρ and σ^2), a likelihood ratio test is not possible. However, the fact that the spherical model has the smallest AIC, AICC, and BIC statistics and the smallest –2 Res Log Likelihood suggests that the spherical model is the best of the three.

11.5.4 Comparing Spatial Models to Nonspatial Models

You can also compare nonspatial models to spatial models. To complete the objectives of this uniformity trial, the complete block and incomplete block models must be evaluated.

Program

You can obtain the relevant model-fitting information from these programs:

```

title 'rcb error model';
proc mixed data=spatvar;
  class rep;
  model yield = ;
  random rep;
run;

title 'incomplete block error model';
proc mixed data=spatvar;
  class bloc;
  model yield = ;
  random bloc;
run;

```

The results are given in Output 11.3.

Results

Output 11.3 Results of Nonspatial Models for Yield Trial Data

a. Complete block (RCB) error model

Covariance Parameter Estimates	
Cov Parm	Estimate
rep	0
Residual	3.1578

Fit Statistics	
-2 Res Log Likelihood	255.4
AIC (smaller is better)	257.4
AICC (smaller is better)	257.5
BIC (smaller is better)	256.8

b. Incomplete block error model

Covariance Parameter Estimates	
Cov Parm	Estimate
bloc	1.3553
Residual	1.8671

Fit Statistics	
-2 Res Log Likelihood	242.7
AIC (smaller is better)	246.7
AICC (smaller is better)	246.9
BIC (smaller is better)	248.3

Interpretation

The following points are relevant to the objective of the uniformity trial.

- The estimated REP variance component is 0 and the –2 Res Log Likelihood is 255.4, the same as the independent errors model without blocking. This means that because of spatial variability there is as much variability among the plots *within* the proposed complete blocks as there is variability *among the blocks themselves*. Thus, a randomized complete block design is a poor choice for an experiment conducted in this field. The resulting information would be just as imprecise as a completely random design.

- The estimated BLOC variance is 1.3553, the residual error is reduced from 3.158 to 1.867, and the $-2 \text{ Res Log Likelihood}$ is 242.7. Thus, the incomplete block design would clearly do a better job of controlling field heterogeneity than would a complete block design. However, the $-2 \text{ Res Log Likelihood}$ is much higher than either the exponential or spherical covariance model. Thus, for these data, using the incomplete block model is less desirable than modeling spatial covariance directly. Section 11.6 further examines the consequences of using a blocked model versus a covariance model.

11.5.5 Estimating the Model with a Nugget Effect

You can estimate a model with a nugget effect by modifying the PARMS and REPEATED statements.

Program

For example, the following program allows you to estimate a spherical model with a nugget effect using the R-side formulation (see Section 11.4.1):

```
title 'nugget effect - exponential covariance model';
proc mixed data=spatvar;
  model yield = ;
  repeated / subject = intercept
    type      = sp(exp) (row col)
    local;
  parms (0      to 10      by 2.5 )
        (1      to 10      by 3     )
        (0.05   to 1.05   by 0.25);
  run;
```

The REPEATED statement is identical to the REPEATED statement used to estimate the no-nugget spherical model given in Section 11.5.1, except that the LOCAL option is added to the list of options. The LOCAL option adds the NUGGET parameter σ_l^2 to σ^2 and ρ used in the no-nugget model—i.e., it adds $\sigma_l^2 \mathbf{I}$ to the covariance matrix so that the modified covariance matrix is now $\sigma^2 \mathbf{F} + \sigma_l^2 \mathbf{I}$, as defined in Sections 11.3 and 11.4.

If you use the PARMS statement, you must include some specification for the nugget variance σ_l^2 . The value(s) is given last because that is the order in which PROC MIXED treats σ_l^2 (the order in which you specify values in the PARMS statement corresponds to the order in which the parameters are listed in the “Covariance Parameter Estimates” table of the PROC MIXED output). Output 11.4 presents results using two different PROC MIXED statements. The first is the default:

```
proc mixed data=spatvar;
```

The second uses a modified convergence criterion (CONVG) based on the gradient of the objective function:

```
proc mixed data=spatvar convg=1e-7;
```

Results

Output 11.4 Iteration Results Fitting Nugget Effect – Exponential Covariance Model

a. Default

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
1	3	232.70643686	0.00166292
2	2	232.64887020	0.00058781
3	1	232.60809550	0.00015402
4	1	232.59632092	0.00009456
5	1	232.58895773	0.00006303
6	1	232.58404961	0.00004202
7	1	232.58077756	0.00002802
8	1	232.57859619	0.00001868
9	1	232.57714195	0.00001245
10	1	232.57617245	0.00000830
11	1	232.57552612	0.00000553
12	1	232.57509523	0.00000369
13	1	232.57480798	0.00000246
14	0	232.57480798	0.00000246
15	0	232.57480798	0.00000246
16	0	232.57480798	0.00000246

WARNING: Did not converge.

Covariance Parameter Values At Last Iteration		
Cov Parm	Subject	Estimate
Variance	Intercept	4.0390
SP(EXP)	Intercept	1.7548
Residual		0.000156

b. Using convg=le-7

Dimensions	
Covariance Parameters	3
Columns in X	1
Columns in Z	0
Subjects	1
Max Obs Per Subject	64

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
1	3	232.70643686	0.00357059
2	2	232.64887020	0.00347414
3	1	232.60809550	0.00053060
4	1	232.59632092	0.00002007
5	1	232.58895773	0.00000090
6	1	232.58404961	0.00000064
7	1	232.58077756	0.00000045
8	1	232.57859619	0.00000031
9	1	232.57714195	0.00000021
10	1	232.57617245	0.00000014
11	1	232.57552612	0.00000010

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	Intercept	4.0386
SP(EXP)	Intercept	1.7548
Residual		0.000351

Interpretation

With the default convergence criterion—one based on the Hessian of the optimization problem—PROC MIXED does not achieve convergence. The optimization gets stuck after iteration 13, unable to improve the objective function. With the modified convergence criterion, one based on the gradient (derivative) of the objective function and a lowered tolerance (CONVG=1E-7) convergence is achieved after 11 iterations. In general, you should be careful in relaxing convergence criteria, as the results may be imprecise. The default values for CONVH= and CONVG= depend on your particular hardware; in general, these values are approximately 1E-8. In this example, a small relaxation of the CONVG= criterion prevented PROC MIXED from getting stuck.

Also shown in Output 11.4 is the “Dimensions” table from the second PROC MIXED run. Because of the R-side formulation of the model with SUBJECT=INTERCEPT, PROC MIXED models these data as belonging to a single subject. Since no RANDOM statement was used, there are no columns in Z. You can also fit this model with a G-side formulation (as discussed in Section 11.4.1) with the following program:

```

data spatvar; set spatvar;
  obs = _n_;
run;
proc mixed data=spatvar convg=1e-7;
  class obs;
  model yield = ;

```

```

random obs / type=sp(exp) (row col);
parms (0      to 10    by 2.5 )
      (1      to 10    by 3     )
      (0.05   to 1.05  by 0.25);
run;

```

The results are shown in Output 11.5.

Output 11.5 Exponential Covariance Model with Nugget Effect and RANDOM Statement

Dimensions	
Covariance Parameters	3
Columns in X	1
Columns in Z	64
Subjects	1
Max Obs Per Subject	64

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
1	3	232.70643686	0.00357059
2	2	232.64887020	0.00347414
3	1	232.60809550	0.00053060
4	1	232.59632092	0.00002007
5	1	232.58895773	0.00000090
6	1	232.58404961	0.00000064
7	1	232.58077756	0.00000045
8	1	232.57859619	0.00000031
9	1	232.57714195	0.00000021
10	1	232.57617246	0.00000014
11	1	232.57552613	0.00000010

Convergence criteria met.

Covariance Parameter Estimates	
Cov Parm	Estimate
Variance	4.0386
SP(EXP)	1.7548
Residual	0.000351

Fit Statistics	
-2 Res Log Likelihood	232.6
AIC (smaller is better)	238.6
AICC (smaller is better)	239.0
BIC (smaller is better)	245.1

Interpretation

Key results are as follows.

The “Dimensions” table reveals that PROC MIXED continues to treat the data as from a single subject with 64 observations. In contrast to the R-side model in Output 11.4, there are now $N = 64$ columns in the $\mathbf{Z} = \mathbf{I}$ matrix.

The “Iteration History” table is the same as in the R-side model in Output 11.4. Notice that this holds for the number of function evaluations, the objective function (-2 Res Log Like), and the convergence criterion. The R-side and G-side models with nugget effects are statistically and numerically identical.

From the “Covariance Parameter Estimates” table:

- The “Variance” parameter, whose estimate is 4.0386, is called the partial sill. The sill is the sum of the partial sill and the nugget.
- The estimated range, ρ , appears as “SP(EXP).” The estimated ρ is 1.7548. The practical range, the distance at which the spatial autocorrelation has declined to less than 0.05, is three times this amount, $3 \times 1.7548 = 5.2644$.
- The estimated nugget, σ_i^2 , appears as “Residual.” The estimated σ_i^2 is 0.003994.
- The –2 Res Log Likelihood and information criteria in the “Fit Statistics” table we obtain are almost indistinguishable from the model-fitting statistics obtained for the no-nugget exponential model discussed in Section 11.5.3. This is evidence that the nugget effect is zero. A formal likelihood ratio statistic to test $H_0: \sigma_i^2 = 0$ can be calculated from the difference between the –2 Res Log Likelihood of the nugget and no-nugget models. The test has 1 degree of freedom because the nugget model has three parameters and the no-nugget model has two. In this case, the result seems obvious by inspection—actually doing the calculations does not seem warranted.

11.6 Using Spatial Covariance for Adjustment: Part 1, Regression

PROC MIXED is especially useful for estimating regression and analysis of variance models when the assumption of independent errors is violated, as it is whenever spatial variability is present. Failure to account for correlated errors can result in severely distorted regression estimates or erroneous conclusions about treatment effects. In this section and the next, we present two examples showing how to use PROC MIXED to adjust for spatial correlation. The first is a regression example, and the second is an analysis of variance example (Section 11.7).

11.6.1 A Regression Example

This example is from an environmental study. (We acknowledge Carol Gotway Crawford for providing this example.) The investigator wanted to evaluate water drainage characteristics at a potential hazardous waste disposal site. The data are given as Data Set 11.6, “Hazardous Waste,” in Appendix 2, “Data Sets.” Thirty samples were taken at various locations in the site. Locations were classified by their north-south (NORTHING) and east-west (EASTING) coordinates. Water movement was measured by log-transmissivity (LOGT). The investigator suspected that a linear relationship existed between LOGT and the thickness of a layer of salt (SALT), which was a geological feature of the area. Also, the LOGTs of samples relatively close together were suspected to be more alike than those farther apart.

Spatial Covariance Model

You can describe the linear relationship between LOGT and SALT by the regression model

$$\text{LOGT} = \beta_0 + \beta_1 \text{SALT} + e$$

However, unlike standard regression, inference on this model must take into account spatial correlation among the errors.

Program

An example of an appropriate PROC MIXED program to analyze these data follows:

```
proc mixed data=a;
  model logt = salt / solution;
  repeated / subject = intercept
    type      = sp(exp)(easting northing)
    local;
  parms (1   to   7   by 3   )
        (2   to  17   by 5   )
        (0.1 to  0.5 by 0.2);
  run;
```

The choice of an exponential covariance model with a nonzero nugget effect and the range of starting values used in the PARMS statement is based on the investigators’ prior experience. In the absence of clear-cut guidelines, you can use model comparisons like those presented in Section 11.5.

The results of this analysis are given in Output 11.6.

Results

Output 11.6 Regression Example with Exponential Covariance Model

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	Intercept	2.5252
SP(EXP)	Intercept	11.5305
Residual		0.04117

Fit Statistics	
-2 Res Log Likelihood	78.8
AIC (smaller is better)	84.8
AICC (smaller is better)	85.8
BIC (smaller is better)	88.8

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-5.0993	0.8920	0	-5.72	.
salt	-0.02117	0.006229	28	-3.40	0.0021

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
salt	1	28	11.55	0.0021

Interpretation

The “Covariance Parameter Estimates” table displays the estimate of the partial sill, σ^2 (“Variance”), as 2.5252 and the estimate of the range, ρ (“SP(EXP)”), as 11.53 (the practical range is 3×11.53). The estimate of the nugget effect, σ_l^2 (“Residual”), is 0.04117.

The “Solutions for Fixed Effects” table gives the estimate of the intercept β_0 as -5.0993 with a standard error of 0.892. The estimate of the slope β_1 is -0.02117 with a standard error of 0.00623.

The “Test of Fixed Effects” table shows the F -statistic for the test of $H_0: \beta_1 = 0$ as $F = 11.55$ with a p -value of 0.0021.

The -2 Res Log Likelihood and other fit criteria in the “Model Fitting Information” table can be used to compare the exponential covariance model with other alternatives. We can compare the -2 Res Log Likelihood of this model to that of the independent errors model. This tests $H_0: \rho = 0$ and $\sigma_l^2 = 0$, which is interpreted as a test of the existence of spatial variability. This test is a χ^2 test with 2 degrees of freedom, corresponding to the two parameters tested. Fitting the independent errors model, the -2 Res Log Likelihood is 94.07. The resulting likelihood ratio χ^2 value is $94.07 - 78.79 = 15.28$. Therefore, you conclude that significant spatial variability exists at this site.

11.6.2 Comparison with Independent Errors Model

A final point to consider is how much the regression estimates and conclusions about the effect of SALT thickness on LOGT are affected by accounting for spatial covariance among the errors. You can estimate the regression equation under the assumption of independent errors

using several SAS procedures, such as PROC REG and PROC GLM, as well as PROC MIXED. The PROC MIXED program follows:

```
proc mixed data=a;
   model logt = salt / solution;
run;
```

This results in an estimated intercept, β_0 , and slope, β_1 , of -5.025 and -0.034 , respectively. The F -value to test $H_0: \beta_1 = 0$ is 26.46 with a p -value of 0.0001 . The magnitude of the slope estimate is roughly 50% greater and the F -value is more than twice as large. While the general conclusions *for this particular data set* do not change, it is easy to see that they could be greatly affected in other data sets.

Different Covariance Models

Finally, you could try a different covariance model. For example, if you fit a spherical model with a nugget effect, the -2 Res Log Likelihood is 79.49 , which is only negligibly different from the exponential model. The estimated intercept and slope are -5.110 and -0.021 , respectively. The F -value to test the slope is 10.95 with a p -value of 0.0026 . Again, these are only negligibly different from the results obtained using the exponential model. This is typical of spatially correlated data—the major impact on inference results from using a *reasonable* covariance model. The specific model used is not nearly as important, as long as it is “in the ballpark.”

11.7 Using Spatial Covariance for Adjustment: Part 2, Analysis of Variance

This example is from an agronomic yield trial reported by Stroup, Baenziger, and Mulinze (1994). The investigator wanted to compare the mean yield of 56 varieties of wheat. A randomized complete block (RCB) design was used—all varieties were planted in each of four blocks. The data are given in Data Set 11.7, “Alliance Wheat Yield,” in Appendix 2, “Data Sets.”

The variables in the data set are wheat variety (ENTRY), block (REP), field plot location by row (LATITUDE) and column (LONGITUDE), and YIELD.

11.7.1 Possible Models for Wheat Yield Trial

The standard analysis of variance model for this experiment is

$$Y_{ij} = \mu + r_i + \tau_j + e_{ij}$$

where r_i is the block (REP) effect, τ_j is the variety (ENTRY) effect, and the errors (e_{ij}) are iid normal. As a practical matter, however, 56 field plots per block may be too many to expect within-block homogeneity. In addition, after doing a standard ANOVA, the investigator was troubled by several results that did not make sense. These are discussed below.

Two alternative models for these data are (1) to use the same model equation, but assume the e_{ij} are spatially correlated, or (2) drop the block (REP) effect from the model, and use the model equation

$$Y_{ij} = \mu + \tau_j + e_{ij}$$

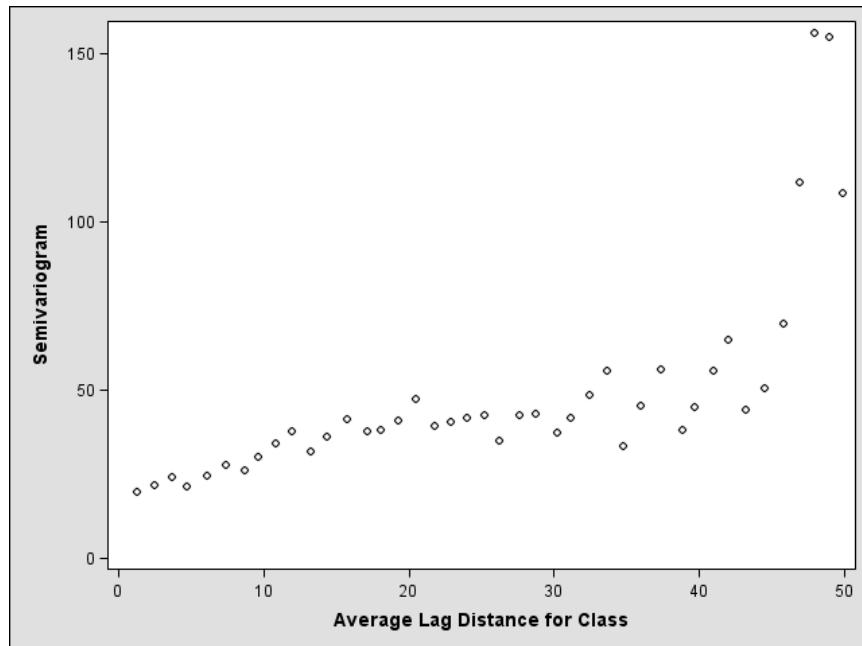
where the errors are assumed to be spatially correlated. This model made biological sense to the investigator because the variability at the location did appear to consist of irregular fertility gradients that would be expected in a spatial correlation process.

The model was fit in two ways, one using REML covariance estimates (obtained as in Sections 11.5 and 11.6.1) and the other using an empirical semivariogram computed with PROC VARIOGRAM to obtain a reasonable spatial covariance model. PROC MIXED does not compute semivariograms or use them in model fitting. However, in many cases (these data are an example) using a semivariogram procedure external to PROC MIXED may produce more reasonable covariance estimates than the REML procedure. Semivariogram procedures are described in, e.g, Journel and Huijbregts (1978), Cressie (1993), and Schabenberger and Gotway (2005).

You can compute the semivariogram using the following SAS statements:

```
proc mixed data=alliance;
  class Rep Entry;
  model Yield = Entry / outp=residuals;
  random Rep;
run;
proc variogram data=residuals outvar=plot_variogram;
  coordinates xc=lng yc=lat;
  compute lagd=1.2 maxlags=50;
  var resid;
run;
```

The program uses the randomized block analysis assuming independent errors to generate residuals. Then PROC VARIOGRAM computes the semivariogram for the residuals. This characterizes the spatial correlation not accounted for by the randomized block model. See the *SAS/STAT User's Guide* for details about the VARIOGRAM procedure. In this example the **lag distance** (LAGD) was 1.2, the size, in meters, of each plot (experimental unit). Consistent with comments in the introduction, these are lattice data with 1.2 meters between the centers of the plots in the field. The MAXLAGS option reflects the size of the field. The VARIOGRAM procedure does not produce listing output. The results are written to the OUTVAR= data set. Figure 11.2 graphs the contents of the data set, specifically, the variable VARIOG, containing the classical estimator of the semivariogram, versus the average distance in the lag class.

Figure 11.2 Empirical Semivariogram

You can discern three main features from the empirical semivariogram (Figure 11.2):

1. The semivariogram intercepts the Y-axis well above zero, indicating the possible presence of a nugget effect.
2. The shape of the semivariogram up through distances in the low 30s have roughly the shape of a spherical covariance model.
3. The semivariogram values are extremely high for the largest distances. These may be spurious, as they involve plots at extreme corners of the field, and may reflect edge effects, the fact that they are based on only a few pairs of plots, or both.

Section 11.7.2 shows the analysis with and without the nugget effect suggested by Figure 11.2. This section also shows the ENTRY means from the standard, independent error randomized block analysis of variance model compared with the means from the spatial model to show how dramatic the adjustment can be. Section 11.7.3 shows an alternative analysis based on truncating the semivariogram (eliminating the larger distances) and estimating the nugget, sill, and range based on the remaining semivariogram.

11.7.2 Analysis Using REML Estimation of Covariance Components

You can use the following program to compute the spatial model assuming the spherical correlation structure:

```

proc mixed data=alliance;
  class Entry Rep;
  model Yield = entry / ddfm=kr;
  random Rep;
  repeated / subject=intercept
    type=sp(sph) (latitude longitude);
  parms (9.88 18 45);
/*
  repeated / subject=intercept
    type=sp(sph) (latitude longitude) local;
  parms (1 18 18 25);
*/
  lsmeans name;
estimate 'arap v brul' entry 1 -1 0;
estimate 'arap v buck' entry 1 0 -1 0;
estimate 'arap v ks83' entry 1 0 0 0 0 0 0 0 0 -1 0;
estimate 'brul v ks83' entry 0 1 0 0 0 0 0 0 0 -1 0;
run;

```

The ESTIMATE statements given here are just examples. The DDFM=KR option computes adjusted standard errors for the fixed effects and appropriate degrees of freedom for tests of significance in order to control Type I error rates. There are two sets of REPEATED statements shown: the first assumes no nugget effect; the second includes the nugget effect (as likely seems necessary from the semivariogram). Note that PARMS statements go along with either REPEATED statement. The PARMS statements replace the default starting values for the REML algorithm with values suggested by the data. The technique by which PROC MIXED determines starting values for covariance parameters is not well suited for time series and spatial models. This can result in suboptimal starting values (see Output 11.7a for the default starting values in this example). The order of the starting values in the PARMS statement must be the same as the order in the output listing. If you are not sure about the order in which to enter covariance parameters in the PARMS statement, run the program without the PARMS statement and see how the covariance estimates are listed in the “Covariance Parameter Estimates” table.

The starting values for the no-nugget model are 9.88 for the REP variance (suggested by the estimate from the independent error analysis that generated residuals for the semivariogram, shown above), 18 for the range, and 45 for the sill (residual variance), the last two suggested by the semivariogram in Figure 11.2. The output for the no-nugget model appears in Output 11.7b.

For the nugget model, the starting value of 9.88 for the REP variance led to unreasonable estimates. Instead, you can reduce it (e.g., to 0.1, as shown in the example SAS code above). This starting value is suggested by the REP variance component estimate, 0, obtained in the no-nugget model. The REP variance estimate from the randomized block model is an artifact of spatial variation that is best modeled directly using the REPEATED statement. Once this is done, the REP variance disappears. Alternatively, you could drop the RANDOM REP statement. The starting values for the sill, range, and nugget parameter (18, 18, and 25, respectively) are suggested by inspection of the semivariogram (Figure 11.2). The results for the no-nugget model are shown in Output 11.7c.

Results

Output 11.7 Wheat Trial Covariance Estimates

a. Estimates Using Default Starting Values – Model with Nugget Effect

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Rep		4.97E-15
Variance	Intercept	10719
SP(SPH)	Intercept	7453.07
Residual		11.9218

b. Estimates for No-Nugget Spherical Model with Starting Values from PARMS Statement

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Rep		0
SP(SPH)	Intercept	21.1601
Residual		150.92

Fit Statistics	
-2 Res Log Likelihood	1106.3
AIC (smaller is better)	1110.3
AICC (smaller is better)	1110.4
BIC (smaller is better)	1109.1

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Entry	55	160	2.86	<.0001

c. Estimates for Spherical Model with Nugget Effect and Starting Values from PARMS Statement

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Rep		4.33E-19
Variance	Intercept	43.4224
SP(SPH)	Intercept	27.4570
Residual		11.4948

Fit Statistics	
-2 Res Log Likelihood	1067.9
AIC (smaller is better)	1073.9
AICC (smaller is better)	1074.0
BIC (smaller is better)	1072.0

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Entry	55	138	1.85	0.0021

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
arap v brul	0.8094	2.9738	150	0.27	0.7859
arap v buck	-8.1894	3.0798	152	-2.66	0.0087
arap v ks83	-0.1974	3.0179	151	-0.07	0.9479
brul v ks83	-1.0068	3.0007	153	-0.34	0.7377

Interpretation

The estimates in Output 11.7a for the “Variance” (sill) and “SP(SPH)” (range), obtained with PROC MIXED default starting values, are not sensible. The estimated sill is $\hat{\sigma}^2 = 10719$, which is nearly 100 times greater than the maximum values on the semivariogram. The estimated range is $\hat{r} = 7453.07$, which is several orders of magnitude longer than the entire field where the trial was conducted, as well as obviously inconsistent with the semivariogram.

The “Fit Statistics” tables for the two models estimated with starting values based on the semivariogram indicate that the model with the nugget effect provides a better fit. For example, its AICC is 1074 versus 1110.4 for the model without a nugget effect, and the –2 Res Log Likelihood is 1067.9 versus 1106.3. Both fit dramatically better than the randomized block analysis of variance model, whose AICC is 1221.8.

The estimated variance (sill) for the no-nugget and nugget models are $\hat{\sigma}^2 = 150.92$ and $\hat{\sigma}^2 = 43.42$, respectively. Both are higher than one would expect from the semivariogram, especially for the no-nugget model. For the no-nugget model, this could be partly a result of the poor fit without the nugget parameter, but for both models it could be a consequence of the high semivariogram values at the greatest distances. Section 11.7.3 considers this further.

The estimated range (denoted “SP(SPH)” in the output) is 21.16 for the no-nugget model and 27.46 for the model with nugget effect. These both appear to be consistent with the semivariogram.

The estimated nugget (denoted “Residual” in Output 11.7c) is 11.49.

The F -values in the “Type 3 Tests of Fixed Effects” tables test the equality of variety means and are affected by choice of covariance model. For the no-nugget model $F = 2.86$, whereas for the nugget model $F = 1.85$. The p -value for both models is highly significant. By contrast, the F -

value for the randomized block ANOVA model is 0.88 with a p -value of 0.7119. Modeling spatial variability would be the difference between concluding no significant variety effect (using ANOVA) and concluding that a statistically significant variety effect does exist (using models with spatial adjustment).

The standard errors of estimated differences between variety (ENTRY) means vary depending on the difference between the plots where the varieties in the pair were located. This is different from the RCB analysis, which assumes equal variance: the standard error of a treatment difference using the RCB model is 4.979. For these data, spatial adjustment substantially improves the precision of estimated treatment differences. Similarly, the standard errors of variety least-squares means are affected by location (see Output 11.10 below).

Like the F -value, the least-squares means and standard errors are substantially affected by spatial adjustment. Section 11.7.4 presents these results in detail. First, however, consider the issue of REML estimation of the covariance parameters.

11.7.3 Alternative Estimate of Covariance Parameters Based on Semivariogram

REML estimation is one possible way to estimate the spatial covariance parameters. The REML estimates given in Output 11.7 showed some inconsistency with the semivariogram (Figure 11.2). What about alternatives?

As noted before, the high values of the semivariogram at the greatest distances, where the number of pairs is relatively small and edge effects may have an impact, is one possible explanation for the REML estimates. Investigating alternative estimation procedures for spatial variation is an area of ongoing statistical research. One possible approach might be to truncate the semivariogram at a distance less than where the very high values begin to occur, and fit the model from the remaining values. The following SAS program uses PROC NLIN to estimate the nugget, range, and sill from the PROC VARIOGRAM output data set. You could also use PROC NLMIXED (see Chapter 15) to run the same model.

```

data variogram1;
  set plot_variogram;
  if distance <= 35;
  if lag=-1 then distance=0;
run;
proc nlin data=plot_variogram;
  parms r=18 nugget=17 sill=25;
  model vario = nugget +
    sill*(1.5*distance/r-0.5*(distance/r)**3)*(distance<r) +
    sill*(distance>=r);
  /*
  model covar=nugget*(distance=0) + sill*(1-1.5*distance/r +
    0.5*(distance/r)**3)*(distance<r);
  */
run;

```

The response variable in the model is VARIOG, the value for the semivariogram in the OUTVAR= data set from PROC VARIOGRAM. The program also shows a model using the value of the empirical covariance between pairs of observations a given distance apart. This is an alternative to using the semivariogram. In this case, the semivariogram yielded better results as measured by a lower AICC value, so in the interest of space, only these results are shown here. The estimated covariance parameters are shown in Output 11.8.

Output 11.8 Covariance Component Estimates from PROC NLIN

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
r	24.2770	3.5818	16.9146	31.6395
nugget	17.0688	2.7290	11.4593	22.6783
sill	25.4534	2.8618	19.5709	31.3359

You can use these parameters in PROC MIXED to reanalyze the wheat trial data with the following SAS statements.

```
proc mixed data=alliance noprofile;
  class entry rep;
  model Yield = entry / ddfm=kr;
  repeated / subject=intercept
    type=sp(sph) (latitude longitude) local;
  parms (25.453 24.277 17.069) / noiter;
  /* parms (20.057 13.974 16.684)/noiter; */
run;
```

Because the REP variance component consistently produced an estimate of 0 with the spatial models, it is dropped from this analysis. The NOPROFILE and NOITER statements act together to request that PROC MIXED not update the covariance parameters given in the PARMS statement, which are the PROC NLIN estimates from Output 11.8. There is a second PARMS statement, commented out, that shows the estimates obtained from the NLIN model using COVAR instead of VARIOG. Output 11.9 shows results from this analysis.

Output 11.9 Analysis of Wheat Yield Data with Covariance Components Estimated from Semivariogram

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	Intercept	25.4530
SP(SPH)	Intercept	24.2770
Residual		17.0690

Fit Statistics	
-2 Res Log Likelihood	1075.1
AIC (smaller is better)	1075.1
AICC (smaller is better)	1075.1
BIC (smaller is better)	1075.1

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Entry	55	71	1.55	0.0420

Compared to the REML estimate of the spherical model with nugget effect in Output 11.7c, the fit, as measured by the likelihood-based fit statistics, is not quite as good. For example, the $-2 \text{ Res Log Likelihood}$ is 1075.1 versus 1067.9 for the REML estimated model. The other fit statistics are not comparable because there is no adjustment for number of parameters when you use the NOPROFILE / NOITER options. The F -value is 1.55, compared with 1.85 for the REML estimated model, and the p -value is 0.0420 versus 0.0021. The least-squares means for the varieties are evaluated in the next section. Which estimating procedure is to be preferred has not been investigated systematically.

11.7.4 Impact of Spatial Adjustment on Treatment Effect Inference

LSMEANS of RCB Analysis Compared with Spatial Covariance Model

Output 11.10 shows the least-squares means and ranks of the variety means for analyses from three models: the randomized block analysis of variance model, the spherical model with nugget effect using REML estimates (Output 11.7c), and the spherical-nugget effect model using estimates from the semivariogram (Output 11.9). Only means for the 37 varieties with the highest yield according to the RCB analysis are shown. One of the major problems with randomized block designs with excessive block size is the tendency for some treatments to be located disproportionately in relatively good or poor plots. The performance of these treatments is thus misrepresented. One of the major advantages of spatial analysis is to adjust for localized effects within blocks.

Figure 11.3 compares the least-squares means and their standard errors for the three analyses and all varieties.

Output 11.10 Rankings of Least-Squares Means—RCB versus Spatial Analyses; Results for the First 37 Varieties according to RCB Analysis

Obs	Entry	rcb_rank	rcb	sp_reml_rank	sp_reml	sp_sv_rank	sp_sv
1	NE86503	1	32.6500	9	27.2240	7	27.3449
2	NE87619	2	31.2625	3	28.4850	3	28.5113
3	NE86501	3	30.9375	25	25.0176	23	25.3693
4	REDLAND	4	30.5000	4	27.9835	4	28.2372
5	CENTURK 78	5	30.3000	14	26.3342	13	26.7938
6	NE83498	6	30.1250	2	28.7051	2	29.1328
7	SIOUXLAND	7	30.1125	20	25.6558	17	25.9198
8	NE86606	8	29.7625	12	26.7965	11	27.0591
9	ARAPAHOE	9	29.4375	13	26.6590	10	27.1846
10	NE87613	10	29.4000	6	27.5617	6	27.7438
11	NE86607	11	29.3250	16	25.9103	18	25.8950
12	LANCER	12	28.5625	36	23.3332	37	23.4533
13	TAM 107	13	28.4000	38	22.7704	38	23.1857
14	CHEYENNE	14	28.0625	28	24.6733	29	24.5934
15	NE87446	15	27.6750	43	22.2195	43	22.6003
16	HOMESTEAD	16	27.6375	50	21.6695	45	22.2817

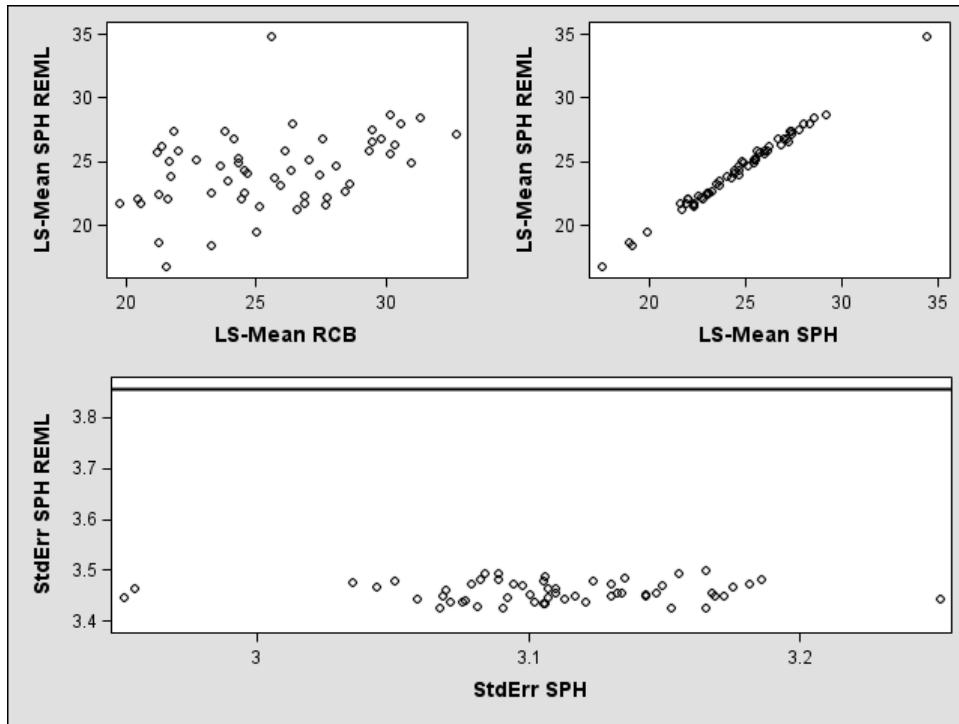
Obs	Entry	rcb_rank	rcb	sp_reml_rank	sp_reml	sp_sv_rank	sp_sv
17	SCOUT 66	17	27.5250	10	26.8777	12	26.9049
18	NE83404	18	27.3875	32	24.0492	28	24.6091
19	COLT	19	27.0000	23	25.1843	24	25.3638
20	NE86509	20	26.8500	42	22.4081	44	22.4749
21	NE87513	21	26.8125	47	21.8250	47	22.2276
22	LANCOTA	22	26.5500	52	21.2915	51	21.5861
23	NE85556	23	26.3875	5	27.9735	5	27.9470
24	NE87408	24	26.3000	30	24.3191	30	24.5666
25	BRULE	25	26.0750	18	25.8496	16	26.0550
26	NE87463	26	25.9125	37	23.2257	36	23.5890
27	NE87615	27	25.6875	34	23.8189	33	24.1829
28	BUCKSKIN	28	25.5625	1	34.8484	1	34.3714
29	NE87403	29	25.1250	51	21.4896	46	22.2405
30	NE87522	30	25.0000	53	19.4826	53	19.8192
31	NE87451	31	24.6125	31	24.1909	31	24.3331
32	NE86582	32	24.5375	39	22.5872	39	23.0104
33	GAGE	33	24.5125	29	24.3215	32	24.3327
34	NORKAN	34	24.4125	46	22.1449	42	22.7155
35	NE86482	35	24.2875	26	24.9743	26	24.7938
36	NE83406	36	24.2750	21	25.3137	22	25.4318
37	KS831374	37	24.1250	11	26.8564	14	26.6204

Interpretation

In Output 11.10, RCB is associated with results for the least-squares means obtained with the randomized blocks ANOVA model. Similarly, SP_REML is associated with results for the spherical model and REML estimation, and SP_SV is associated with results for the least-squares means using the semivariogram estimates of the spherical-nugget effect model. The column to the left and right of each least-squares mean are the rank and standard error, respectively. Two ENTRY lines are shaded: those for NE86503, the highest-ranking mean under the RCB analysis, and BUCKSKIN, the highest-ranking mean under the spatial analysis. You can see that the ranks change considerably from the ANOVA to the spatial models, but are consistent, though not identical, for the two estimation approaches for the spatial model. Which is “correct”?

One of the results that troubled the investigator was that the variety BUCKSKIN was known to be a superior performer. This was confirmed by observations in the field. BUCKSKIN performed noticeably better than the varieties in nearby plots. However, in the RCB analysis, BUCKSKIN had the 28th highest yield and certainly would not have been identified as a desirable variety. Using the spatial covariance model, BUCKSKIN’s least-squares mean yield was almost 7 units higher than the variety with the second best yield. Also, the varieties BRULE and KS831374 were known to be genetically similar. Although not statistically significant, their yields differed by 1.95 in the RCB analysis. This difference was 1.12 in the spatial correlation model. These were two of several reasons why the spatial correlation model gave more plausible results to the investigator than the RCB ANOVA.

Figure 11.3 Comparison of Least-Squares Means and Standard Errors for Three Analyses



The two upper panels in Figure 11.3 show how poorly the LS-means of the RCB and the spatial analysis agree. The large discrepancy between the RCB and the spatial LS-mean for variety BUCKSKIN is represented by the uppermost point in the left LS-mean panel. The right LS-mean panel shows how well the two spatial analyses agree. Whether the spatial covariance parameters are estimated by REML or held fixed at the estimates obtained from nonlinear least-squares fitting of a semivariogram model has very little influence on the estimates of varietal effects. The standard errors of the least-squares means in the REML analysis are somewhat higher compared to the analysis that holds the covariance parameters fixed. Note that the standard errors of the least-squares means are not constant in the spatial analysis, but are constant in the RCB analysis (horizontal line in lower panel of Figure 11.3). The spatial analysis takes into account small-scale local variation that is reflected in estimates and their standard errors.

Before we leave this application one final comment is necessary:

Spatial analysis is not a cure-all. Good experimental design is essential.

Although spatial analysis was able to “salvage” a plausible analysis from the wheat yield trial, the fact remains that the experimental design used for this trial was not a good choice. In general, randomized block designs should never be used for experiments with “large” (rules of thumb are difficult, but say >12) numbers of treatments. Stroup (2002) has demonstrated that even spatial analysis on large randomized block designs is relatively inefficient compared to well-known alternatives. Specifically, incomplete block designs should be used. They better capture local variation, making the randomized block ANOVA more likely to yield accurate information. In addition, when spatial analysis is necessary—that is, when even small

incomplete blocks still have within-block heterogeneity—Incomplete block design produces more efficient estimates of the spatial model. Chapter 12 shows how to use mixed model power analysis as a tool to design such experiments.

11.8 Example: Spatial Prediction—Kriging

Spatial prediction consists of computing predicted values for the analysis variable Y at observed or unobserved locations. This problem is different from estimation of fixed quantities, since Y is a random variable. For example, in the spatial regression example in Section 11.6.1, evaluating the regression equation at the estimates for the slope and intercept

$$\hat{\beta}_0 + \hat{\beta}_1 \text{SALT}$$

produces an estimate of the mean log-transmissivity. When data are uncorrelated, the best predictor of the log-transmissivity is also the best estimate of the mean log-transmissivity. The difference is reflected only in the precision of the quantities. In prediction problems we are interested in the mean square prediction error, which is larger than the mean square error. When data are correlated, the best predictor of Y is no longer equal to the best estimate of $E[Y]$.

“The” best predictor is typically not available without making some restrictions. For example, we commonly require that predictors are linear in the data and unbiased. The resulting best linear unbiased predictors (BLUPs) and their prediction standard errors can be computed in the spatial case with mixed model software if the spatial covariance structure can be modeled with G-side components. In the case of R-side specification, the typical formula needs to be modified. However, these modifications are incorporated in PROC MIXED, which enables you to perform best linear unbiased prediction (= linear kriging) in spatial models. This example demonstrates the computation of kriging estimates in two important scenarios. Section 11.8.1 discusses the case of ordinary kriging where the mean of the analysis variable is not known but is known to be constant throughout the spatial domain. Prediction at unobserved locations then does not require any further information other than the covariance parameter estimates and the coordinates of observed and unobserved data. Section 11.8.2 presents an example of universal kriging, where the mean of the analysis variable depends on other covariates. In many applications this dependence is expressed as a polynomial function in the coordinates themselves. Here, we consider the case of a true regressor variable.

The data for both analyses were kindly provided by Dr. Thomas G. Mueller of the Department of Agronomy at the University of Kentucky. They consist of 195 observations on the total soil carbon percentage (TC), the total soil nitrogen percentage (TN), and the C/N ratio (CN). In addition to these 195 observations—which correspond to a chisel-plowed portion of a field—observations on the variable TN are available at 200 further locations in the field. The data appear as Data Set 11.8, “C-N Ratios,” in Appendix 2, “Data Sets.”

11.8.1 Ordinary Kriging

Our first goal is to produce a map of the C/N ratios across the entire study area by **ordinary kriging**. This method of spatial prediction assumes that the mean of the spatial process is unknown but constant. The following program appends to the 195 observations for which the variable CN was observed a data set of prediction locations and performs the analysis.

Program

```

data predlocs;
  do x=0 to 500 by 12.5;
    do y=0 to 300 by 7.5;
      output;
    end;
  end;
run;
data fitthis;
  set cnratio(where=(tc ne .)) predlocs;
run;
proc mixed data=fitthis;
  model cn = / outp=OKpredictions;
  repeated / subject=intercept type=sp(exp)(x y);
run;

```

The spatial coordinates range from 0 to 500 and 15 to 300 in these data. The data set PREDLOCS contains a grid of $41 \times 41 = 1681$ prediction locations. It is combined with the analysis data in the second DATA step. Because the PREDLOCS data set does not contain the variable CN, missing values are assigned to all observations in data set FITTHIS that belong to the PREDLOCS data set. A missing value for the analysis variable is the “trigger” for PROC MIXED to compute kriging predictions in models with only R-side components. The kriging predictions are found in the OUTP= data set. The right-hand side of the “empty” MODEL statement corresponds to the ordinary kriging assumption of an unknown but constant mean. The mean is estimated by PROC MIXED as an intercept.

Output 11.11 Results from Spatial Analysis of C/N Ratios

Dimensions	
Covariance Parameters	2
Columns in X	1
Columns in Z	0
Subjects	1
Max Obs Per Subject	195

Number of Observations	
Number of Observations Read	1876
Number of Observations Used	195
Number of Observations Not Used	1681

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(EXP)	Intercept	13.8664
Residual		0.3187

Fit Statistics	
-2 Res Log Likelihood	277.8
AIC (smaller is better)	281.8
AICC (smaller is better)	281.9
BIC (smaller is better)	288.3

Interpretation

Key results are as follows:

- From the “Dimensions” and “Number of Observations” tables it is seen that PROC MIXED reads all 1876 observations in the FITTHIS data set, but only 195 observations are used in the analysis. These are the observations that correspond to complete records for CN, X, and Y. All observations are used, however, when predicted values are computed in the OUTP= data set.
- The –2 Res Log Likelihood is 277.8. The corresponding value for a model without spatial covariance is 321.3 (not shown). The difference is a χ^2 variable with 1 degree of freedom. A test for spatial autocorrelation is highly significant.
- The estimate of the range parameter ρ is 13.8664, and the practical range is $3 \times 13.8664 = 41.5992$ distance units. The estimate of the sill, the variance of the process, is 0.3187.

To demonstrate prediction versus estimation and a property of kriging without nugget effect, the following statements print observations in the OUTP= data set for a particular location.

Program

```
proc print data=OKpredictions (where=((x=100) and (y=300)));
run;
```

Output 11.12a Predicted Values in OUTP= Data Set

Obs	x	y	TN	TC	CN	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
53	100	300	0.067	0.7225	10.7872	10.8183	0.063466	194	0.05	10.6931	10.9435	-0.031093
564	100	300	.	.	.	10.7872	0.000000	194	0.05	.	.	.

Interpretation

This set of coordinates was represented once among the observed data and once in the PREDLOCS data set. In Output 11.12a the predicted value for the observed value is 10.8183. Because the CN value is not missing for OBS 53, PROC MIXED does not invoke the BLUP formula (note that it would use the BLUP formula if a RANDOM statement had been used). The predicted value of 10.8183 corresponds to the intercept estimate. All observations in the OUTP= data set with non-missing CN have this same predicted value. If CN is missing, however, PROC MIXED invokes the best linear unbiased prediction formula. Notice that the predicted value for OBS 564 is identical to the observed value for OBS 53. In the absence of a nugget effect, kriging predictions are said to “honor the data.” That is, they interpolate the observed data. Consequently, the prediction standard error is zero at the observed location.

Output 11.12b shows the contents of the OUTP= data set for all observations that have an X coordinate of 100.

Output 11.12b Predicted Values in OUTP= Data Set for X =100

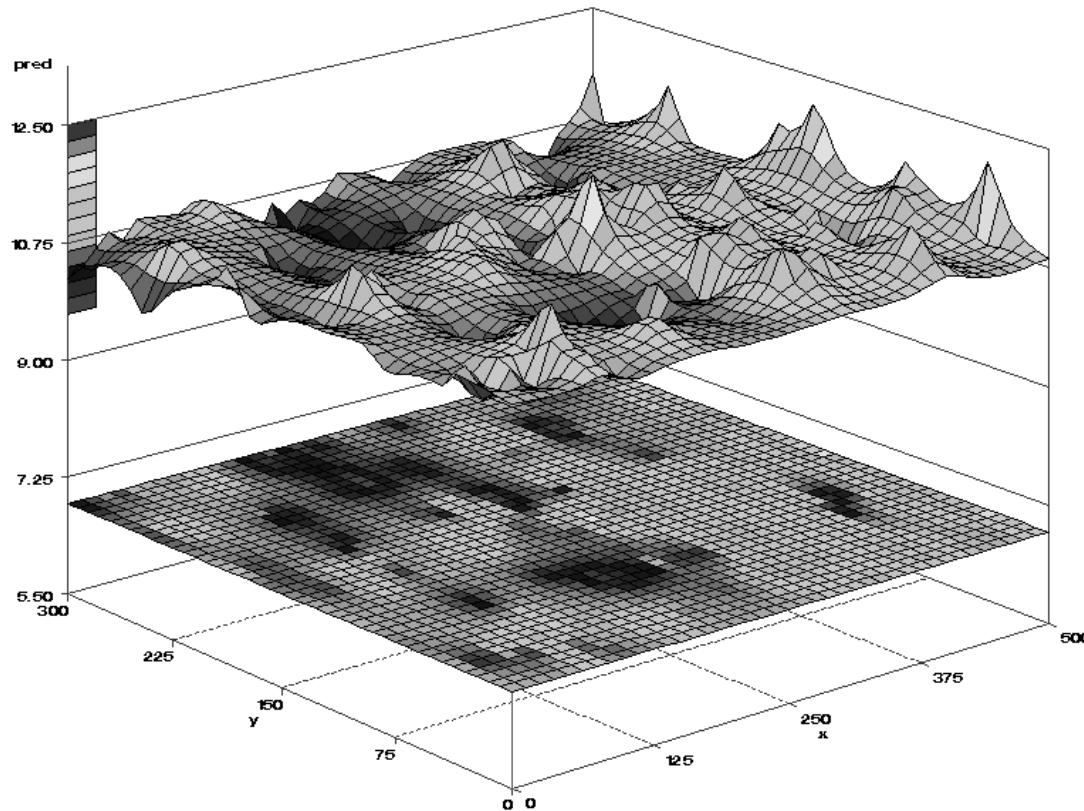
Obs	x	y	TN	TC	CN	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
44	100	27.5	0.0785	0.8245	10.5030	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.31529
45	100	35.0	0.0740	0.8090	10.9642	10.8183	0.06347	194	0.05	10.6931	10.9435	0.14591
46	100	75.0	0.0790	0.8525	10.7929	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.02539
47	100	95.0	0.0610	0.6190	10.1475	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.67079
48	100	145.0	0.0890	0.9310	10.4621	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.35619
49	100	170.0	0.0795	0.8830	11.1179	10.8183	0.06347	194	0.05	10.6931	10.9435	0.29961
50	100	175.0	0.0890	0.9055	10.1756	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.64269
51	100	230.0	0.0765	0.7875	10.2912	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.52709
52	100	240.0	0.0795	0.8560	10.7661	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.05219
53	100	300.0	0.0670	0.7225	10.7872	10.8183	0.06347	194	0.05	10.6931	10.9435	-0.03109
524	100	0.0	.	.	.	10.7884	0.53485	194	0.05	9.7335	11.8432	.
525	100	7.5	.	.	.	10.7562	0.49739	194	0.05	9.7752	11.7372	.
526	100	15.0	.	.	.	10.7095	0.45339	194	0.05	9.8153	11.6037	.
527	100	22.5	.	.	.	10.6304	0.38254	194	0.05	9.8759	11.3849	.
528	100	30.0	.	.	.	10.6795	0.27365	194	0.05	10.1398	11.2192	.
529	100	37.5	.	.	.	11.0298	0.30355	194	0.05	10.4311	11.6284	.
530	100	45.0	.	.	.	11.0934	0.46916	194	0.05	10.1680	12.0187	.
531	100	52.5	.	.	.	11.0332	0.51768	194	0.05	10.0122	12.0542	.
532	100	60.0	.	.	.	10.9536	0.51682	194	0.05	9.9343	11.9729	.
533	100	67.5	.	.	.	10.8803	0.45264	194	0.05	9.9876	11.7730	.
534	100	75.0	.	.	.	10.7929	0.00000	194	0.05	.	.	.
535	100	82.5	.	.	.	10.6306	0.42913	194	0.05	9.7842	11.4769	.
536	100	90.0	.	.	.	10.3848	0.39057	194	0.05	9.6145	11.1552	.

Obs	x	y	TN	TC	CN	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
537	100	97.5	.	.	.	10.2605	0.30780	194	0.05	9.6534	10.8675	.
538	100	105.0	.	.	.	10.4986	0.47982	194	0.05	9.5523	11.4449	.
539	100	112.5	.	.	.	10.6355	0.52575	194	0.05	9.5986	11.6724	.
540	100	120.0	.	.	.	10.7086	0.53641	194	0.05	9.6507	11.7666	.
541	100	127.5	.	.	.	10.7294	0.52254	194	0.05	9.6988	11.7600	.
542	100	135.0	.	.	.	10.6833	0.47759	194	0.05	9.7413	11.6252	.
543	100	142.5	.	.	.	10.5384	0.30778	194	0.05	9.9314	11.1454	.
544	100	150.0	.	.	.	10.6376	0.39832	194	0.05	9.8520	11.4232	.
545	100	157.5	.	.	.	10.8303	0.47758	194	0.05	9.8884	11.7722	.
546	100	165.0	.	.	.	10.9962	0.39836	194	0.05	10.2105	11.7819	.
547	100	172.5	.	.	.	10.6513	0.23839	194	0.05	10.1811	11.1214	.
548	100	180.0	.	.	.	10.3393	0.39638	194	0.05	9.5575	11.1210	.
549	100	187.5	.	.	.	10.4627	0.49435	194	0.05	9.4877	11.4377	.
550	100	195.0	.	.	.	10.5225	0.52518	194	0.05	9.4867	11.5583	.
551	100	202.5	.	.	.	10.5545	0.54078	194	0.05	9.4879	11.6211	.
552	100	210.0	.	.	.	10.5460	0.53996	194	0.05	9.4811	11.6110	.
553	100	217.5	.	.	.	10.4868	0.51034	194	0.05	9.4803	11.4933	.
554	100	225.0	.	.	.	10.3806	0.40138	194	0.05	9.5890	11.1722	.
555	100	232.5	.	.	.	10.4085	0.28854	194	0.05	9.8394	10.9776	.
556	100	240.0	.	.	.	10.7661	0.00000	194	0.05	.	.	.
557	100	247.5	.	.	.	10.7652	0.45052	194	0.05	9.8766	11.6537	.
558	100	255.0	.	.	.	10.7872	0.52157	194	0.05	9.7585	11.8158	.
559	100	262.5	.	.	.	10.8162	0.54534	194	0.05	9.7407	11.8918	.
560	100	270.0	.	.	.	10.8473	0.54857	194	0.05	9.7653	11.9292	.
561	100	277.5	.	.	.	10.8762	0.53707	194	0.05	9.8169	11.9354	.
562	100	285.0	.	.	.	10.8880	0.51074	194	0.05	9.8806	11.8953	.
563	100	292.5	.	.	.	10.8591	0.44738	194	0.05	9.9767	11.7414	.
564	100	300.0	.	.	.	10.7872	0.00000	194	0.05	.	.	.

Interpretation

The values for PRED are identical for the observations where CN is not missing. These predictions are **not** kriging predictions. For observations with a missing value for CN, the variable PRED contains the kriging prediction and STDERRPRED is the prediction standard error. The square of this value is known as the **kriging variance** in geostatistical parlance. You can easily identify which prediction locations correspond to locations where data were observed: their prediction standard error is exactly zero.

Figure 11.4 shows a plot of the predicted values at the prediction locations and a contour projection.

Figure 11.4 Predicted C/N-Ratio Surface and Contour

11.8.2 Universal Kriging

In Section 11.8.1 it was assumed that the mean C/N ratio is constant across the field. The spatial covariance structure was used to estimate the spatial dependency and to produce predictions of the actual (not average) C/N ratio everywhere on the field. In many applications it is not reasonable to assume that the mean is spatially not changing. This is particularly true when analysis extends over larger regions.

Obtaining spatial predictions in PROC MIXED when the mean depends on other variables is no more difficult than computing ordinary kriging predictions. The following program computes kriging predictions in a spatial regression of total carbon percentage (TC) on total nitrogen percentage (TN). However, keep in mind that PROC MIXED cannot produce predicted values if the variables on the right side of the MODEL statement are missing. You need to supply values for all model effects in order to obtain predictions; only the response variable can be missing.

Program

```
proc mixed data=cnratio;
  model tc = tn / solution outp=UKpredictions;
  repeated / subject = intercept
             type      = sp(expa)(x y);
run;
```

Recall that the data comprise 195 observations for which TC and TN are available. These are the same observations as in the analysis in 11.8.1. In addition there are 200 observations in the data set for which a TN observation is available, and the TC value is missing. The MODEL

statement relates the total carbon and total nitrogen percentages linearly. The OUTP= option requests predicted values and specifies the output data set. The REPEATED statement chooses an anisotropic covariance model for the analysis. In this model the covariance between two observations (with coordinates x and y) is computed as

$$\sigma^2 \exp\left\{-\theta_1 |x_i - x_j|^{p_1} -\theta_2 |y_i - y_j|^{p_2}\right\}$$

This covariance model was determined to be appropriate in this case in Schabenberger and Gotway (2005) based on a method proposed by Brownie and Gumpertz (1997).

The results are displayed in Output 11.13.

Output 11.13 Analysis of Spatial Regression with Anisotropic Covariance Structure

Dimensions	
Covariance Parameters	5
Columns in X	2
Columns in Z	0
Subjects	1
Max Obs Per Subject	195

Number of Observations	
Number of Observations Read	395
Number of Observations Used	195
Number of Observations Not Used	200

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(EXPA) x	Intercept	0.7301
SP(EXPA) y	Intercept	0.4068
Power x	Intercept	0.2004
Power y	Intercept	0.2009
Residual		0.001547

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-0.00838	0.02188	0	-0.38	.
TN	10.9238	0.2585	193	42.25	<.0001

Interpretation

The “Dimensions” and “Number of Observations” tables show that the 200 observations for which TC is missing were not used in the analysis. They are used, however, when predicted values are computed in the OUTP= data set.

The SP(EXPA) parameters in the “Covariance Parameter Estimates” table represent θ_1 and θ_2 in the covariance function. The estimates are quite different in the two directions—evidence of anisotropy in these data. The exponents of the distances are quite similar, however.

The estimates of the intercept and slope in the spatial regression of TC on TN are -0.00838 and 10.9238, respectively (“Solution for Fixed Effects” table). The slope is highly significant (p -value < 0.0001).

Output 11.14 displays the contents of the OUTP= data set for those observations where X = 5.

Output 11.14 Predicted Values in OUTP= Data Set for X = 5

Obs	x	y	TC	TN	Pred	StdErrPred
12	5	35	0.9460	0.0880	0.95291	0.011120
13	5	40	0.8790	0.0785	0.84913	0.010591
14	5	295	0.6505	0.0650	0.70166	0.010797
15	5	300	0.5800	0.0570	0.61427	0.011432
207	5	5	.	0.1075	1.16313	0.034181
208	5	10	.	0.1205	1.30556	0.034688
209	5	260	.	0.0600	0.62507	0.033345
210	5	265	.	0.0640	0.66845	0.033138

Interpretation

For observations where the response variable TC is not missing, the PRED variable contains the estimate of the mean. For example, for the first observation displayed in Output 11.12, $-0.00838 + 10.9238 \times 0.088 = 0.95291$.

For observations where TC is missing, but a value for TN is present, the variable PRED contains the kriging prediction and STDEERRPRED is the prediction standard error. The kriging variance for OBS 207, for example, is 0.034181².

11.9 Summary

This chapter showed how to analyze mixed models with spatial variability. PROC MIXED allows you to work with spatially correlated errors. Sections 11.3 and 11.4 described spatial correlation models and showed how you can include them in mixed models. Section 11.5 and 11.6 presented examples using PROC MIXED, showing how to write the needed programs and how to interpret the results. Section 11.8 made some cautious remarks regarding the reliability of REML estimation.



Power Calculations for Mixed Models

12.1 Introduction	479
12.2 Power Analysis of a Pilot Study	480
12.3 Constructing Power Curves	483
12.4 Comparing Spatial Designs.....	486
12.5 Power via Simulation.....	489
12.6 Summary.....	495

12.1 Introduction

Many of the results and conclusions in this book are based on a Fisherian consideration of a p -value corresponding to a null hypothesis of no treatment differences. We now turn our attention to a Neyman-Pearson testing framework formulated in terms of null and alternative hypotheses. The **power** of a statistical test is the probability of rejecting the null hypothesis when the alternative is in fact true. Power equals one minus the probability of a Type II error, and is also known as **sensitivity** or the **true positive rate**.

Our focus is on **prospective** power calculations, referring to the power of statistical hypothesis tests for new experiments that are yet to be conducted. Such calculations can be critical in determining the size and structure of a new experimental design and in optimizing information gain from experimental units. Along with Lenth (2001) and Hoenig and Heisey (2001), we recommend against **retrospective** power calculations, in which power statistics are used to embellish analysis of a data set in hand. Conclusions from such calculations are misleading, and in fact, the power estimates themselves are generally just a function of the corresponding p -values. However, power calculations on current data sets can be useful from a pilot study perspective, in the sense that reasonable estimates for required parameters can be obtained from existing data in order to perform an appropriate prospective power calculation.

There is a well-developed literature for power and sample size calculations for a variety of simple statistical models, and for general linear models in particular. Please refer to Castelloe and O'Brien (2000), Verbeke and Molenberghs (2000), Muller and Fetterman (2002), and SAS

Institute Inc. (2004), as well as the documentation for the POWER and GLMPOWER procedures in SAS/STAT, for details and references.

Power calculations for mixed models are more difficult due to their more complex covariance structure (Helms 1992, Stroup 2002, Tempelman 2006, Rosa, Steibel, and Tempelman 2006). Assuming the hypothesis test of interest is formed as a linear combination $\mathbf{K}'\boldsymbol{\beta}$, recall that our general t - and F -statistics can be written using the variance matrix $\mathbf{K}'[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{K}$. So the power associated with such tests is a function of the following:

- the magnitude of $\mathbf{K}'\boldsymbol{\beta}$, also known as the effect size
- the design matrix \mathbf{X} , including the number of its rows (the sample size) and the structure of its columns (from the fixed effects)
- the values of the variance and covariance parameters in \mathbf{V}
- the *test size*, commonly known as α , the probability of a Type I error, or one minus the *specificity* of the test

In the next section we show how to compute power in a pilot study for mixed model F -tests using a simple analytical approximation based on the noncentral F -distribution. The following two sections then discuss prospective power calculations via the same type of analytical approximation or via simulation, respectively. The analytical approximation techniques make use of an *exemplary* data set, where PROC MIXED is used to compute relevant statistics from an existing data set.

12.2 Power Analysis of a Pilot Study

Several authors (Helms 1992, Stroup 2002, Tempelman 2005, Rosa, Steibel, and Tempelman 2005) have shown how to compute a simple approximation to the power of an F -test by using an observed F -statistic to compute the noncentrality parameter needed for the power calculation. Consider the mice experiment example from Chapter 4 as a pilot study, and the following code produces Output 12.1.

Results

```

proc mixed data=mice nobound cl;
  class cage condition diet;
  model gain= condition|diet / solution ddfm=satterth;
  random cage cage*condition;
  ods output tests3=t3;
run;

data f_power;
  set t3;
  Noncen = NumDF*FValue;
  Alpha   = 0.05;
  FCrit   = finv(1-Alpha,NumDF,DenDF,0);
  Power   = 1 - probf(FCrit,NumDF,DenDF,Noncen);
run;

proc print data=f_power;
run;

```

The calculations employ the following steps:

1. Create an exemplary data set.
2. Run PROC MIXED on the data set, perhaps specifying alternative covariance parameters and holding them fixed with a PARMS statement. Save the F -statistics into an output data set.
3. In a subsequent DATA step, compute the approximate noncentrality parameter $\phi = (F \text{ computed by MIXED}) \times \text{rank}(K)$.
4. Compute the critical F (F_{crit}) as the value satisfying $\Pr(F_{\text{rank}(K),v,0} > F_{\text{crit}}) = \alpha$, using a specified value for α and the FINV function. Note that the third argument of $F_{\text{rank}(K),v,0}$ refers to the noncentrality parameter.
5. Compute power = $\Pr(F_{\text{rank}(K),v,\phi} > F_{\text{crit}})$ using the PROBF function.

The code produces the results shown in Output 12.1.

Results

Output 12.1 Results for Power Calculations for the Mice Experiment

Covariance Parameter Estimates				
Cov Parm	Estimate	Alpha	Lower	Upper
cage	5.0292	0.05	-4.2118	14.2703
cage*condition	-6.2415	0.05	-15.7836	3.3006
Residual	31.5567	0.05	17.5039	73.0938

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	4.72	5.16	0.0589
diet	2	16	0.82	0.4561
condition*diet	6	16	1.52	0.2334

Obs	Effect	NumDF	DenDF	FValue	ProbF	Alpha	Noncen	FCrit	Power
1	condition	3	4.72	5.16	0.0589	0.05	15.4900	5.66723	0.57952
2	diet	2	16	0.82	0.4561	0.05	1.6496	3.63372	0.16644
3	condition*diet	6	16	1.52	0.2334	0.05	9.1388	2.74131	0.42796

Interpretation

Consider the test for the interaction between condition and diet. The power value of 0.42796 indicates that for a data set having the following characteristics, the probability of observing a CONDITION \times DIET F -test p -value less than $\alpha = 0.05$ is approximately 0.42796:

- exactly the same \mathbf{X} matrix as the present one
- exactly the same $\boldsymbol{\beta}$ vector with values equal to those in the “Solution for Fixed Effects” table
- exactly the same \mathbf{V} matrix with values equal to the unbounded REML estimates as shown in the “Covariance Parameter Estimates table”

We now ask: what would the power for each F -test be if we had doubled the number of cages in the study, assuming the fixed effects and covariance parameter estimates stay the same?

Our strategy for computing these new powers is to run the same code as before but to increase the denominator degrees of freedom for each F -test by an appropriate amount. A convenient way to calculate these degrees of freedom is to run PROC MIXED on a doubled data set as follows, producing the results shown in Output 12.2.

Program

```

data mice1;
  set mice;
  cage = cage + 6;
data mice2;
  set mice mice1;
run;

proc mixed data=mice2 nobound cl;
  class cage condition diet;
  model gain=condition|diet / ddfm=satterth;
  random cage cage*condition;
run;

```

Results

Output 12.2 Approximate Degrees of Freedom for “Doubled” Mice Experiment

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	3	11.9	13.91	0.0003
diet	2	40	2.06	0.1405
condition*diet	6	40	3.81	0.0043

Interpretation

The denominator degrees of freedom for the CONDITION main effect have increased from 4.72 to 11.9, and those for DIET and CONDITION \times DIET have increased from 16 to 40.

We do not perform the power calculations using the doubled data set because the covariance parameter estimates are no longer the same as the original one. Instead, we rerun the code on the original pilot data and specify the new denominator degrees of freedom using the DDF= option as follows, producing the results shown in Output 12.3.

Program

```

proc mixed data=mice nobound cl;
  class cage condition diet;
  model gain = condition|diet / ddf=11.9,40,40;
  random cage cage*condition;
  ods output tests3=t3;
run;

data f_power2;
  set t3;
  Alpha  = 0.05;
  Noncen = NumDF*FValue;
  FCrit  = finv(1-Alpha,NumDF,DenDF,0);
  Power  = 1 - probf(FCrit,NumDF,DenDF,Noncen);
run;

proc print data=f_power2;
run;

```

Results

Output 12.3 Power Calculations for “Doubled” Mice Experiment

Obs	Effect	NumDF	DenDF	FValue	ProbF	Alpha	Noncen	FCrit	Power
1	condition	3	11.9	5.16	0.0162	0.05	15.4900	3.49914	0.80776
2	diet	2	40	0.82	0.4456	0.05	1.6496	3.23173	0.18124
3	condition*diet	6	40	1.52	0.1956	0.05	9.1388	2.33585	0.52161

Interpretation

The approximate power for the CONDITION \times DIET F -test has increased from 0.427 to 0.521. The power for CONDITION has increased even more, from 0.579 to 0.807, while that for DIET increased from 0.166 to only 0.181 due to its small F -statistic.

12.3 Constructing Power Curves

An analytical approximation to power makes it feasible to quickly compute entire power curves. While several different kinds of curves can be constructed by holding all but two variables fixed, a classical power curve plots effect size along the horizontal axis and power along the vertical axis. To facilitate this kind of calculation, we move from an F -distribution to a t -distribution in order to be able to specify a range of effect sizes in the numerator of the t -distribution noncentrality parameters.

The following SAS macro computes the power for an arbitrary set of ESTIMATE statements via the noncentral t -distribution. The GPLOT code following the macro plots the power curves (Figure 12.1). The macro also generates a JMP scripting language file that you can optionally run in JMP to produce the power curves (not shown).

Program

The input variables for this macro are as follows:

- InData specifies the exemplary input data set.
- ProcMixedStatements provides PROC MIXED statements, enclosed within %str(). The statements must contain one or more ESTIMATE statements corresponding to the single-degree-of-freedom contrasts for which you want to compute power curves.
- ProcMixedOptions specifies options for the PROC MIXED statement.
- AlphaValues specifies a list of alpha values over which to compute distinct power curves.
- EffectSizes specifies a list of effect sizes that form the X-axis values of the power curve plots.
- OutPath specifies the directory to which output should be written.

```
%macro MixedTPower;

  * run proc mixed;
  proc mixed data=&InData &ProcMixedOptions;
    &ProcMixedStatements
    ods output estimates=ests;
  run;

  libname OutLib "&OutPath";

  * compute power curves for each difference
  using a noncentral t-distribution;
  data OutLib.MixedPower;
    set ests;
    do alpha = &AlphaValues;
      do effectsize = &EffectSizes;
        tcrit = tinv(1-alpha/2,df);
        noncen = effectsize/stderr;
        power = sum(1, -cdf("t", tcrit,df,noncen),
                    cdf("t",-tcrit,df,noncen));
        output;
      end;
    end;
  run;

  * create JMP scripting code;
  filename jsllib "&OutPath.mixedtpower.jsl";

  data _null_;
    file jsllib;
    put "%str(data=open(%".\MixedPower.sas7bdat%"); ) ";
    put " ";
    put "Overlay Plot(X( :effectsize), Y( :power),
      Grouping( :Label, :alpha), ";
    put "Connect Thru Missing(1));";
  run;

%mend MixedTPower;
```

Using our mice experiment example again, we invoke the macro as follows:

```
%let InData = mice;
%let ProcMixedStatements=
%str(
  class cage condition diet;
  model gain = condition|diet / ddfm=satterth;
  random cage cage*condition;
  estimate "est1: condition 2-1" condition -1 1 0 0;
  estimate "est2: diet restrict-normal" diet -1 1 0;
  estimate "est3: condition*diet 1 restr-nor"
    diet -1 1 0
    condition*diet -1 1 0 0 0 0 0 0 0 0 0 0 0 0 0;
);
%let ProcMixedOptions = nobound;
%let AlphaValues = 0.05;
%let EffectSizes = 0 to 10 by 0.5;
%let OutPath = C:\temp\;

%MixedTPower;

data power; set outlib.mixedpower;
  symbol = substr(label,4,1)+0;
run;
goptions reset=all hsize=9in vsize=8in;
symbol1 color=black value=none i=join l=1 r=1 w=2;
symbol2 color=black value=none i=join l=2 r=1 w=2;
symbol3 color=black value=none i=join l=21 r=1 w=2;
axis1 minor =none
  order =(0 to 1 by 0.1)
  label =(font='Arial Black' Height=1.5 angle=90 'Power')
  offset=(0.3in,0.3in)
  value =(font='Arial Black' Height=1.3);
axis2 minor =none
  order =(-2 to 12 by 2)
  label =(font='Arial Black' Height=1.5 'Effect Size')
  offset=(0.3in,0.3in)
  value =(font='Arial Black' Height=1.3);
legend1 across =1
  cborder =black
  mode =protect
  position=(top left inside)
  label = none
  value =(font='Arial Black' Height=2
    'Condition 2-1'
    'Diet restrict-normal'
    'Condition*diet 1 restr-nor');

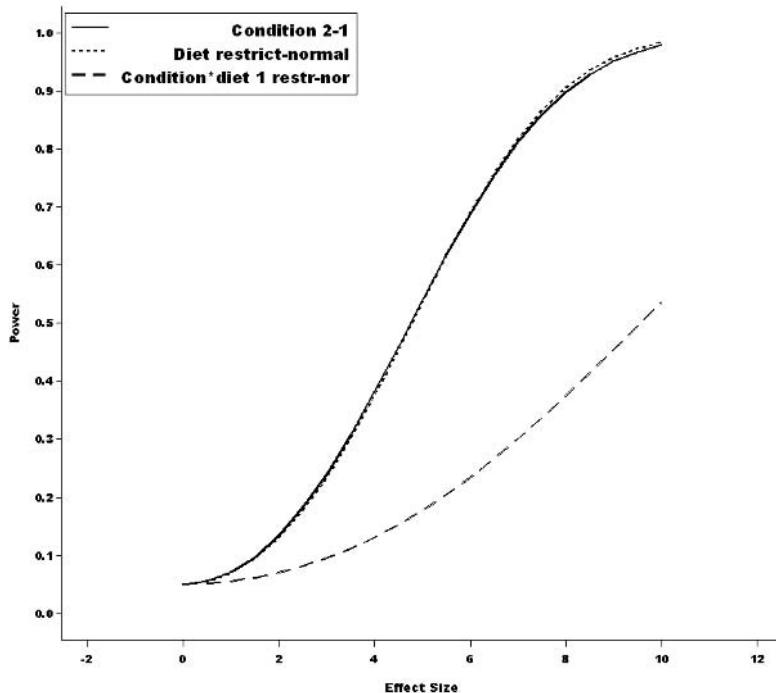
proc gplot data=power;
  plot power*effectsize=symbol /
    noframe
    vaxis=axis1 haxis=axis2
    legend=legend1;
run;
quit;
```

This code constructs three different power curves corresponding to the main effect contrasts for CONDITION and DIET and one interaction contrast for CONDITION \times DIET.

In contrast to earlier examples that used the F -distribution, this macro does not use the estimate of β during the power calculations because you specify effect sizes directly. So the response variable GAIN is used only to compute values of the covariance parameters. You can alternatively use a PARMS / HOLD= statement in order to directly specify the covariance parameters, in which case the response variable is not used at all in the power calculations.

Results

Figure 12.1 Power Curves for Selected Estimates from the Mice Experiment Pilot Study



Interpretation

Note that CONDITION and DIET have very similar power curves, even though the former is a whole-plot effect and the latter a split-plot effect. This is because DIET has a larger estimated standard error for this example but more degrees of freedom than CONDITION. The interaction estimate has low power compared to a similar range of effect sizes for the main effects.

12.4 Comparing Spatial Designs

The power calculations illustrated previously can be applied to any covariance structure you select in PROC MIXED. One interesting application is to use these calculations to help decide between potential experimental designs. In this section, we consider two different designs for a spatial field layout and compare power calculations for a representative set of contrasts.

The following code performs the comparison. The first design naively arranges the 16 treatments in systematic fashion across the field. The second is a classical 4×4 lattice design obtained from Cochran and Cox (1957). A spherical spatial model is assumed with shape parameter 3 and variance 16.

Program

```

data layout;
  do row=1 to 8;
    do col=1 to 8;
      output;
    end;
  end;
run;

/*---design option 1, 16 trts, 4 blocks, no randomization---*/
data td;
  input trt @@;
  datalines;
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16
;
;

/*---specify treatment effects as the data---*/
data design1;
  merge layout td;
  mu=-6*(trt=1)-2*(trt>1 and trt<5)+3*(trt>12);
run;

proc mixed noprofile;
  class trt;
  model mu=trt/ddf=48;
  parms (16) (3) /noiter;
  repeated / subject=intercept type=sp(sph)(row col);
  lsmeans trt/diff;
  contrast '1 vs 2' trt 1 -1 0;
  contrast '1 vs 4' trt 1 0 0 -1 0;
  contrast '1 vs 5' trt 1 0 0 0 -1 0;
  contrast '1 vs 10' trt 1 0 0 0 0 0 0 0 -1 0;
  contrast '1 vs 13' trt 1 0 0 0 0 0 0 0 0 0 0 -1 0;
  contrast '1 vs 16' trt 1 0 0 0 0 0 0 0 0 0 0 0 0 0 -1;
  ods output diff=trtdiff contrasts=b;
run;

proc means data=trtdiff;
  var estimate;
run;

data power1;
  set b;
  alpha=0.05;
  ncparm=fvalue;
  fcrit=finv(1-alpha,numdf,dendf,0);
  power=1-probf(fcrit,numdf,dendf,ncparm);
run;

proc print;
  var label numdf dendf alpha ncparm fcrit power;
run;

```

```

/*---design option 2; layout from Cochran & Cox, Experimental
Designs, plan 10.2
 4x4 lattice incomplete block design---*/

data td;
  input trt @@;
  datalines;
  1 2 3 4 1 5 9 13
  5 6 7 8 2 6 10 14
  9 10 11 12 3 7 11 15
  13 14 15 16 4 8 12 16
  1 6 11 16 1 8 11 14
  2 7 12 13 2 5 12 15
  3 8 9 14 3 6 9 16
  4 5 10 15 4 7 10 13
;

data design2;
  merge layout td;
  mu=-6*(trt=1)-2*(trt>1 and trt<5)+3*(trt>12);
run;

proc mixed noprofile;
  class trt;
  model mu=trt/ddf=48;
  parms (16) (3) /noiter;
  repeated / subject=intercept type=sp(sph) (row col);
  lsmeans trt/diff;
  contrast '1 vs 2' trt 1 -1 0;
  contrast '1 vs 4' trt 1 0 0 -1 0;
  contrast '1 vs 5' trt 1 0 0 0 -1 0;
  contrast '1 vs 10' trt 1 0 0 0 0 0 0 0 -1 0;
  contrast '1 vs 13' trt 1 0 0 0 0 0 0 0 0 0 0 -1 0;
  contrast '1 vs 16' trt 1 0 0 0 0 0 0 0 0 0 0 0 0 0 -1;
  ods output diffs=trtdiff contrasts=b;
run;

proc means data=trtdiff;
  var estimate;
run;

data power2;
  set b;
  alpha=0.05;
  ncparm=fvalue;
  fcrit=finv(1-alpha,numdf,dendf,0);
  power=1-probf(fcrit,numdf,dendf,ncparm);
run;

proc print;
  var label numdf dendf alpha ncparm fcrit power;
run;

```

Results

The results from the power calculations for the two designs are shown in Outputs 12.4 and 12.5.

Output 12.4 Results from Power Calculations for Spatial Design 1

Obs	Label	NumDF	DenDF	alpha	nparm	fcrit	power
1	1 vs 2	1	48	0.05	3.61279	4.04265	0.46131
2	1 vs 4	1	48	0.05	1.72960	4.04265	0.25167
3	1 vs 5	1	48	0.05	3.89495	4.04265	0.48970
4	1 vs 10	1	48	0.05	7.86350	4.04265	0.78457
5	1 vs 13	1	48	0.05	8.76264	4.04265	0.82652
6	1 vs 16	1	48	0.05	8.74409	4.04265	0.82573

Output 12.5 Results from Power Calculations for Spatial Design 2

Obs	Label	NumDF	DenDF	alpha	nparm	fcrit	power
1	1 vs 2	1	48	0.05	5.2505	4.04265	0.61235
2	1 vs 4	1	48	0.05	4.1617	4.04265	0.51566
3	1 vs 5	1	48	0.05	10.2209	4.04265	0.87950
4	1 vs 10	1	48	0.05	7.1501	4.04265	0.74543
5	1 vs 13	1	48	0.05	17.8239	4.04265	0.98522
6	1 vs 16	1	48	0.05	16.7661	4.04265	0.97989

Interpretation

The power for the second design is uniformly higher for all contrasts than that for the first design. Under an identical spatial covariance structure for exactly the same number of observations, the lattice is a clearly superior design. This example illustrates that choice of experimental design remains a vitally important aspect of scientific experimentation.

12.5 Power via Simulation

While the analytical approximations to power described in the previous sections are quite useful, they are not as comprehensive, nor necessarily as accurate, as computing power via direct simulation. Simulation is a definitive way to check the power of any test under known mixed model assumptions. With today's computing hardware and the power of the SAS language, the ability to conduct quick simulation studies is a valuable tool for any mixed modeler.

Several authors have conducted comprehensive simulation studies for certain mixed model scenarios. These include the following:

- Murray and Wolfinger (1994) and Murray et al. (1998) simulate power for *group randomized trials*—that is, data for which treatments are allocated to experimental units, like cities, and then many observations are collected from each experimental unit.
- Wright and Wolfinger (1997), Keselman et al. (1998, 1999a, 1999b), and Kowalchuk et al. (2004) focus on repeated measures designs and consider covariance structure selection and degrees of freedom adjustments

- Wolfinger et al. (2001), Chu, Weir, and Wolfinger (2004), and Feng et al. (2006) consider power simulations in the context of microarray data.

We return to the mice example from the beginning of this chapter to illustrate how to perform power calculations via simulation. This example is a little more complex than standard mixed model scenarios because we want to simulate a negative variance component. To do this, we use SAS/IML to compute the Cholesky root of the 3×3 matrix formed using the variance component for CAGE \times CONDITION and the residual variance, and then hand-code the results into a SAS DATA step. The following program accomplishes this and simulates the noise structure for 1000 simulations.

Program

```

/*---simulate noise structure from the mice data, IML is used
   in order to accommodate the negative variance component---*/
%let seed      = 71853048;
%let var_cage  = 5.03;
%let var_cc    = -6.24;
%let var_resid = 31.56;
%let nsim      = 1000;

proc iml;
  cov = &var_cc * j(3,3) + &var_resid * i(3);
  covh = half(cov)`;
  print cov covh;
run;

data noise;
  retain i j;
  do sim = 1 to &nsim;
    i = 1;
    j = 2;
    do cage = 1 to 6;
      u_cage = sqrt(&var_cage) * rannor(&seed);
      do condition = i,j;
        e1 = rannor(&seed);
        e2 = rannor(&seed);
        e3 = rannor(&seed);
        diet = "normal ";
        noise = u_cage + 5.0318983*e1;
        output;
        diet = "restrict";
        noise = u_cage - 1.240089*e1 + 4.8766977*e2;
        output;
        diet = "suppleme";
        noise = u_cage - 1.240089*e1 - 1.594895*e2 + 4.6085237*e3;
        output;
      end;
      j = j + 1;
      if (j = 5) then do;
        i = i + 1;
        j = i + 1;
      end;
    end;
  end;
run;

```

This program uses the I and J variables to help index the conditions within cages. Note that each unique pair of the four conditions is present within one of the six cages. The values of the Cholesky root obtained and printed from SAS/IML are hand-coded into the DATA step to create an appropriate error structure for the three observations within a unique cage and condition. This code simulates only the random effects and residual errors for the data; that is, no fixed effects have yet been added.

In any power simulation, it is important to verify that the data are simulated correctly and that the tests you are investigating have the appropriate size; that is, under the null hypothesis of no treatment effect, the *p*-values should have an approximate uniform distribution. One way to check these two things is to fit the desired mixed model to data simulated under the null and check the distributions of the estimated variance components and see how often *p*-values fall below a nominal value such as 0.05.

The following program does this using the noise data just generated.

Program

```
/*---verify test sizes are okay under the null hypothesis
   of no treatment effect---*/
ods exclude all;
ods noresults;

proc mixed data=noise nobound;
  by sim;
  class cage condition diet;
  model noise=condition|diet;
  random cage cage*condition;
  parms &var_cage &var_cc &var_resid;
  ods output covparms=cpsim tests3=t3sim;
run;

ods exclude none;
ods results;

proc sort data=cpsim;
  by covparm;
run;

proc means data=cpsim;
  by covparm;
  var estimate;
run;

data t3sim;
  set t3sim;
  if probf < 0.05 then sig = 1;
  else sig = 0;
proc sort data=t3sim;
  by effect;
proc means data=t3sim mean;
  by effect;
  var sig;
run;
```

Results

The results of the program are shown in Output 12.6.

Output 12.6 Results from Fitting to Simulated Noise Data Based on the Mice Experiment**Cov Parm=Residual**

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	31.5224199	11.0925121	8.6519013	86.2920458

Cov Parm=cage

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	4.7117467	5.8798187	-20.3398459	36.1323573

Cov Parm=cage*condition

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	-5.9348334	5.5319303	-27.0796760	29.9070699

Effect=condition

Analysis Variable : sig
Mean
0.0580000

Effect=condition*diet

Analysis Variable : sig
Mean
0.0410000

Effect=diet

Analysis Variable : sig
Mean
0.0420000

Interpretation

While the average estimate of the residual variance over the 1000 simulations is close to the true value of 31.56, the estimates of the variance components for CAGE and CAGE \times CONDITION are on average shrunk toward zero (for CAGE, the true value is 5.03 and the average estimate is 4.71; for CAGE \times CONDITION, the true value is -6.24 and the average estimate is -5.93). This could perhaps be due to the relatively small number of cages (six) in this design. These estimates are also perhaps introducing a small amount of bias in the test sizes: that for CONDITION is slightly liberal (0.058), whereas those for DIET and CONDITION \times DIET are slightly conservative (0.041 and 0.042, respectively), compared to a nominal level of 0.05.

We now perform the power simulation. We first need to add in values for the fixed effects in order to establish an alternative hypothesis, and we accomplish this by adding predicted mean values from the original data set. In order to compare results with the analytical calculations performed earlier, we use the degrees of freedom obtained from the “doubled” data set. The following program does this using the generated noise data.

Program

```

/*---simulate power under alternative as estimated
   from the pilot data---*/
proc mixed data=mice nobound;
   class cage condition diet;
   model gain=condition|diet / outpm=opm;
   random cage cage*condition;
run;

data opml;
   set opm;
   drop gain stderrpred df alpha lower upper resid;
run;

proc sort data=noise;
   by cage condition diet;
run;

data mice_sim;
   merge noise opml;
   by cage condition diet;
   y = pred + noise;
run;

proc sort data=mice_sim;
   by sim;
run;

ods exclude all;
ods noresults;

proc mixed data=mice_sim nobound;
   by sim;
   class cage condition diet;
   model y = condition|diet / ddf=11.9,40,40;
   random cage cage*condition;
   parms &var_cage &var_cc &var_resid;
   ods output covparms=cpsim tests3=t3sim;
run;

ods exclude none;
ods results;

```

```

proc sort data=cpsim;
  by covparm;
run;

proc means data=cpsim;
  by covparm;
  var estimate;
run;

data t3sim;
  set t3sim;
  if probf < 0.05 then sig = 1;
  else sig = 0;
proc sort data=t3sim;
  by effect;
proc means data=t3sim mean;
  by effect;
  var sig;
run;

```

Results

The results of the program are shown in Output 12.7.

Output 12.7 Results for Simulated Power for Mice Experiment

Cov Parm=Residual

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	31.5224199	11.0925121	8.6519013	86.2920458

Cov Parm=cage

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	4.7117467	5.8798187	-20.3398459	36.1323573

Cov Parm=cage*condition

Analysis Variable : Estimate				
N	Mean	Std Dev	Minimum	Maximum
1000	-5.9348334	5.5319303	-27.0796760	29.9070699

Effect=condition

Analysis Variable : sig
Mean
0.8600000

Effect=condition*diet

Analysis
Variable : sig
Mean
0.5130000

Effect=diet

Analysis
Variable : sig
Mean
0.2060000

Interpretation

The estimated covariance components are identical to those from the noise data in Output 12.6; that is, adding in the fixed effects parameters did not affect the estimation of the variance components. The simulated powers agree with the analytical approximations obtained previously in Output 12.3: for CONDITION, analytical = 0.808, simulated = 0.860; for DIET, analytical = 0.181, simulated = 0.206; for CONDITION \times DIET, analytical = 0.522, simulated = 0.513.

12.6 Summary

In this chapter we described power calculations for mixed models, which depend on the design, the magnitudes of the variance components, the effect size in the alternative hypothesis, and the desired specificity. A simple analytical approximation to the noncentrality parameter of the F - and t -distributions provides an effective way to compute approximate power for mixed models using an exemplary data set. You can also simulate power for mixed model tests relatively easily using BY-group processing.



Some Bayesian Approaches to Mixed Models

13.1 Introduction and Background	497
13.2 P-Values and Some Alternatives	499
13.3 Bayes Factors and Posterior Probabilities of Null Hypotheses.....	502
13.4 Example: Teaching Methods	507
13.5 Generating a Sample from the Posterior Distribution with the PRIOR Statement	509
13.6 Example: Beetle Fecundity	511
13.7 Summary.....	524

13.1 Introduction and Background

Most readers are likely aware of the strong resurgence of Bayesian methods over the past two decades due largely to the development and availability of efficient algorithms for generating samples from Bayesian posterior probability distributions, such as Markov Chain Monte Carlo (Gelfand et al. 1990, Tierney 1994, Robert 1994). Computations are now less of a critical difficulty to implementing Bayesian methods; hence, debates have increased over the foundations and broader philosophical contexts of statistical inference (Swinburne 2002, Hartley 2004, D'Agostini 1998, Hively 1996). Common frequentist-based statistics like the p -value have come under renewed criticisms on various fronts (e.g., Johnson 1999, Goodman 1999a,b, Gill 2002, Gigerenzer 2004).

In this chapter we retain focus on the mixed linear model, but from a Bayesian viewpoint. This chapter differs in tone, therefore, from the others in this book and in some respects may even call into question the intent, manner, and interpretation of analyses they present. This should throw into sharp relief some of the similarities and differences between Bayesian and frequentist methods in a mixed modeling context and provide you with some examples with which you can explore these differences yourself. Some situations in which you will find Bayesian techniques especially appropriate are as follows:

- You have informative prior information about the parameters in your mixed model and you want to incorporate it directly in your analysis.
- Since p -values are very often misinterpreted, you want to compute the posterior probabilities of hypotheses, or at least bounded approximations of them.
- You need to make statistical inferences about nonstandard or nonlinear functions of mean and/or covariance parameters.
- You want to account for uncertainty of variance component estimation in the case where one or more of the likelihood-based estimates are zero.

Many of the statistical results presented in other chapters can be interpreted in a Bayesian fashion as long as you are aware of the underlying assumptions and structure of the problem. Sufficient, though not always necessary, assumptions for correspondences between frequentist and Bayesian statistical outcomes include the following:

- flat or uninformative prior distributions
- unbounded sample space
- unbounded parameter space
- symmetric likelihood function

These assumptions are reasonable in most mixed model settings, and we make them in the foregoing developments except where otherwise noted. They effectively allow us to interpret the likelihood function (from REML or ML) as a posterior distribution, which is formed by combining the likelihood function of the data with a certain form of prior distributions on the parameters. We thereby follow the classical Bayesian formulation

$$\text{Prior} \times \text{Likelihood} \propto \text{Posterior}$$

The Bayesian viewpoint can actually simplify interpretation of mixed models. Posterior probabilities and other Bayesian constructs are often more natural for experimentalists to consider than p -values or Neyman-Pearson critical regions. In addition, let us revisit the distinction between fixed and random effects. Random effects are simply those effects in the model whose parameters have a Gaussian prior distribution, whereas fixed effects have an improper flat prior equal to 1 across the real line.

Now, strict Bayesian practice states that prior distributions should be fully specified as proper probability distributions independently of the data. Under this paradigm, standard PROC MIXED analyses are not strictly Bayesian because the parameters or hyperparameters of the prior distributions of the random effects are actually estimated from the data via REML, ML, or method-of-moments. This style of analysis falls under the broad heading of *empirical Bayes* (e.g., Carlin and Louis 1996). This mild violation of rigid Bayesian principles is often the most practical way to proceed in many mixed model scenarios. Operating initially under the empirical Bayes paradigm, we first focus attention on the use of *p*-values for statistical inference, and some potential alternatives.

13.2 P-Values and Some Alternatives

The *p*-value is arguably the most used and most misused quantity in all of statistical practice. The use of *p*-values has been criticized severely by a number of authors (Goodman 1999a,b, Johnson 1999, Thompson 2001, Gill 2002, Gigerenzer 2004). They often quote Jeffreys (1961), that “a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred.” These and others argue that current practice of using *p*-values would satisfy neither Fisher nor Neyman and Pearson, due in part to the historical, virulent disagreements between them. The current use of *p*-values assumes they can help us to both (1) control long-term error rates and (2) judge whether conclusions from individual experiments are true (Berger 2003, Goodman 1999a).

In this chapter we attempt to strike a balanced view and continue to work with *p*-values as potentially valuable quantities for statistical inference in mixed model contexts, but also consider some viable alternatives. We stress it is certainly and critically important to properly interpret *p*-values. To this end, let us first review some *p*-value fundamentals.

- A *p*-value is *not* a posterior probability of the corresponding null hypothesis.
- A *p*-value is *not* the probability of falsely rejecting the null hypothesis; that is, it is not the probability of a Type I error.
- A *p*-value is the conditional probability of observing a statistic as extreme as or more extreme than the one computed from the current data, across hypothetical repetitions of the experiment, conditioned upon the following:
 - null hypothesis
 - statistical model
 - experimental design
 - sample size
 - sampling space from which observations arise

As with any conditional probability, it is imperative to consider carefully what is being conditioned upon. When the conditions change, the probability itself will also very likely change, often substantially.

Are *p*-values evidence? According to the dictionary definition, and with the common usage of “evidence” in forensic methods, *p*-values would certainly qualify as evidence, although more formal definitions are certainly possible in a statistical context (e.g., Good 1988). Smaller *p*-values convey stronger evidence, although again, as in crime scenes, context is critical for proper interpretation and weighing of available evidence. One factor that is particularly important in assessing the evidential weights of *p*-values is sample size (Royall 1986); a given *p*-value usually signifies more evidence against the tested hypothesis when the sample size is small than it does when the sample size is large. On the other hand, Neyman-Pearson hypothesis testing results usually conveys stronger evidence in a large study than in a small study.

It is therefore important for practitioners to realize that since *p*-values are conditional probabilities, they are not necessarily directly comparable across different hypotheses, models, designs, sample sizes, and sampling spaces. In other words, a *p*-value of, say, 0.01 can imply a different weight of evidence against a null hypothesis in two different experiments, or even within the same experiment when different power functions apply (e.g., a split-plot experiment). It is for this reason that we now consider some alternatives to *p*-values that are potentially less problematic.

The correspondences possible between frequentist and Bayesian results may be illustrated in the case of evaluating a one-sided hypothesis $H_0: \theta \leq 0$, based on an observed sample mean \bar{X} that, given the scalar parameter θ , is distributed as $N(\theta, \sigma^2/n)$, where σ^2 is unknown. Frequentist significance testing of H_0 proceeds by calculating the corresponding *t*-statistic, $T = t$, which, assuming $\theta = 0$, has a *t*-distribution with $n - 1$ degrees of freedom. Then the *p*-value equals $\Pr(T_r \geq t)$, where T_r is a test statistic from an imagined repetition of the experiment; namely, the *p*-value is an upper tail area. Bayesian inference about H_0 under noninformative priors, on the other hand, is that the posterior distribution of θ is a non-central *t*-distribution with $n - 1$ degrees of freedom and mean t (Box and Tiao 1973, pp. 95–97). The posterior tail area below 0 equals the *p*-value (an upper tail area), and so there is an exact correspondence in this case. Refer to Casella and Berger (1987) for a thorough explanation and more general discussion.

The situation above typifies many similar situations that are frequently handled within PROC MIXED as single-degree-of-freedom testing and estimation problems. For instance, the parameter θ may be a single element from the fixed effects parameter vector, a difference between two least-squares means, or a custom linear combination of fixed and random effects constructed with an ESTIMATE statement. In such cases, agreement between Bayesian and frequentist results is realized not only between one-sided posterior probabilities (assuming uniform priors and unbounded parameter and sample spaces) and *p*-values, but also between two-sided highest posterior density intervals and confidence intervals (e.g., Schield 1997, Perkins and Wang 2004).

In cases where the null hypothesis involves more than one degree of freedom, then it can become unwieldy to compute a simultaneous set of univariate confidence limits or multivariate confidence regions. In such situations, the classical p -values from F -tests, as computed in the “Type 3 Tests of Fixed Effects” in PROC MIXED, are often valuable results for quickly judging “what is going on” in a mixed model analysis. It is certainly then reasonable to explore indications from these p -values by looking at interesting linear combinations of the parameters via ESTIMATE and LSMEANS statements, both of which have a CL option for computing confidence limits. Also, new for PROC MIXED in SAS®9 is the LCOMPONENTS option in the MODEL statement, which prints a complete set of univariate linear combinations derived from the aforementioned F -tests. Of course you can compute additional informative F -tests via CONTRAST statements as well as LSMEANS statements with the SLICE= option.

Chapter 13 of Westfall et al. (1999) describes several useful techniques for Bayesian multiple comparisons and multiple tests. These include the following:

- simultaneous Bayesian intervals for a set of estimates or differences based (with accompanying %BayesIntervals macro)
- computation of average differences of one-sided loss functions (Waller-Duncan) for a series of estimates
- posterior probabilities of meaningful effects, in which you specify the minimum effect deemed to be scientifically “significant” and then compute posterior probabilities of exceeding that minimum for a set of hypotheses
- multiple Bayesian tests of point null hypotheses, using a prior correlation matrix and prior probabilities on the individual or joint null hypotheses (with accompanying %BayesTests macro)

The first three of these are based on generating a simulated sample from the posterior distribution, and Westfall et al. (1999) use an incomplete block example for illustration. See the final sections of this chapter for more details on posterior distribution simulation for mixed models.

Direct Bayesian-frequentist connections exist for simple-versus-simple hypothesis tests (e.g., Royall 1986, Berger and Sellke 1987); however, these are typically not useful in practice because of the composite character of most reasonable alternative hypotheses. In the much more common case of point null and composite alternative hypotheses, another approach is to view hypothesis testing as an exercise in model selection (e.g., Lahiri 2001). The basic concept is that a null hypothesis entails a model simplification of some sort, and so the idea is to fit the model under each of a series of null hypothesis and then compare model-fitting statistics. This opens a range of possibilities for considering null hypotheses that are likely true, including likelihood ratio tests and a variety of model-selection criteria like AIC, AICC, and BIC (Burnham and Anderson 2002). Such an approach is appealing but can become tedious when there are many null and alternative models to consider. A certain degree of automation is available via macros like COMPMIX, available at <http://ftp.sas.com/techsup/download/stat/>.

A related possibility is the extension of R -squared concepts (proportion of variability explained by an effect) for the mixed model. Muller and Fetterman (2002) describe the foundational issues in a univariate context. Kramer (2005) provides a recent review and advocates for an R -squared based on a likelihood ratio test. The ideas relate back to classical ANOVA, in which corrected sums of squares are partitioned into components attributable to known sources of variability in the model. A possible modern approach would be to declare all effects to be random (even ones known to be fixed) and compute REML variance component estimates for each; refer to Jin et al. (2001) for an application in a microarray context. Note that the sum of the variance components in this case would be an estimate of total variability, but this sum will generally be larger than the simple sample variance of the response because of the assumed covariance structure. Further research is needed to fully develop these ideas for rigorous mixed model applications.

Another potential idea in need of further research is the use of ordinates of sampling distributions rather than tail areas, a concept proposed in Jeffreys (1961). The ORD option in the PROC MIXED statement produces ordinates of test statistics in several output tables and is available to help you explore this style of inference.

We now investigate direct measures of the posterior probability of a null hypothesis computed under reasonable prior specifications. We discuss these in relation to a useful quantity: the Bayes factor.

13.3 Bayes Factors and Posterior Probabilities of Null Hypotheses

Berger and colleagues, although critical of p-values in papers such as Berger and Sellke (1987), have provided thoughtful commentary on best statistical practices. In particular, they advocate the approximation $-e p \log(p)$ (when $p < 1/e$) as a useful lower bound on the **Bayes factor** (Sellke, Bayarri, and Berger 2001). The Bayes factor is a generalized likelihood ratio, and is the multiplication factor for changing prior odds to posterior odds:

$$\text{Prior Odds of } H_0 \times BF_{01} = \text{Posterior Odds of } H_0$$

where

$$BF_{01} = \frac{\Pr(\text{data} | \text{model under } H_0)}{\Pr(\text{data} | \text{model under } H_a)}$$

H_0 is the null hypothesis and H_a is some appropriate alternative hypothesis. Many of the aforementioned authors and others have advocated the use of Bayes factors as alternatives to p-values; refer to Kass and Raftery (1995) and Good (1988, 1992) for

thorough overviews. Some general guidelines for interpreting Bayes factors were originally provided by Jeffreys (1961) and are along the following lines:

Bayes Factor	Strength of Evidence
0.2 = 1/5	Weak
0.1 = 1/10	Moderate
0.033 = 1/30	Strong
0.01 = 1/100	Very Strong

Instead of working extensively with Bayes factors in this chapter, we compute them as an intermediate step in approximating bounds on posterior probabilities of tested hypotheses. The word “approximate” is important here, as there are some deeper analytical challenges in making the transition that we do not discuss here. We rely heavily on the fact that the aforementioned $-e p \log(p)$ criterion provides a bound on the Bayes factor no matter what alternative is tacitly assumed or what weights may be applied to each parameter value within each hypothesis.

One important point made by Lavine and Schervish (1999) interestingly describes a common danger of interpretation shared by both Bayes factors and p -values. It is this: Neither Bayes factors nor p -values are necessarily monotone in terms of the size of the parameter space covered by a null hypotheses. For instance, if

$$BF_{ij} = \frac{\Pr(x | H_i)}{\Pr(x | H_j)}$$

for composite hypotheses H_i and H_j , $i, j = 1, 2, 3$, then possibly $BF_{12} < BF_{32}$, even though $H_3 \Rightarrow H_1$. This means that, if Bayes factors are used as test statistics, then H_1 but not H_3 could be rejected, even though $\Pr(H_1) \geq \Pr(H_3)$. Bayesians refer to such an anomaly as “incoherence.” In this sense neither the Bayes factor nor the p -value represents direct measures of credibility for a hypothesis; rather, these statistics describe how you should use the data to *change* your prior opinions or beliefs about hypotheses of interest. In other words, neither quantity directly estimates probabilities of hypotheses, but rather how such probabilities should be altered given results from a statistical model of the current data.

This viewpoint should be intuitively clear by considering extremes. First, suppose your prior experience tells you that a certain null hypothesis is almost certainly false, yet you encounter a p -value that is not small, say, $p = 0.23$. In this case you would not radically change your opinion immediately, but rather further question the design, size, model, and sampling frame. For example, the model may be inappropriate for the data, or the power of the test generating this p -value may be low because of a small sample size. Conversely, suppose you have extensive experience informing you that a certain null hypothesis is very likely true, yet you are confronted with $p = 1E-4$. Again, you would carefully recheck the conditions upon which the p -value was computed and further explore the discrepancies with what you already know or believe to be true. You would substantially modify your opinion only after very careful consideration of the new results and reconciliation with old ones.

The degree to which you should alter your prior opinions given data from a new experiment is formally described by the Bayes rule; however, the difficulty with applying this rule in practice is finding a way to fully capture all of your prior insights and intuitions into a statistical prior distribution. In most cases our prior opinions are somewhat vague and uncertain, and this uncertainty is often the *raison d'être* for conducting experiments in the first place. In this common situation it is advisable to work with noninformative priors and rely heavily on the data to help form a more informative judgment concerning new hypotheses of interest. The constant interplay between our prior knowledge and new data is how we learn and form increasingly intelligent and informed views in any discipline. Throughout the discovery and confirmation cycles, we and our colleagues undoubtedly want to reduce our uncertainties about the truth or falsehood of hypotheses. Casting our uncertainties as posterior probabilities instead of p -values can potentially help us make more accurate decisions around the cycle (Box and Tiao 1973, pp. 1–6).

Returning to the Sellke-Bayarri-Berger approximate Bayes factor lower bound $-e p \log(p)$, we can use this approximation to compute a comparable lower bound on the posterior probability of the null hypothesis of interest. As mentioned before, please keep in mind this probability depends critically upon your context and prior probabilities. Consider the following SAS code, which generates the results shown in Output 13.1 and Figure 13.1.

Program

```

data p;
do p=0.2,0.1,0.05,0.01,0.005,1e-3,1e-4,1e-5,1e-6;
  log10_p = log10(p);
  bayes_factor = -exp(1)*p*log(p);
  do prior_prob_h0 = 0.9,0.5,0.1;
    prior_odds_h0 = prior_prob_h0 / (1-prior_prob_h0);
    post_odds_h0 = prior_odds_h0 * bayes_factor;
    post_prob_h0 = post_odds_h0 / (1+post_odds_h0);
    log10_pph = log10(post_prob_h0);
    output;
  end;
end;
run;

proc print data=p;
run;

goptions reset=all
  hsize   = 5.25 in
  vsize   = 5.25 in
  htext   = 3.0 pct
  htitle  = 3.5 pct
  vorigin = 0 in
  horigin = 0 in
  ftext   = swiss
  lfactor = 1;

axis1 order=(-6 to 0)
  value=(font=swissl height=1)
  label=(font=swissl h=1)
  major=(h=0.5)
  minor=none;

```

```

axis2 order=(-6 to 0)
      value=(font=swiss1 height=1)
      label=(font=swiss1 h=1)
      major=(h=1)
      minor=none;
symbol1 i=join r=1 color=black line=1;
symbol2 i=join r=1 color=black line=2;
symbol3 i=join r=1 color=black line=25;
proc gplot;
  plot log10_pph * log10_p = prior_prob_h0 /
    haxis=axis1 vaxis=axis2;
run;

```

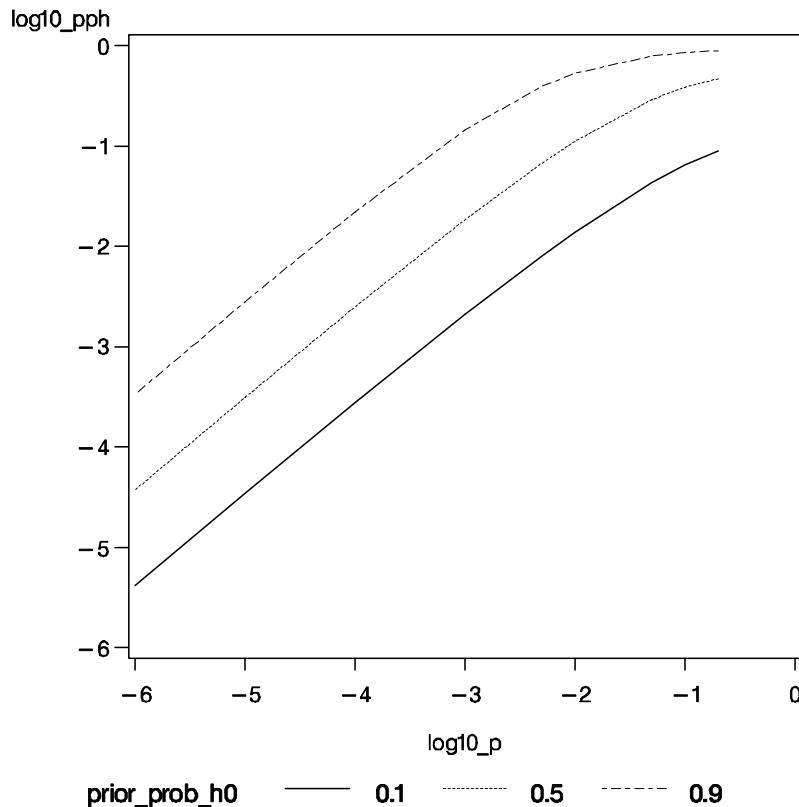
Results

Output 13.1 Comparison of *P*-Values and Various Bayesian Statistics via the Approximation of Sellke, Bayarri, and Berger (2001)

Obs	p	log10_p	bayes_factor	prior_prob_h0	prior_odds_h0	post_odds_h0	post_prob_h0	log10_pph
1	0.20000	-0.69897	0.87498	0.9	9.00000	7.87483	0.88732	-0.05192
2	0.20000	-0.69897	0.87498	0.5	1.00000	0.87498	0.46666	-0.33100
3	0.20000	-0.69897	0.87498	0.1	0.11111	0.09722	0.08861	-1.05254
4	0.10000	-1.00000	0.62591	0.9	9.00000	5.63317	0.84924	-0.07097
5	0.10000	-1.00000	0.62591	0.5	1.00000	0.62591	0.38496	-0.41459
6	0.10000	-1.00000	0.62591	0.1	0.11111	0.06955	0.06502	-1.18693
7	0.05000	-1.30103	0.40716	0.9	9.00000	3.66446	0.78561	-0.10479
8	0.05000	-1.30103	0.40716	0.5	1.00000	0.40716	0.28935	-0.53858
9	0.05000	-1.30103	0.40716	0.1	0.11111	0.04524	0.04328	-1.36369
10	0.01000	-2.00000	0.12518	0.9	9.00000	1.12663	0.52977	-0.27591
11	0.01000	-2.00000	0.12518	0.5	1.00000	0.12518	0.11125	-0.95368
12	0.01000	-2.00000	0.12518	0.1	0.11111	0.01391	0.01372	-1.86270
13	0.00500	-2.30103	0.07201	0.9	9.00000	0.64810	0.39324	-0.40534
14	0.00500	-2.30103	0.07201	0.5	1.00000	0.07201	0.06717	-1.17280
15	0.00500	-2.30103	0.07201	0.1	0.11111	0.00800	0.00794	-2.10030
16	0.00100	-3.00000	0.01878	0.9	9.00000	0.16900	0.14456	-0.83994
17	0.00100	-3.00000	0.01878	0.5	1.00000	0.01878	0.01843	-1.73445
18	0.00100	-3.00000	0.01878	0.1	0.11111	0.00209	0.00208	-2.68152
19	0.00010	-4.00000	0.00250	0.9	9.00000	0.02253	0.02204	-1.65686
20	0.00010	-4.00000	0.00250	0.5	1.00000	0.00250	0.00250	-2.60252
21	0.00010	-4.00000	0.00250	0.1	0.11111	0.00028	0.00028	-3.55579
22	0.00001	-5.00000	0.00031	0.9	9.00000	0.00282	0.00281	-2.55150
23	0.00001	-5.00000	0.00031	0.5	1.00000	0.00031	0.00031	-3.50466
24	0.00001	-5.00000	0.00031	0.1	0.11111	0.00003	0.00003	-4.45878
25	0.00000	-6.00000	0.00004	0.9	9.00000	0.00034	0.00034	-3.47124

Obs	p	log10_p	bayes_factor	prior_prob_h0	prior_odds_h0	post_odds_h0	post_prob_h0	log10_pph
26	0.00000	-6.00000	0.00004	0.5	1.00000	0.00004	0.00004	-4.42535
27	0.00000	-6.00000	0.00004	0.1	0.11111	0.00000	0.00000	-5.37958

Figure 13.1 Comparison of log10 P-Values and log10 of a Lower Bound on the Posterior Probability of the Null Hypothesis Based on the Approximation of Sellke, Bayarri, and Berger (2001)



Output 13.1 and Figure 13.1 portray the relationship between the common p -value and the lower bound on the posterior probability based on the Sellke-Bayarri-Berger approximation. In Figure 13.1, note that both axes are on a log10 scale, and the three lines correspond to prior probabilities of 0.1, 0.5, and 0.9 for H_0 . A candidate for a default “neutral” case is when H_0 is assigned a prior probability of 0.5. In this case, a general rule of thumb for determining the approximate lower bound on the posterior probability is to simply add 1 to the $\log_{10}(p\text{-value})$ when $-5 < \log_{10}(p) < -2$, and add 1.5 when $\log_{10}(p) < -5$. Interestingly, in this case the computed posterior probability can also be interpreted as the probability of a Type I error (Sellke, Bayarri, and Berger 2001).

Note the preceding lower bound depends upon the sample size of the experiment in the same way the p -value does. Experienced statisticians often develop an intuitive calibration for p -values in problems they encounter, and they recognize that a p -value of

0.001 carries a stronger weight of evidence against a null hypothesis in a small experiment than it does in a large one. In fact, Good (1992) recommends scaling p -values to a common sample size of 100 via the formula

$$p_s = \min \left\{ \frac{1}{2}, \frac{p\sqrt{n}}{10} \right\}$$

If you are in a situation where you are comparing p -values across experiments with widely different sample sizes, you may want to apply this standardization before converting to posterior probabilities.

Another way to incorporate sample size is to work directly with the test statistics from which p -values are computed. Johnson (2004) presents a useful way to calibrate F -statistics into Bayes factors and hence posterior probabilities. For an $F > 1$ with n numerator degrees of freedom and d denominator degrees of freedom,

$$\text{Bayes factor} = \left[\frac{d/n+1}{d/n+F} \right]^{(n+d)/2} F^{n/2}$$

This approximation is based on the marginal maximum likelihood estimate (MMLE) of parameters implicit to the alternative hypothesis, and so is in the same spirit as the Sellke-Bayarri-Berger lower bound computed from the p -value.

13.4 Example: Teaching Methods

We revisit the teaching methods example discussed in Chapter 7 (Section 7.7). Recall that this is a split-plot design with a covariate measured on the large-size experimental unit. The following program fits a model with unequal slopes for each gender and computes the Bayes factor and posterior probability approximations described in the previous section, producing the results shown in Output 13.2.

Program

```

proc mixed data=Ex_7_7sp covtest cl;
  class teacher met gen;
  model score = met gen gen*met y_ex y_ex*gen / ddfm=kr;
  random teacher(met);
  ods output tests3=t3;
run;

data t3_bayes;
  set t3;
  if (probft < 1/exp(1)) then bf_sbb=-exp(1)*probft*log(probft);
  else bf_sbb = 1;
  if fvalue > 1 then bf_j = ((dendf/numdf + 1) /
    (dendf/numdf + fvalue))**((numdf+dendf)/2) *
    fvalue**((numdf/2));
  else bf_j = 1;
  do prior_prob_h0 = 0.9,0.5,0.1;
    prior_odds_h0      = prior_prob_h0/(1-prior_prob_h0);
    post_odds_h0_sbb = prior_odds_h0 * bf_sbb;
    post_odds_h0_j     = prior_odds_h0 * bf_j;
    post_prob_h0_sbb = post_odds_h0_sbb/(1+post_odds_h0_sbb);
  end;

```

```

post_prob_h0_j      = post_odds_h0_j / (1+post_odds_h0_j) ;
output;
end;
run;

proc print data=t3_bayes;
id effect;
format _numeric_ best6.;
run;

```

Results

Output 13.2 shows the results, which include the analysis of variance for the remaining effects in the model.

Output 13.2 Approximate Bayes Factors and Posterior Probabilities of the Null Hypotheses for the Teaching Methods Example

Effect	NumDF	DenDF	FValue	ProbF	bf_sbb	bf_j	prior_prob_h0	prior_odds_h0
met	2	8	77.682	575E-8	0.0002	0.0001	0.9	9
met	2	8	77.682	575E-8	0.0002	0.0001	0.5	1
met	2	8	77.682	575E-8	0.0002	0.0001	0.1	0.1111
gen	1	80	5.0013	0.0281	0.2729	0.3173	0.9	9
gen	1	80	5.0013	0.0281	0.2729	0.3173	0.5	1
gen	1	80	5.0013	0.0281	0.2729	0.3173	0.1	0.1111
met*gen	2	80	63.902	26E-18	27E-16	18E-16	0.9	9
met*gen	2	80	63.902	26E-18	27E-16	18E-16	0.5	1
met*gen	2	80	63.902	26E-18	27E-16	18E-16	0.1	0.1111
y_ex	1	8	0.7532	0.4107	1	1	0.9	9
y_ex	1	8	0.7532	0.4107	1	1	0.5	1
y_ex	1	8	0.7532	0.4107	1	1	0.1	0.1111
y_ex*gen	1	80	4.2033	0.0436	0.3714	0.4262	0.9	9
y_ex*gen	1	80	4.2033	0.0436	0.3714	0.4262	0.5	1
y_ex*gen	1	80	4.2033	0.0436	0.3714	0.4262	0.1	0.1111

Effect	post_odds_h0_sbb	post_odds_h0_j	post_prob_h0_sbb	post_prob_h0_j
met	0.0017	0.0006	0.0017	0.0006
met	0.0002	0.0001	0.0002	0.0001
met	21E-6	742E-8	21E-6	742E-8
gen	2.4564	2.8556	0.7107	0.7406
gen	0.2729	0.3173	0.2144	0.2409
gen	0.0303	0.0353	0.0294	0.0341
met*gen	24E-15	16E-15	24E-15	16E-15
met*gen	27E-16	18E-16	27E-16	18E-16

Effect	post_odds_h0_sbb	post_odds_h0_j	post_prob_h0_sbb	post_prob_h0_j
met*gen	3E-16	2E-16	3E-16	2E-16
y_ex	9	9	0.9	0.9
y_ex	1	1	0.5	0.5
y_ex	0.1111	0.1111	0.1	0.1
y_ex*gen	3.3426	3.8358	0.7697	0.7932
y_ex*gen	0.3714	0.4262	0.2708	0.2988
y_ex*gen	0.0413	0.0474	0.0396	0.0452

Interpretation

This example was chosen because it produces p -values ranging from nonsignificant to very significant. The results for the Sellke-Bayarri-Berger lower bound (with suffix _sbb) and Johnson lower bound (with suffix _j) are similar for each test. We consider three prior probabilities for each null hypothesis: 0.1, 0.5, and 0.9. Of particular interest is the final test of the $y_{ex} \times gen$ effect, testing for the equality of the gender slopes for the covariate. The Bayes factor lower bounds ($bf_{sbb} = 0.37$ and $bf_j = 0.43$) are not very strong in this case, and so the resulting posterior probabilities depend greatly upon the assumed prior probability of equal slopes. For example, if you were a priori 90% sure that the slopes were equal, then the data would reduce this probability to 77–79%. On the other hand, if you were only 10% sure, then your posterior probability would drop to 4–5%.

13.5 Generating a Sample from the Posterior Distribution with the PRIOR Statement

We now move from analytical Bayesian approximations to the modern approach of simulating a random sample from the posterior distribution of all model parameters. While somewhat computationally intensive, this general methodology is very flexible and hundreds of papers in the literature show various forms and usage. We restrict attention to variance component models, building on early analytical work by Box and Tiao (1973) and Searle, Casella, and McCulloch (1992), and employing techniques described in Gelfand et al. (1990), Schervish (1992), Ghosh (1992), Wolfinger and Rosner (1996), Wolfinger (1998), Paudler, Wakefield, and Kass (1999), Westfall et al. (1999), and Wolfinger and Kass (2000).

Adding a PRIOR statement to your PROC MIXED code enables you to carry out a sampling-based Bayesian analysis of your variance component model. The analysis produces a SAS data set containing a pseudo-random sample from the joint posterior density of the variance components and fixed effects parameters in your mixed model. The posterior sample is generated after all other PROC MIXED computations, and in fact, results from the MIVQUE0, REML, or ML fit are used to help create the sample. By default, PROC MIXED uses an independence chain algorithm in order to generate the sample (Tierney 1994). This algorithm works by generating a pseudo-random proposal from a convenient base distribution, chosen to be as close as possible to the posterior.

The proposal is then retained in the sample with probability proportional to the ratio of weights constructed by taking the ratio of the true posterior to the base density. If a proposal is not accepted, then a duplicate of the previous observation is added to the chain. This algorithm is customized for mixed models and so is preferable to more general sampling schemes like Gibbs sampling or Metropolis-Hastings. It generates an independent or near-independent sample, and so no burn-in or interval sampling is necessary. In selecting the base distribution, PROC MIXED makes use of the fact that the fixed effects parameters can be analytically integrated out of the joint posterior, leaving the marginal posterior density of the variance components. In order to better approximate the marginal posterior density of the variance components, PROC MIXED transforms them using the MIVQUE(0) equations. The density of the transformed parameters is then approximated by a product of inverted gamma densities (refer to Gelfand et al. 1990). To determine the parameters for the inverted gamma densities, PROC MIXED evaluates the logarithm of the posterior density over a grid of points in each of the transformed parameters. PROC MIXED then performs a linear regression of these values on the logarithm of the inverted gamma density. At the end of the sampling, PROC MIXED displays the acceptance rate computed as the number of accepted samples divided by the total number of samples generated.

The output data set from the PRIOR statement contains the posterior sample. This data set automatically includes all variance component parameters (labeled COVP1–COVP n), Type 3 F -statistics constructed as in Ghosh (1992) discussing Schervish (1992) (labeled T3F1–T3F1 n), the log values of the posterior (labeled LOGF), the log of the base sampling density (labeled LOGG), and the log of their ratio (labeled LOGRATIO). If you specify the SOLUTION option in the MODEL statement, the data set also contains the fixed effects parameters (labeled BETA1–BETA n), and if you specify the SOLUTION option in the RANDOM statement, it contains the random effects parameters (labeled GAM1–GAM n). PROC MIXED also generates additional variables corresponding to any CONTRAST, ESTIMATE, or LSMEANS statement that you specify.

The output from the posterior sample is then explored using graphical and descriptive tools, such as SAS/INSIGHT, PROC UNIVARIATE, PROC KDE, SAS/GRAPH procedures, SAS Enterprise Guide, and JMP. The prior density of the variance components is, by default, a noninformative version of Jeffreys' prior (Box and Tiao 1973). You can also specify informative priors with the DATA= option or a flat (equal to 1) prior for the variance components. The prior density of the fixed effects parameters is assumed to be flat (equal to 1), and the resulting posterior is conditionally multivariate normal (conditioning on the variance component parameters) with mean $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and variance $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$.

Although we are considering only variance component models in this chapter, the basic techniques for constructing the sample as described above can be extended to more general mixed models like random coefficient models and repeated measures models with an unstructured covariance matrix via an inverted Wishart distribution (Daniels and Kass 2001).

13.6 Example: Beetle Fecundity

We consider an example recently discussed by Fry (2004). Quoting from that chapter:

The use of PROC MIXED for a half-sib design with one trait will be illustrated by a data set on seed beetles (data courtesy of Frank Messina, Utah State University). The data come from one of the six blocks in the “seeds present” treatment of Messina and Fry (2003). In this experiment, each of 24 sires was mated to each of four or five dams (different for each sire), and 5 female progeny from each dam were measured for two traits, mass at eclosion and lifetime fecundity.... Here, we will use just the fecundity data.

The data for this experiment appear as Data Set 13.6, “Beetle Fecundity,” in Appendix 2, “Data Sets.”

A traditional variance component model (Sokal and Rohlf 1981) applied to the fecundity data could be written as follows:

$$Y_{ijk} = \mu + S_i + D_{j(i)} + W_{k(j)}$$

Here, Y_{ijk} is an observation, μ is the population mean (a fixed effect), S_i is the random effect of the i^{th} sire, $D_{j(i)}$ is the random effect of the j^{th} dam within the i^{th} sire, and $W_{k(j)}$ is the random effect of the k^{th} individual within the j^{th} family. The three random effects are assumed to be independent and normally distributed with means zero and variances σ_S^2 , σ_D^2 , and σ_W^2 , respectively. In quantitative genetics, these three parameters are sometimes called “observational” components of variance, because they can be directly estimated by a conventional ANOVA. In contrast, we will be interested in estimating and making inferences about three “causal” components of variance, namely, the additive genetic variance V_A , the common environment or maternal-effect variance V_M , and the within-family environmental variance V_E . From quantitative genetics theory, if the sires and dams are randomly sampled from a random-mating population, and if certain other assumptions are met, then the observational components have the following interpretations in terms of the causal components (Falconer and Mackay 1996):

$$\begin{aligned}\sigma_S^2 &= \frac{1}{4}V_A \\ \sigma_D^2 &= \frac{1}{4}V_A + \frac{1}{4}V_D + V_M \\ \sigma_W^2 &= \frac{1}{2}V_A + \frac{3}{4}V_D + V_E\end{aligned}$$

For simplicity, we will assume $V_D = 0$. In this case, we can solve for the causal components to obtain the following:

$$\begin{aligned}V_A &= 4\sigma_S^2 \\ V_M &= \sigma_D^2 - \sigma_S^2 \\ V_E &= \sigma_W^2 - 2\sigma_S^2\end{aligned}$$

The following program reads the data, performs a basic variance components analysis, and produces the results shown in Output 13.3.

Program

```
proc mixed data=beetle covtest cl asycov;
  where trait = 'fec';
  class sire dam;
  model value = / solution;
  random sire dam(sire);
run;
```

Results

Output 13.3 Results from a Standard Variance Components Analysis for the Beetle Fecundity Example

Model Information	
Data Set	WORK.BEETLE
Dependent Variable	value
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
sire	24	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
dam	5	1 2 3 4 5

Dimensions	
Covariance Parameters	3
Columns in X	1
Columns in Z	134
Subjects	1
Max Obs Per Subject	549

Number of Observations	
Number of Observations Read	549
Number of Observations Used	549
Number of Observations Not Used	0

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	4725.58171669	
1	2	4572.06799839	0.00000000

Convergence criteria met.

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
sire	31.7912	19.8135	1.60	0.0543	0.05	12.5172	184.27
dam(sire)	114.61	23.0327	4.98	<.0001	0.05	80.1138	177.52
Residual	177.49	11.9787	14.82	<.0001	0.05	156.17	203.52

Asymptotic Covariance Matrix of Estimates					
Row	Cov Parm	CovP1	CovP2	CovP3	
1	sire	392.58	-116.52	-0.01690	
2	dam(sire)	-116.52	530.50	-28.6296	
3	Residual	-0.01690	-28.6296	143.49	

Fit Statistics	
-2 Res Log Likelihood	4572.1
AIC (smaller is better)	4578.1
AICC (smaller is better)	4578.1
BIC (smaller is better)	4581.6

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	55.4665	1.6425	23	33.77	<.0001

Interpretation

Output 13.3 is the same as that from Fry (2004). Point estimates of the causal components are as follows:

$$V_A = 4\sigma_S^2 = 4 \times 31.79 = 127.16$$

$$V_M = \sigma_D^2 - \sigma_S^2 = 114.61 - 31.79 = 82.82$$

$$V_E = \sigma_W^2 - 2\sigma_S^2 = 177.49 - 2 \times 31.79 = 113.91$$

Fry (2004) goes on to show how to derive approximate standard errors for these estimates based on the “Asymptotic Covariance Matrix of Estimates” table. However, these standard errors can often be biased because they are based on large-sample normal theory, whereas the small-sample distribution of the variance components is skewed right based on chi-square distributions.

Instead, we use a Bayesian sampling approach to directly construct a sample from the joint posterior distribution of the causal components. To accomplish this, simply add a PRIOR statement to the preceding code as follows.

Program

```
proc mixed data=beetle covtest cl asycov;
  where trait = 'fec';
  class sire dam;
  model value = / solution;
  random sire dam(sire);
  prior / nsample=1e4 seed=3102834 out=post
    psearch ptrans;
run;
```

The PRIOR statement requests a Bayesian sample of size 10,000 to be saved in the SAS data set POST. The SEED= option fixes the random number seed so that results can be duplicated exactly. The PSEARCH and PTRANS options produce the extra output tables explained below. This code produces the same output as in Output 13.3, plus the tables shown in Output 13.4.

Results

Output 13.4 Results from a Bayesian Sampling Analysis for the Beetle Fecundity Example

Posterior Sampling Information	
Prior	Jeffreys
Algorithm	Independence Chain
Sample Size	10000
Seed	3102834

Transformation for Covariance Parameters			
TCovP	CovP1	CovP2	CovP3
1	23.113	4.9927	1
2	0	4.9927	1
3	0	0	1

Transformed Parameter Search				
Parm	TCovP1	TCovP2	TCovP3	Log Posterior
1	859.46	749.72	177.49	-2296.8252
1	976.66	749.72	177.49	-2296.0643
1	1093.9	749.72	177.49	-2295.623
1	1211.1	749.72	177.49	-2295.3934
1	1328.3	749.72	177.49	-2295.3084
1	1445.5	749.72	177.49	-2295.3247
1	1562.7	749.72	177.49	-2295.413
1	1679.8	749.72	177.49	-2295.5532
1	1797	749.72	177.49	-2295.7309
1	1914.2	749.72	177.49	-2295.9361
1	2031.4	749.72	177.49	-2296.1611
1	2148.6	749.72	177.49	-2296.4003
1	2265.8	749.72	177.49	-2296.6497
1	2383	749.72	177.49	-2296.9059
1	2500.2	749.72	177.49	-2297.1667
1	2617.4	749.72	177.49	-2297.4301
1	2734.6	749.72	177.49	-2297.6947
1	2851.8	749.72	177.49	-2297.9594
1	2969	749.72	177.49	-2298.2233
2	1484.5	374.86	177.49	-2307.8435
2	1484.5	434.05	177.49	-2302.5682
2	1484.5	493.23	177.49	-2299.2817
2	1484.5	552.42	177.49	-2297.2664
2	1484.5	611.61	177.49	-2296.0982
2	1484.5	670.8	177.49	-2295.5122
2	1484.5	729.99	177.49	-2295.3361
2	1484.5	789.17	177.49	-2295.454
2	1484.5	848.36	177.49	-2295.7858
2	1484.5	907.55	177.49	-2296.2743
2	1484.5	966.74	177.49	-2296.8787
2	1484.5	1025.9	177.49	-2297.5687
2	1484.5	1085.1	177.49	-2298.3219
2	1484.5	1144.3	177.49	-2299.1212
2	1484.5	1203.5	177.49	-2299.9538
2	1484.5	1262.7	177.49	-2300.8098
2	1484.5	1321.9	177.49	-2301.6814
2	1484.5	1381.1	177.49	-2302.5628
2	1484.5	1440.2	177.49	-2303.4492

Transformed Parameter Search				
Parm	TCovP1	TCovP2	TCovP3	Log Posterior
3	1484.5	749.72	88.744	-2362.0154
3	1484.5	749.72	102.76	-2334.4746
3	1484.5	749.72	116.77	-2317.1635
3	1484.5	749.72	130.78	-2306.4041
3	1484.5	749.72	144.79	-2300.0182
3	1484.5	749.72	158.81	-2296.6453
3	1484.5	749.72	172.82	-2295.3993
3	1484.5	749.72	186.83	-2295.6827
3	1484.5	749.72	200.84	-2297.0818
3	1484.5	749.72	214.85	-2299.3026
3	1484.5	749.72	228.87	-2302.1328
3	1484.5	749.72	242.88	-2305.416
3	1484.5	749.72	256.89	-2309.0355
3	1484.5	749.72	270.9	-2312.9031
3	1484.5	749.72	284.92	-2316.9516
3	1484.5	749.72	298.93	-2321.1293
3	1484.5	749.72	312.94	-2325.396
3	1484.5	749.72	326.95	-2329.7203
3	1484.5	749.72	340.96	-2334.0779
3	1484.5	749.72	354.98	-2338.4494

Base Densities			
Density	Type	Parm1	Parm2
1	ig	11.265	16709
2	ig	42.967	32216
3	ig	219.55	38966

Acceptance Rates	
Boundary Constraints	Sampling
0.99	1.00

Interpretation

In Output 13.4, the first table echoes specifications. Here a Jeffreys noninformative prior is used for the variance components. This prior has the advantage of being invariant to transformations of the parameters. For a data set of this size, it produces little change on the likelihood function.

The “Transformation for Covariance Parameters” table is produced by the PTRANS option and lists the transformation that is used to create approximately independent sampling parameters. Note the residual variance is not transformed, and the other two variance components are transformed in a way similar to that used to create expected mean squares. As indicated above, the purpose of this transformation is to create approximately independent parameters so that the sampling can be performed from univariate inverted gamma densities.

The “Transformed Parameter Search” table is produced by the PSEARCH option. This search is used to determine the parameters for the three independent inverted gamma densities that are used to sample from the posterior distribution. The “Base Densities” table then lists the inverted gamma parameters estimated from this search.

The “Acceptance Rates” table lists the rates that each independent draw was retained in the final sample. 100% of the initial samples were accepted, but 1% of these were dropped because they violated the nonnegativity constraint on the original variance components.

You can construct the causal components of interest by using the following code. Note this code deletes observations that produce negative estimates of the maternal or environmental variance. The code also computes the percent of total variance explained by each causal component. As mentioned previously, you can use several different SAS modules to explore the posterior sample, including SAS/INSIGHT, PROC UNIVARIATE, PROC KDE, SAS/GRAFH, and SAS Enterprise Guide. Here we use the UNIVARIATE procedure to compute the 2.5 and 97.5 percentiles of the distribution (the approximate 95% confidence limits) and the median (50th percentile).

Program

```

data post;
  set post;
  Var_Additive      = 4*covp1;
  Var_Maternal      = covp2 - covp1;
  Var_Environmental = covp3 - 2*covp1;
  if Var_Maternal < 0 or Var_Environmental < 0 then delete;
  Var_Total          = Var_Additive + Var_Maternal +
                        Var_Environmental;
  Percent_Var_Additive = 100*Var_Additive/Var_Total;
  Percent_Var_Maternal = 100*Var_Maternal/Var_Total;
  Percent_Var_Environmental = 100*Var_Environmental/Var_Total;
  rename covp1 = Var_Sire
        covp2 = Var_Dam_Sire
        covp3 = Var_Residual
        beta1 = Intercept;
run;

```

```

proc univariate data=post noprint;
var Var_Sire Var_Dam_Sire Var_Residual Intercept;
output out=confInt pctlpts=2.5,50,97.5
        pctlpre = Sire_DamSire_Residual_Intercept_
        pctlname= lower Median upper
        mean    = Sire_mean DamSire_mean
                  Residual_Mean Intercept_Mean
        std     = Sire_Std DamSire_Std Residual_Std
                  Intercept_Std;
run;
proc print data=confInt label noobs; var sire: ;
proc print data=confInt label noobs; var damsire: ;
proc print data=confInt label noobs; var residual: ;
proc print data=confInt label noobs; var intercept:;
run;

```

Results

Output 13.5 displays descriptive statistics about the distributions for the three variance components and the intercept.

Output 13.5 Posterior Distribution Results of the Three Variance Components and Overall Intercept for the Beetle Fecundity Example

the mean, Var_Sire	the standard deviation, Var_Sire	the 2.5000 percentile, Var_Sire	the 50.0000 percentile, Var_Sire	the 97.5000 percentile, Var_Sire
35.0545	18.9152	4.81942	32.6380	77.7916

the mean, Var_Dam_Sire	the standard deviation, Var_Dam_Sire	the 2.5000 percentile, Var_Dam_Sire	the 50.0000 percentile, Var_Dam_Sire	the 97.5000 percentile, Var_Dam_Sire
117.813	23.2714	78.0940	115.755	169.550

the mean, Var_Residual	the standard deviation, Var_Residual	the 2.5000 percentile, Var_Residual	the 50.0000 percentile, Var_Residual	the 97.5000 percentile, Var_Residual
178.530	12.1389	156.402	178.037	204.215

the mean, Intercept	the standard deviation, Intercept	the 2.5000 percentile, Intercept	the 50.0000 percentile, Intercept	the 97.5000 percentile, Intercept
55.4623	1.69218	52.1191	55.4685	58.8140

In Output 13.5, note that the Bayesian sampling-based posterior median, mean, and standard deviation of each parameter agrees well with the REML estimates shown in Output 13.3.

However, the approximate 95% confidence interval for the sire variance component is (12.5, 184.3) from Output 13.3, but the corresponding Bayesian posterior interval is somewhat narrower: (4.82, 77.79). The latter is obtained by simply reading the 2.5% and 97.5% quantiles

from Output 13.5. The intervals for the other two variance components are in better agreement. The posterior distribution for the intercept agrees very closely with the frequentist-based results since they are both t -distributions.

Similar code as previously describes the distributions of the causal components.

Program

```
proc univariate data=post noprint;
  var Var_Additive Var_Maternal Var_Environmental;
  output out=confInt pctlpts=2.5,50,97.5
          pctlpre=Additive_ Maternal_ Environ_
          pctlname=lower Median upper
          mean=Additive_mean Maternal_mean Environ_Mean
          std=Additive_Std Maternal_Std Environ_Std;
run;
proc print data=confInt label noobs; var Additive:;
proc print data=confInt label noobs; var Maternal:;
proc print data=confInt label noobs; var Environ:;
run;
```

Results

The results appear in Output 13.6.

Output 13.6 Posterior Distributions of the Three Causal Components for the Beetle Fecundity Example

the mean, Var_Additive	the standard deviation, Var_Additive	the 2.5000 percentile, Var_Additive	the 50.0000 percentile, Var_Additive	the 97.5000 percentile, Var_Additive
140.218	75.6608	19.2777	130.552	311.167

the mean, Var_Maternal	the standard deviation, Var_Maternal	the 2.5000 percentile, Var_Maternal	the 50.0000 percentile, Var_Maternal	the 97.5000 percentile, Var_Maternal
82.7585	32.7962	21.3500	81.7124	151.004

the mean, Var_Environment al	the standard deviation, Var_Environment a	the 2.5000 percentile, Var_Environment	the 50.0000 percentile, Var_Environment	the 97.5000 percentile, Var_Environment
108.421	39.6220	21.0155	112.293	174.641

Interpretation

Output 13.6 displays information about the marginal posterior distributions of the three causal components. Their posterior medians and means agree fairly well with the REML-based point estimates computed after Output 13.3. It is interesting to compare the posterior standard deviations obtained here with the approximate ones computed in Fry (2004). Table 13.1 shows that for this example they are in good agreement, with the Bayesian results uniformly smaller by a small percentage. The Bayesian results are

more informative in the sense of being able to easily compute any function or moment of interest as a statistic from the posterior sample.

Table 13.1 Comparison of Asymptotic and Bayesian Standard Errors for the Causal Components from the Beetle Data

Component	Asymptotic SE (Fry, 2004)	Bayesian Sample Std Dev
Additive	79.2	75.7
Maternal	34.0	32.8
Environmental	41.4	39.6

Bayesian sampling also makes it easy for you to explore multivariate relationships between parameter estimates. Output 13.7 and Figure 13.2 display results about the bivariate relationships between the three causal components, obtained with the following program.

Program

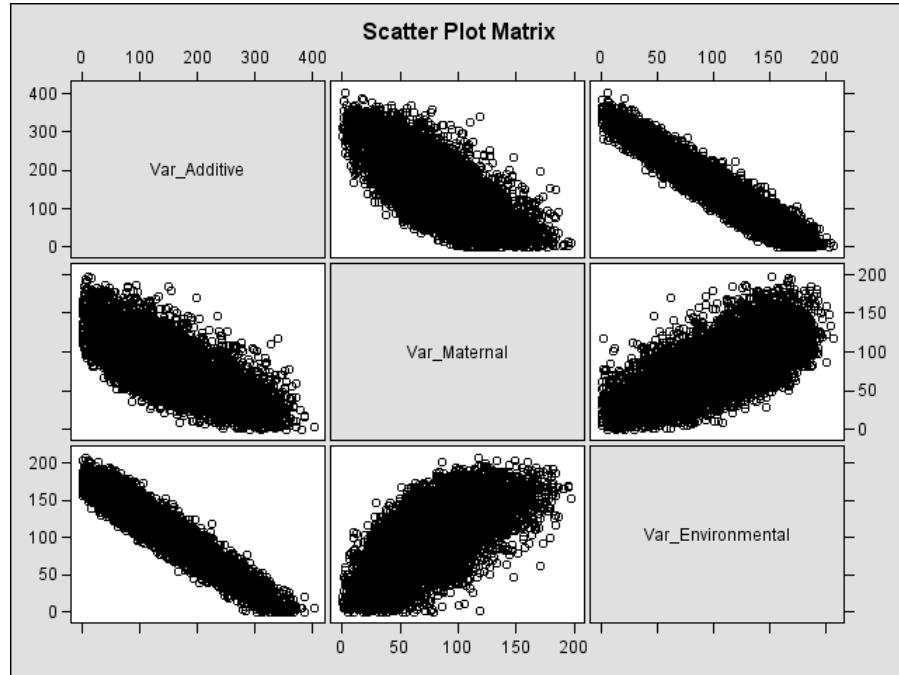
```
ods html;
ods graphics on;
proc corr data=post plots=matrix;
  var Var_Additive Var_Maternal Var_Environmental;
run;
ods graphics off;
ods html close;
```

Results

Output 13.7 Bivariate Relationships of the Three Causal Components for the Beetle Fecundity Example

Pearson Correlation Coefficients, N = 9617 Prob > r under H0: Rho=0			
	Var_Additive	Var_Maternal	Var_Environmental
Var_Additive	1.00000	-0.71881 <.0001	-0.95192 <.0001
Var_Maternal	-0.71881 <.0001	1.00000	0.66587 <.0001
Var_Environmental	-0.95192 <.0001	0.66587 <.0001	1.00000

Figure 13.2 Bivariate Relationships of the Three Causal Components for the Beetle Fecundity Example



Interpretation

It is immediately clear that the additive component is strongly negatively correlated with the other two components (-0.72 with maternal and -0.95 with environmental).

Output 13.8 shows statistics regarding the marginal posterior distributions of the percent variability explained by each component. The distributions are very similar to those of the causal components themselves shown in Output 13.6, but there is a bit more spread due to division by the total variance. Point estimates for percent variability are approximately 40% additive, 25% maternal, and 35% environmental, but there is considerable uncertainty surrounding these estimates.

Output 13.8 Posterior Distributions of Percent Variability Explained by the Three Causal Components for the Beetle Fecundity Example

the mean, Percent_Var_Ad ditive	the standard deviation, Percent_Var_Ad di	the 2.5000 percentile, Percent_Var_Ad d	the 50.0000 percentile, Percent_Var_Ad d	the 97.5000 percentile, Percent_Var_Ad d
41.6631	20.7439	6.10699	39.9593	85.5509

the mean, Percent_Var_Ma ternal	the standard deviation, Percent_Var_Ma te	the 2.5000 percentile, Percent_Var_Ma t	the 50.0000 percentile, Percent_Var_Ma t	the 97.5000 percentile, Percent_Var_Ma t
25.0125	9.56674	6.17635	25.1084	43.2337

the mean, Percent_Var_En vironmental	the standard deviation, Percent_Var_En vi	the 2.5000 percentile, Percent_Var_En v	the 50.0000 percentile, Percent_Var_En v	the 97.5000 percentile, Percent_Var_En v
33.3244	12.9753	5.72254	34.3798	55.3018

This example should provide you with a basic feel for a few standard analyses possible with the PRIOR statement. Space prohibits more detailed results here. Chapter 13 of Westfall et al. (1999) provides some additional discussion, with a special focus on multiple comparisons and multiple testing issues. Rosa and Tempelman (2004) provide additional discussion of genetic concepts along with an in-depth treatment and SAS/IML code for simulating from Bayesian posterior distributions for genetic linkage and quantitative trait loci (QTL) analyses.

We conclude this example by showing how to incorporate informative prior information. For variance component models like the one here, the PRIOR statement requires the prior to be specified in terms of independent inverted gamma densities on the *transformed* parameters, not on the original variance components. So for this example, we specify priors on the following:

$$\text{TCovP1} = 23.113 \text{ Var(Sire)} + 4.9927 \text{ Var(Dam(Sire))} + \text{Var(Residual)}$$

$$\text{TCovP2} = 4.9927 \text{ Var(Dam(Sire))} + \text{Var(Residual)}$$

$$\text{TCovP3} = \text{Var(Residual)}$$

Again, the transformation is used so that the priors for each parameter are approximately independent. For sake of illustration, we assume the prior is exactly the same as the posterior from the same data, and so we effectively double the information on the variance components.

The following SAS code sets up the priors in the data set PV and then specifies this data set in the PRIOR statement. The code produces the results shown in Output 13.9.

Program

```

data pv;
  input type $ Parm1 Parm2;
  datalines;
  ig 11.265 16709
  ig 42.967 32216
  ig 219.55 38966
run;

proc mixed data=beetle covtest cl asycov;
  where trait = 'fec';
  class sire dam;

```

```

model value = / solution;
random sire dam(sire);
prior data=pv / nsample=1e4 seed=3102834 out=post1
               psearch ptrans;
run;

data post1;
  set post1;
  Var_Additive      = 4*covp1;
  Var_Maternal      = covp2 - covp1;
  Var_Environmental = covp3 - 2*covp1;
  if Var_Maternal < 0 or Var_Environmental < 0 then delete;
  Var_Total = Var_Additive + Var_Maternal + Var_Environmental;
  Percent_Var_Additive      = 100*Var_Additive/Var_Total;
  Percent_Var_Maternal      = 100*Var_Maternal/Var_Total;
  Percent_Var_Environmental = 100*Var_Environmental/Var_Total;
  rename covp1 = Var_Sire
         covp2 = Var_Dam_Sire
         covp3 = Var_Residual
         beta1 = Intercept;
run;

proc means data= post1 mean std;
run;

```

Results

Output 13.9 Results after Using Informative Priors on the Variance Components for the Beetle Fecundity Example

Variable	Mean	Std Dev
Var_Sire	34.0972410	14.8136930
Var_Dam_Sire	116.3058820	16.5544867
Var_Residual	178.0940800	8.5160790
Intercept	55.4663669	1.6711953
logf	-3990.33	1.2319004
logg	-3387.97	1.2323084
logratio	-602.3597075	0.0036638
sample	4999.00	2887.12
Var_Additive	136.3889641	59.2547719
Var_Maternal	82.2086410	24.7087838
Var_Environmental	109.8995980	30.8084501
Var_Total	328.4972030	20.5470599
Percent_Var_Additive	41.0965422	16.4546891
Percent_Var_Maternal	25.0826584	7.4240690
Percent_Var_Environmental	33.8207994	10.2123180

Note from Output 13.9 that the posterior means of the variance components and functions of them are all nearly the same as those from the original data, whereas their standard deviations are significantly smaller. The posterior distribution for the Intercept remains the same because its standard deviation is dependent upon the means of the variance components, which have not changed from before.

13.7 Summary

In this chapter we considered several ways you can adopt a Bayesian approach to mixed model analysis. Section 13.2 discussed some of the controversies surrounding p -values and their proper conditional interpretation. Section 13.3 discussed computation of lower bounds on Bayes factors and posterior probabilities of null hypotheses based on either a p -value or an F -statistic. Section 13.4 applied these techniques to the teaching methods example from Chapter 7. Section 13.5 discussed generating a random sample from the posterior distribution of a variance components model and analyzing the results using exploratory data analysis techniques. Section 13.6 applied this approach to a beetle fecundity example.



Generalized Linear Mixed Models

14.1 Introduction	526
14.2 Two Examples to Illustrate When Generalized Linear Mixed Models Are Needed	527
14.2.1 Binomial Data in a Multi-center Clinical Trial.....	527
14.2.2 Count Data in a Split-Plot Design	528
14.3 Generalized Linear Model Background.....	529
14.3.1 Fixed Effects Generalized Linear Models	529
14.3.2 Probability Distributions	531
14.3.3 Link Functions and Variance Structure.....	534
14.3.4 Predicting Means from the Inverse Link Function.....	535
14.3.5 Deviance	536
14.3.6 Inference Using Estimable Functions	537
14.4 From GLMs to GLMMs	538
14.4.1 Incorporating Random Effects	538
14.4.2 The Pseudo-likelihood Approach.....	538
14.4.3 The Role of the Scale Parameter.....	540
14.4.4 Estimation Methods in PROC GLIMMIX	541
14.5 Example: Binomial Data in a Multi-center Clinical Trial	542
14.5.1 Analysis of Binomial Data Using Logit (Canonical) Link.....	542
14.5.2 Obtaining Predicted Values	552
14.5.3 Alternative Link Functions	553
14.6 Example: Count Data in a Split-Plot Design.....	557
14.6.1 Standard Analysis of Poisson GLMM	557
14.6.2 Analysis of Poisson GLMM with Overdispersion.....	561
14.7 Summary.....	566

14.1 Introduction

The linear mixed models discussed in the previous chapters have two defining features in common. First, the errors and random effects are assumed to be normally distributed. Second, the response variable is modeled directly as a linear combination of fixed and random effects. However, in many practical situations, the response variable of interest may not have a normal distribution. In other cases, there may be restrictions on the range of allowable values for predictable functions that direct modeling of the response variable cannot address. One approach to handling nonnormality is to use various data transformations in conjunction with standard linear model methods. The more commonly used transformations are discussed in introductory statistical methods texts such as Steel and Torrie (1980) and Snedecor and Cochran (1989). Box and Cox (1964) discuss such transformations in considerable depth.

For models without random effects, Nelder and Wedderburn (1972) present a comprehensive alternative—the **generalized linear model**. In this class of models the mean of the data is expressed in terms of a monotonic transform of a linear model and the parameters are estimated by maximum likelihood. Originally developed for members of the exponential family of distributions, generalized linear models have been extended to a much broader range of applications. For example, **quasi-likelihood** methods for independent data enable statistical estimation and inference based on only the mean and variance model of the data, without the specification of a joint distribution. Similarly, likelihood-based techniques have been developed for cases where the joint distribution of the data is unknown or difficult to determine, but a conditional distribution can be specified. This scenario is important for developments discussed below, where the distribution of the data is known conditional on random effects. The two examples in the following section illustrate models of this type.

In practice, generalized linear model estimation strongly resembles generalized least squares, the fixed effects component of the mixed model procedure. In the 1990s, a number of articles extended the generalized linear model to various special-purpose mixed model applications. Breslow and Clayton (1993) presented a unifying approach tying together the underlying principles of the **generalized linear mixed model (GLMM)**. Independently, Wolfinger and O'Connell (1993) developed a similar methodology for GLMMs and implemented it in the GLIMMIX macro. Recently, the functionality of this macro has been enhanced and incorporated as a SAS/STAT procedure, PROC GLIMMIX.¹

The purpose of this chapter is to introduce the basic concepts of the GLMM and to show how to use PROC GLIMMIX to fit basic GLMMs. Section 14.2 presents two examples illustrating when generalized linear mixed models are of interest. Section 14.3 is an introduction to generalized linear models. Section 14.4 shows how random effects are incorporated to form generalized linear **mixed** models. Sections 14.5 and 14.6 present the GLIMMIX procedure statements and associated output for the two examples described in Section 14.2.

¹ The GLIMMIX procedure is an add-on in SAS 9.1 to SAS/STAT for the (32-bit) Windows platform. It does not ship with SAS 9.1. You can obtain the GLIMMIX procedure for SAS 9.1 as a download from www.sas.com/statistics. This site also contains the documentation for the GLIMMIX procedure.

14.2 Two Examples to Illustrate When Generalized Linear Mixed Models Are Needed

As mentioned above, many practical applications call for mixed model methods in circumstances where the assumptions of the standard mixed linear model do not hold. For example, these assumptions do not hold when the distribution of the data—given the random effects—is not normal, when the mean is nonlinearly related to the model parameters, or when the variances of the data are related to the mean. This section presents two common examples to illustrate when generalized linear mixed models are needed.

14.2.1 Binomial Data in a Multi-center Clinical Trial

Beitler and Landis (1985) discuss data from a clinical trial involving 2 drug treatments and 8 clinics. The 8 clinics represented a sample from a larger target population. At the j^{th} clinic ($j=1,2,\dots,8$), n_{1j} subjects were randomly assigned to receive treatment 1, and n_{2j} subjects were randomly assigned to treatment 2. Each subject was classified as having a “favorable” or “unfavorable” response to the treatment. The number of “favorable” responses to the i^{th} treatment at the j^{th} clinic was denoted y_{ij} , and the sample proportion of “favorable” responses for the ij^{th} treatment-clinic combination was denoted $p_{ij} = Y_{ij}/n_{ij}$.

The objective of the trial was to determine the effect of treatment and clinic on the probability of a “favorable” response. If we denote the probability of having a “favorable” response for the ij^{th} treatment-clinic combination as π_{ij} , then a possible linear model for this trial is

$$\begin{aligned}\pi_{ij} &= \mu + \tau_i + c_j + (\tau c)_{ij} \\ p_{ij} &= \pi_{ij} + e_{ij}\end{aligned}\tag{14.1}$$

where

- μ is the overall mean
- τ_i is the i^{th} treatment effect
- c_j is the j^{th} clinic effect
- $(\tau c)_{ij}$ is the ij^{th} treatment-by-clinic interaction
- e_{ij} is error

Because clinics were sampled from the population, c_j and $(\tau c)_{ij}$ are considered random effects.

There are three serious problems with using conventional normal linear modeling methods with model 14.1. First, the estimates of $(\tau c)_{ij}$ and e_{ij} are confounded. Confounding occurs because there is only one p_{ij} per treatment-clinic combination. This precludes evaluating the clinic-by-treatment interaction, which, if it exists, is of interest. Second, the data are binomial, violating standard linear mixed model assumptions, e.g., normality and homogeneity of variance. Finally, although the probability of a favorable response clearly must be between 0 and 1, predicted probabilities using a linear model are not similarly restricted. A predicted probability that is negative or greater than one is entirely possible.

Beitler and Landis (1985) present a procedure for analyzing these data that addresses the confounding of $(\tau c)_{ij}$ and e_{ij} and the normality issues. Their method anticipates the generalized linear mixed model, although they did not develop their approach within that framework. Also, their model did not address the need for restrictions on predicted probabilities. Section 14.5 presents a generalized linear mixed model for these data and shows how to use the GLIMMIX procedure for the analysis.

14.2.2 Count Data in a Split-Plot Design

A split-plot experiment was conducted to compare various treatments for improving damaged rangeland. The whole-plot treatments were various management methods. They were applied in randomized complete blocks. Each whole plot was split into 4 split plots; the split-plot treatments were different seed mixes. The response variable of interest was botanical composition, measured by the number of plants of various species present in a given plot.

This experiment is similar to the split-plot experiments described in Chapter 2. The model is

$$\begin{aligned}\mu_{ijk} &= R_i + \tau_j + (R\tau)_{ij} + \delta_k + (\tau\delta)_{jk} \\ Y_{ijk} &= \mu_{ijk} + e_{ijk}\end{aligned}\tag{14.2}$$

where

μ_{ijk} is the expected number of plants of a given species in the ijk^{th} block-management method-seed mix combination

Y_{ijk} is the observed number of plants of a given species in the ijk^{th} block-management method-seed mix combination

μ is the overall mean or intercept

R_i is the i^{th} block effect

τ_j is the j^{th} management method effect

$(R\tau)_{ij}$ is the whole-plot error (equivalent to the block-management interaction)

δ_k is the effect of the k^{th} seed mix

$(\tau\delta)_{jk}$ is the management-seed mix interaction

e_{ijk} is the split-plot error

The whole-plot errors, $(R\tau)_{ij}$, are clearly random effects. The block effects, R_i , are typically considered random as well.

Plausible assumptions are as follows:

- $R_i \sim iid N(0, \sigma^2_R)$
- $(R\tau)_{ij} \sim iid N(0, \sigma^2_{RM})$

In a standard split-plot model, you would also assume that $e_{ijk} \sim iid N(0, \sigma^2)$, and, as a consequence, Y_{ijk} must be normally distributed. However, because the ijk^{th} observation is a count, the e_{ijk} 's are not normally distributed. Moreover, a count cannot be negative, but negative predicted counts from the model above are possible.

Traditionally, counts have been presumed to follow a Poisson distribution. However, the Poisson distribution assumes that the mean and variance are equal. Recent work in a number of disciplines suggests that count data are typically **overdispersed**; that is, the variance is larger—often considerably larger—than the mean. Alternatives to the Poisson, such as the negative binomial distribution, often provide better models of variation among the counts and hence of the errors in a linear model.

The question is how to simultaneously take into account the random model effects and the nonnormally distributed errors. For count data, one approach is to analyze the square root of Y_{ijk} , or the square root of $Y_{ijk} + 1$ if zeros are prevalent. See, for example, Snedecor and Cochran (1989). However, transforming the data may be counterproductive. For example, a transformation can distort the distribution of the random model effects or the linearity of the model. More importantly, transforming the data still leaves open the possibility of negative predicted counts. Consequently, inference from a mixed model using transformed data is highly suspect.

Section 14.6 shows how to construct a generalized linear mixed model that is consistent with the data from this experiment and how to use the GLIMMIX procedure to perform the analysis.

14.3 Generalized Linear Model Background

Generalized linear models are extensions of fixed effects linear models to cases where data are independent and standard linear model assumptions are violated. This section presents an introduction to generalized linear models.

A complete introduction to the generalized linear model, with all the theoretical development and technical details, is beyond the scope of this book. The goal here is to present the main ideas for fixed effects models so that (1) the methods presented in Sections 14.5 and 14.6 are accessible and (2) the available statements and options of the GLIMMIX procedure have an adequate context. References containing additional detail are provided throughout this section.

Section 14.4 shows how random effects are incorporated to produce *generalized linear mixed models*.

14.3.1 Fixed Effects Generalized Linear Models

Standard linear models are defined directly in terms of the observations. For example, the fixed effects linear model, assuming normal errors, is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (14.3)$$

Another way to present this model, which is helpful in making the transition to generalized linear models, is in terms of the distribution of the data, means, and variances. Model (14.3) can be expressed equivalently by stating that \mathbf{Y} has a normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\sigma^2 \mathbf{I}$, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$, $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, $\text{Var}[\mathbf{Y}] = \sigma^2 \mathbf{I}$.

In this model, $E[\mathbf{Y}]$, the expected values of the observations, is a linear combination of the model fixed effects $\boldsymbol{\beta}$. The primary functions of $\mathbf{X}\boldsymbol{\beta}$ are to provide, for example, predicted values of $E[\mathbf{Y}]$ for a given \mathbf{X} , as in regression analysis, or comparisons among the $E[\mathbf{Y}]$ for various \mathbf{X} , as in analysis of variance where functions of the $E[\mathbf{Y}]$ represent treatment means or other estimable functions of interest. Note that we did not specify explicitly the nature of the variance of the data. The form of the variance is given by the choice of the distribution. For

example, if we specify that \mathbf{Y} has a Poisson distribution with mean μ , then it is determined that $\text{Var}[\mathbf{Y}] = \mu$.

Standard linear model methods—that is, models that apply least-squares procedures to direct models of \mathbf{Y} —are based on the assumption of normal errors. Standard methods include ordinary least squares, generalized least squares, and the mixed model equations. As noted in Section 14.1, although linear models are extremely useful for nonnormal data as well as for normal data, standard methods do not necessarily produce usable results. Problems include inefficient or inaccurate estimates, estimates outside the range of permissible values, misleading p -values in hypothesis tests, and so on.

The main idea of the *generalized* linear model is to specify the model as above in terms of the distribution of the data and the mean function, and to express the mean in terms of some transformation of $\mathbf{X}\beta$. Specifically,

- the distribution—also termed the **random component** of the model—is chosen as a member of the exponential family of distributions (see Section 14.3.2)
- the relationship between the mean $\mu = E[\mathbf{Y}]$ and the **linear predictor** $\mathbf{x}'\beta$ must follow a monotonic transform

It is common to denote the linear predictor as $\mathbf{x}'\beta = \eta$ and the transformation of the mean as $g(\mu)$. Nelder and Wedderburn (1972) called $g(\mu)$ the **link function** because it links the linear model to the mean of Y . Because the link function is monotonic, the relationship between μ and η (and thus between μ and β) can be expressed in terms of the link or the inverse link function:

$$\begin{aligned} g(\mu) &= \mathbf{x}'\beta = \eta \\ \mu &= g^{-1}(\mathbf{x}'\beta) \end{aligned}$$

The first form is illustrative because it emphasizes that generalized linear models use transformations (mappings) of the mean; they do not involve transformations of the data. The second formulation is helpful because it shows how predictions of the mean can be obtained, following estimation of β : compute the estimate of η and apply the inverse link function,

$$\hat{\mu} = g^{-1}(\hat{\eta}) = g^{-1}(\mathbf{x}'\hat{\beta})$$

Criteria for selecting an appropriate link function for a given data set vary. Link functions may follow directly from the form of the probability distribution or from some physical or biological model of μ . Various link function strategies are discussed below.

Nelder and Wedderburn (1972) showed that the maximum likelihood estimates for β can be obtained by iteratively solving

$$\mathbf{X}'\mathbf{W}\mathbf{X}\beta = \mathbf{X}'\mathbf{W}\mathbf{y}^*$$

where

$$\begin{aligned}\mathbf{W} &= \mathbf{D}'\mathbf{R}^{-1}\mathbf{D} \\ \mathbf{y}^* &= \hat{\boldsymbol{\eta}} + (\mathbf{y} - \hat{\boldsymbol{\mu}})\mathbf{D}^{-1} \\ \mathbf{D} &= \partial\boldsymbol{\mu}/\partial\boldsymbol{\eta} \\ \mathbf{R} &= \text{Var}[\mathbf{Y}] \\ \boldsymbol{\mu} &= E[\mathbf{Y}]\end{aligned}$$

Notice that \mathbf{y}^* is a working response variable, also termed **pseudo-data**. In the generalized linear model this working response can be motivated by a Taylor series expansion of the mean about a current estimate of \mathbf{b} . Taylor series and pseudo-data methods play an important role in extending the class of generalized linear models, for example, by including random effects (Section 14.4).

In practice, estimates of \mathbf{D} and \mathbf{R} are used in place of \mathbf{D} and \mathbf{R} . For standard linear models, $\boldsymbol{\eta} = E[\mathbf{Y}] = \boldsymbol{\mu}$, and hence $\mathbf{D} = \mathbf{I}$. Consequently, the solution for $\boldsymbol{\beta}$ reduces to generalized least squares. That is, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$.

Note that this formulation of generalized linear models avoids expressions that introduce explicit residual error terms. For example, unless the data are normally distributed, expressions such as

$$Y = \mu + e = g^{-1}(\mathbf{x}'\boldsymbol{\beta}) + e$$

can be confusing. Suppose, for example, that Y is a binary variable, taking on only the values 0 and 1. Since μ is continuous in the $(0,1)$ interval, the distribution of e would have to have zero mean, can take on only two values, $-\mu$ and $1-\mu$, and must have variance $\text{Var}[e] = \mu(1-\mu)$. The introduction of this shifted binary variable does not lead to any new insights. The model is completely specified through the distribution of Y , the choice of the linear predictor, and the link function. In the binary example, we would write

$$Y \sim \text{Binary}(\mu), \quad \mu = g^{-1}(\mathbf{x}'\boldsymbol{\beta})$$

14.3.2 Probability Distributions

The essential elements for estimating $\boldsymbol{\beta}$ in the generalized linear model are as follows:

- the **link function**, which determines $\boldsymbol{\eta}$ and \mathbf{D}
- the **probability distribution**, which determines the mean $\boldsymbol{\mu}$ and variance, $\text{Var}[\mathbf{Y}] = \mathbf{R}$. We use \mathbf{R} here to denote the matrix of variance functions for consistency with the notation used elsewhere in this book.

You can better understand the process for choosing a link function and the structure of the mean and variance by examining the probability distribution, or, more specifically, the likelihood function. Consider three distributions, the binomial, Poisson, and normal distribution. These are among the most widely used and familiar probability distributions, and they present the essential features of generalized linear models.

Binomial

For the binomial distribution with n trials and a success probability of π , the familiar form of the probability distribution function is

$$f(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$$

and the log-likelihood function is thus

$$l(\pi; y) = y \log\left(\frac{\pi}{1-\pi}\right) + n \log(1-\pi) + \log\binom{n}{y}$$

A sample proportion, y/n , thus has log-likelihood function

$$l(\pi; y) = ny \log\left(\frac{\pi}{1-\pi}\right) + n \log(1-\pi) + \log\binom{n}{ny}$$

The mean of y/n is π and the variance is $\pi(1-\pi)/n$.

Poisson

For the Poisson distribution, the probability distribution is

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

The log-likelihood function is thus

$$l(\lambda; y) = y \log(\lambda) - \lambda - \log(y!)$$

The mean and the variance of the Poisson are both equal to λ .

Normal

For the normal distribution, the probability density function is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$

and the log-likelihood function is thus

$$l(\mu, \sigma^2; y) = -\frac{1}{2\sigma^2}(y-\mu)^2 - \left(\frac{1}{2}\right)\log(2\pi\sigma^2)$$

where μ is the mean and σ^2 is the variance.

Common Features

The log-likelihood functions for these three distributions have a common form,

$$l(\theta, \phi; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

where

θ is termed the “natural” parameter

ϕ is a scale parameter

You can see that θ is a function of the mean. Denote this function as $\theta(\mu)$. Also, the variance can be expressed as a function of the mean and $a(\phi)$. Specifically,

$$\text{Var}[Y] = V(\mu) a(\phi)$$

where $V(\mu)$ denotes the **variance function**—the function of the mean involved in $\text{Var}[Y]$. Note that if $a(\phi) = 1$, then the variance function is also the variance of an observation. This is the case for the binomial and Poisson distributions. For the three distributions, the various functions can be summarized as shown in Table 14.1.

Table 14.1 Functions for the Three Distributions

	Binomial/ n	Poisson	Normal
Mean	π	λ	μ
$\theta(\mu)$	$\log[\pi/(1-\pi)]$	$\log(\lambda)$	μ
$a(\phi)$	$1/n$	1	σ^2
$V(\mu)$	$\pi(1-\pi)$	λ	1
$\text{Var}[Y]$	$\pi(1-\pi)/n$	λ	σ^2

Notice that for the binomial and Poisson distributions the mean also appears in the variance function. This is commonplace in the exponential family of distributions (see next). The independence of the mean and the variance for the normal distribution is an exception. In other words, you can determine the degree of variability independently from the mean in the normal case only. For other distributions in the exponential family, knowing the mean also determines the degree of variability, at least up to the function $a(\phi)$. This has far-reaching consequences for model formulation and estimation. For example:

- As noted in the previous section, using model formulations where the data are expressed in terms of the mean and additive error terms makes sense only in the normal case.
- Parameter estimation is usually an iterative process when means and variances are linked.
- You cannot combine arbitrary correlation models with nonnormal distributions. The resulting model may be impossible.

The Exponential Family of Distributions

A distribution whose log likelihood has the general form

$$l(\theta, \phi; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

is said to be a member of the **exponential family of distributions**. Generalized linear models apply to data that are independent and whose distribution belongs to the exponential family.

Consequently, if Y_1, Y_2, \dots, Y_n is a random sample from a distribution in the exponential family, then the log likelihood of Y_i is

$$l(\theta_i, \phi_i; y_i) = \frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i)$$

and the joint log likelihood of y_1, y_2, \dots, y_n is

$$l(\theta, \phi; y_1, y_2, \dots, y_n) = \sum \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right]$$

You can write the joint log likelihood in matrix form as

$$l(\boldsymbol{\theta}, \mathbf{N}; \mathbf{y}) = \mathbf{y}' \mathbf{A}^{-1} \boldsymbol{\theta} - (\mathbf{b}_{\boldsymbol{\theta}})^{\frac{1}{2}}' \mathbf{A}^{-1} \mathbf{b}_{\boldsymbol{\theta}}^{\frac{1}{2}} + \mathbf{1}' \mathbf{c}$$

where

\mathbf{A} is a diagonal matrix whose i^{th} diagonal element is $a(\phi_i)$

$\boldsymbol{\theta}$ is a vector whose i^{th} element is θ_i

$\mathbf{b}_{\boldsymbol{\theta}}$ is a vector whose i^{th} element is $b(\theta_i)$

\mathbf{c} is a vector whose i^{th} element is $c(Y_i, \phi_i)$

$\mathbf{1}$ is an n -dimensional vector of ones

The notation $\mathbf{b}_{\boldsymbol{\theta}}^{\frac{1}{2}}$ denotes a vector whose elements are the square roots of the corresponding elements of $\mathbf{b}_{\boldsymbol{\theta}}$.

14.3.3 Link Functions and Variance Structure

The general form of the log-likelihood function reveals some of the basic considerations in the generalized linear model.

Link Functions

First, the observations are linear with respect to the **natural parameter**, θ . For normally distributed data, $\theta = \mu$. Because the likelihood is linear in μ , it is reasonable to fit a model directly to μ , or equivalently to Y .

On the other hand, for Poisson data the observations are linear in $\theta = \log(\lambda)$. Thus, fitting a linear model to $\log(\lambda)$ is more reasonable than fitting a model directly to λ (or Y). Fitting a model to $\log(\lambda)$ gives rise to the **log-linear model**.

Following the strategy of using θ as a guideline, you fit a linear model to $\log[\pi/(1-\pi)]$ for binomial data. Fitting a regression model to $\log[\pi/(1-\pi)]$, or **logit**, gives rise to **logistic regression**.

In each of these examples, $\theta(\mu)$, the function that relates μ to the natural parameter, is used as the link function. This function is called the **canonical link function**. It is common to choose the canonical link function as *the* link in the analysis, but this is neither required nor necessarily recommended. The canonical link does not guarantee that inversely linked values are in the support of the response variable. For example, the canonical link function for the gamma distribution is the reciprocal link, $\mu = 1/\eta$. In order for predictions to be meaningful, η must be constrained to be positive, because a gamma distributed random variable cannot take on negative values. The log link function, $\mu = \exp\{\eta\}$, is a more reasonable choice in gamma regression models.

Variance Structure

You can describe the structure of the variance-covariance matrix of the observations in terms of the scale parameter and the variance function. Specifically,

$$\text{Var}[\mathbf{Y}] = \mathbf{R} = \mathbf{R}_{\mu}^{1/2} \mathbf{A} \mathbf{R}_{\mu}^{1/2}$$

where

\mathbf{R}_{μ} is a diagonal matrix whose i^{th} diagonal element is $V(\mu_i)$, the variance function for the i^{th} observation

$\mathbf{R}_{\mu}^{1/2}$ is a diagonal matrix of square roots of the corresponding elements of \mathbf{R}_{μ}

\mathbf{A} is the scale parameter matrix defined earlier

For the three example distributions, normal, binomial/ n (sample proportion), and Poisson, \mathbf{R}_{μ} and \mathbf{A} are shown in Table 14.2.

Table 14.2 Three Example Distributions

Distribution	\mathbf{R}_{μ}	\mathbf{A}
Normal	\mathbf{I}	$\sigma^2 \mathbf{I}$
Binomial/ n	$\text{diag}[\pi_i(1-\pi_i)]$	$\text{diag}(1/n_i)$
Poisson	$\text{diag}(\lambda_i)$	\mathbf{I}

14.3.4 Predicting Means from the Inverse Link Function

The **inverse link function** is defined as $g^{-1}(\eta) = \mu$. You use the inverse link to obtain predicted values of μ from the estimated β vector:

$$\hat{\mu} = g^{-1}(\mathbf{x}'\hat{\beta})$$

For the normal distribution, $g^{-1}(\mathbf{x}'\beta) = \mathbf{x}'\beta$, since $\eta = \mu$. For the Poisson distribution with canonical link, $\eta = \log(\lambda)$ and hence $\lambda = g^{-1}(\mathbf{x}'\beta) = \exp\{\mathbf{x}'\beta\}$. For the binomial, $\eta = \log[\pi/(1-\pi)]$ and hence $\pi = g^{-1}(\mathbf{x}'\beta) = \exp\{\mathbf{x}'\beta\}/[1+\exp\{\mathbf{x}'\beta\}]$.

14.3.5 Deviance

The **deviance** is defined as

$$D(\hat{\mu}; y) = 2[l(y; y) - l(\hat{\mu}; y)]$$

where

$l(y; y)$ is the value of the log likelihood evaluated at $\mu = y$

$l(\hat{\mu}; y)$ is the value of the log likelihood for the model under consideration

The deviance is a generalization of the SS(error) in analysis of variance and the likelihood ratio χ^2 in contingency tables. The deviance is equal to SS(error) for normal models, and the likelihood ratio χ^2 is equal to the deviance for Poisson models. You can use the deviance to evaluate **goodness of fit** and to test hypotheses.

Goodness of Fit

The deviance is a function of θ and ϕ , and ϕ is known for many distributions. When ϕ is known, the deviance is approximately distributed χ^2 with $n-p$ degrees of freedom, where n is the number of observations and p is the rank of \mathbf{X} . You can use the deviance as a χ^2 statistic to test the goodness of fit of the model.

Hypothesis Testing

If β is partitioned into two vectors, β_1 and β_2 , so that the model can be written $\mathbf{X}\beta = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$, you can use the difference between the deviance of the full model and the deviance of the model fitting $\mathbf{X}_1\beta_1$ alone as a likelihood ratio statistic to test $H_0: \beta_2 = 0$. The likelihood ratio statistic is approximately distributed as a χ^2 with $p - p_1$ degrees of freedom, where p_1 is the rank of \mathbf{X}_1 .

When ϕ is not known, you can estimate it and use it to compute the **scaled deviance**, defined as

$$D^*(\hat{\mu}; y) = D(\hat{\mu}; y) / \hat{\phi}$$

You can use the scaled difference to test hypotheses. For example, in testing $H_0: \beta_2 = 0$ with a normal errors model, the difference between the deviance of the model $\mathbf{X}\beta$ and the deviance of $\mathbf{X}_1\beta_1$ is $SS(\beta_2 | \beta_1)$. The estimate of ϕ is $\hat{\sigma}$, that is, MS(error). The resulting scaled deviance, adjusting for the degrees of freedom for $SS(\beta_2 | \beta_1)$, is

$$F = \frac{MS(\beta_1 | \beta_2)}{MSE}$$

The F -ratio is distributed F with $p - p_1$ numerator degrees of freedom and $n - p$ denominator degrees of freedom.

14.3.6 Inference Using Estimable Functions

In addition to the deviance, you can use estimable functions of β for inference in the generalized linear model. The basic result is

$$\text{Var}[\hat{\beta}] = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

It follows that

- the variance of an estimable function $\mathbf{K}'\hat{\beta}$ is

$$\text{Var}[\mathbf{K}'\hat{\beta}] = \mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}$$

- the Wald statistic for $H_0: \mathbf{K}'\beta = \mathbf{K}'\beta_0$ is

$$(\mathbf{K}'\hat{\beta} - \mathbf{K}'\beta_0)' [\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}]^{-1} (\mathbf{K}'\hat{\beta} - \mathbf{K}'\beta_0)$$

Using the Wald Statistic for χ^2 Tests

Recall that $\mathbf{W} = \mathbf{D}'(\mathbf{R}_\mu^{-1/2} \mathbf{A} \mathbf{R}_\mu^{-1/2})^{-1} \mathbf{D}$. If \mathbf{A} is known (e.g., as for the Poisson or binomial distribution) then the Wald statistic is approximately distributed χ_v^2 , where $v = \text{rank}(\mathbf{K})$.

F-Tests

When \mathbf{A} depends on an unknown scale parameter ϕ (e.g., as it does for the normal distribution), then \mathbf{W} depends on an estimate of \mathbf{A} . In such cases, the Wald statistic divided by $\text{rank}(\mathbf{K})$ has an approximate $F_{(v_1, v_2)}$ distribution, where $v_1 = \text{rank}(\mathbf{K})$ and v_2 are the degrees of freedom associated with the estimate of ϕ . You can use the ratio of the Wald statistic and $\text{rank}(\mathbf{K})$ to test $H_0: \mathbf{K}'\beta = \mathbf{0}$.

Relation between Inference Using Generalized Linear Models versus Standard Linear Models

In the standard linear model with normal errors, the observation vector, \mathbf{y} , has a normal distribution with mean $\mathbf{X}\beta$ and variance $\sigma^2\mathbf{I}$. The log-likelihood function of this model can be written as

$$\begin{aligned} l(\beta, \sigma^2; \mathbf{y}) &= \left(-\frac{1}{2\sigma^2} \right) (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) - \frac{n}{2} \log(2\pi\sigma^2) \\ &= \frac{1}{\sigma^2} \left(\mathbf{y}'\mathbf{X}\beta - \frac{1}{2} \beta'\mathbf{X}'\mathbf{X}\beta \right) - \left[\frac{\mathbf{y}'\mathbf{y}}{2\sigma^2} + \frac{n}{2} \log(2\pi\sigma^2) \right] \end{aligned}$$

The natural parameter for this model is $\mathbf{X}\beta$ and the scale parameter is σ^2 . Using the canonical link, the generalized linear model is

$$\eta = \mathbf{X}\beta$$

which corresponds to the standard linear model. A link function with $\mu = \eta$ is called an **identity link**. The F -ratios used in the analysis of variance are identical to the Wald/rank(\mathbf{K}) F -statistics defined above.

14.4 From GLMs to GLMMs

14.4.1 Incorporating Random Effects

Extending the model formulation of Section 14.3 to accommodate random effects is a simple matter. Instead of the linear predictor $\eta = \mathbf{X}\beta$, we now include random effects,

$$\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

and specify that the conditional distribution of $\mathbf{Y}|\mathbf{u}$ is in the exponential family with mean $E[\mathbf{Y}|\mathbf{u}] = g^{-1}(\eta)$. You apply the same basic strategy to form a generalized linear mixed model as for GLMs, but focusing on the conditional moments. That is, the form of the conditional likelihood determines the **variance function** and contains the natural parameter. The conditional mean function $\theta(\mu)$ can be used as a **canonical link**, or some other function of μ can be used. The difficulty of fitting GLMMs does not lie in the mathematical formulation of the models. What can be difficult is the calculation of the likelihood of \mathbf{Y} as a basis for statistical inference.

The fixed effects generalized linear models discussed in Section 14.3 are based on the **likelihood function** of the data. In the linear mixed model, estimation and inference are also based on the **marginal** log-likelihood or residual log-likelihood function of the data. The marginal distribution is obtained by integrating the joint distribution of data and random effects over the random effects. When the random effects are normal and the data are normal, obtaining the marginal distribution is simple; it is a normal distribution with mean $\mathbf{X}\beta$ and variance $\mathbf{ZGZ}' + \mathbf{R}$. When the random effects are normal and the distribution of $\mathbf{Y}|\mathbf{u}$ is *not* normal, obtaining the marginal distribution is generally difficult. You can carry the integration out numerically; this is the approach taken by the NLMIXED procedure. See Chapter 15 in this book or Chapter 12 in Fitzmaurice, Laird, and Ware (2004). Alternatively, you can apply linear mixed model estimation repeatedly based on an approximated model. This is the **pseudo-likelihood** approach taken by the GLIMMIX procedure.

14.4.2 The Pseudo-likelihood Approach

In Section 14.3.1 we described a method of fitting a generalized linear model by iteratively solving the linear model weighted least-squares problem

$$\mathbf{X}'\mathbf{W}\mathbf{X}\beta = \mathbf{X}'\mathbf{W}\mathbf{y}^*$$

where \mathbf{y}^* is a working dependent variable, the so-called **pseudo-data**. The process is iterative because the pseudo-data, as well as the weight matrix \mathbf{W} , depend on the current estimates. One approach to motivate the fitting method is to apply a first-order Taylor series to the mean model $g^{-1}(\mathbf{x}\beta)$ for the mean. The expansion depends on the value β^* that is chosen in the linearization. The result is a linear regression model with unequal variances that depend on β^* . This model can be fit with standard regression software (e.g., PROC REG), which yields estimates of β . Based on these new estimates, the pseudo-data and the variance weights are updated and a new linear model is derived. The process continues until estimates of the fixed effects no longer change.

The same basic idea can be applied in the case of a generalized linear **mixed** model. The nonlinearity is removed by applying a first-order Taylor expansion to $g^{-1}(\mathbf{X}\beta + \mathbf{Z}\mathbf{u})$ about current values of β and \mathbf{u} . This results in a weighted linear mixed model that can be estimated

with standard linear mixed model methods. The estimates of β and predictors of \mathbf{u} so obtained are used to update the pseudo-data, and the process repeats.

Generalized Mixed Model Equations

Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) show that solutions for β and \mathbf{u} can be obtained by iteratively solving the equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{y}^* \\ \mathbf{Z}'\mathbf{W}\mathbf{y}^* \end{bmatrix}$$

where \mathbf{W} and \mathbf{y}^* are defined as in the solution equations for the fixed effects generalized linear model given above. That is,

$$\begin{aligned} \mathbf{W} &= \mathbf{D}\mathbf{R}^{-1}\mathbf{D} \\ \mathbf{y}^* &= \hat{\eta} + (\mathbf{y} - \hat{\mu})\mathbf{D}^{-1} \\ \mathbf{D} &= [\partial\mu/\partial\eta] \\ \mathbf{R} &= (\mathbf{R}_\mu^{1/2}\mathbf{A}\mathbf{R}_\mu^{1/2}) \end{aligned}$$

The Breslow-Clayton and Wolfinger-O'Connell procedures are similar in that they both use the generalized mixed model equations. The primary difference between what Breslow and Clayton (1993) term **penalized quasi-likelihood** (PQL) and what Wolfinger and O'Connell (1993) term **pseudo-likelihood** (PL) or **restricted pseudo-likelihood** (REPL) lies in the estimation of the parameter ϕ . In the Breslow-Clayton procedure the scale parameter is fixed at $\phi = 1$, whereas in the Wolfinger-O'Connell procedure it is always estimated.

Predictable Functions

As in standard mixed models, the primary tool of inference in generalized linear mixed models is the predictable function, $\mathbf{K}'\beta + \mathbf{M}'\mathbf{u}$. The logic leading from the objectives of a particular study to the selection of predictable functions that address those objectives is identical to that of standard mixed models. Letting $\mathbf{L}' = [\mathbf{K}' \mathbf{M}']$, the prediction error variance of a predictable function is

$$\text{Var}[\mathbf{K}'\hat{\beta} + \mathbf{M}'(\hat{\mathbf{u}} - \mathbf{u})] = \mathbf{L}'\mathbf{C}\mathbf{L}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}$$

Note that relatively little research has been done on the small-sample properties of inference statistics for the generalized linear mixed model. The test statistics are basically reasonable-looking extensions of standard tests for mixed models and generalized linear models. More work is needed to either validate these procedures or develop needed corrections or alternatives.

Wald Statistics

With this caution in mind, hypotheses about predictable functions can be tested using Wald statistics or F -statistics. The basic form of the Wald statistic is

$$F_w = (\mathbf{L}'\hat{\mathbf{a}})'(\mathbf{L}'\mathbf{C}\mathbf{L})^{-1}(\mathbf{L}'\hat{\mathbf{a}})$$

where $\hat{\alpha} = [\hat{\beta}' (\hat{\mathbf{u}} - \mathbf{u})']'$.

The Wald statistic is approximately distributed as χ^2_v , where $v = \text{rank}(\mathbf{L})$. When the conditional variance, \mathbf{R} , depends on a known scale parameter (i.e., \mathbf{A} is known), then the χ^2 test may be used.

F-Tests

If \mathbf{R} depends on an unknown scale parameter, the Wald statistic divided by $\text{rank}(\mathbf{L})$ is preferable:

$$F = F_w / \text{rank}(\mathbf{L})$$

is approximately distributed as $F_{(v_1, v_2)}$, where $v_1 = \text{rank}(\mathbf{L})$ and v_2 are the degrees of freedom used to estimate $\mathbf{L}'\mathbf{C}\mathbf{L}$. In simple cases, v_2 corresponds to the degrees of freedom required to estimate ϕ (e.g., σ^2 in normal errors models). In more complex cases, you must approximate v_2 —for example, by using a Satterthwaite-type procedure.

You can see that the PQL procedure discussed above is appropriate when the scale parameter ϕ is known. When ϕ is unknown (as in the normal errors case) or when nominal assumptions about ϕ are violated (as in the example in Section 14.6), you should use a procedure capable of estimating ϕ , such as the PL and REPL procedures.

14.4.3 The Role of the Scale Parameter

The discussion of test statistics in the previous section has shown that it is not reasonable to set $\phi = 1$ for distributions in the exponential family that contain a free scale parameter—for example, the normal, inverse Gaussian, and gamma distribution. On the other hand, if the conditional distribution does not contain a free scale parameter (Bernoulli, binomial, Poisson, etc.) it should not necessarily be added to the estimation by default.

A multiplicative scale parameter on the variance function is a common device to account for overdispersion in generalized linear models. For example, if you fit a Poisson regression with log link and determine that there is overdispersion in the data, you may replace the Poisson variance function $V(\mu) = \exp\{\mathbf{x}'\boldsymbol{\beta}\}$ with $\phi V(\mu)$. The overdispersion parameter ϕ then “adjusts” for the fact that the data are more variable than what is expected under the assumption of Poisson data.

Adding scale parameters to the variance function has some important consequences. First, overdispersion as a phenomenon is only meaningful relative to a baseline distribution. Data that are overdispersed relative to a Poisson distribution may not be overdispersed relative to a negative binomial distribution. Second, it is important to investigate the reasons for overdispersion. It may point to an important breakdown in the model that should not be “patched up” or glossed over by scaling the variance function. The following list presents some common causes for overdispersed data.

- **Correlated data.** Positive association among the observations creates a situation where the number of effective observations is reduced relative to a set of independent data of the same size. In other words, n correlated data do not deliver the same amount of information as n independent observations. The variability of statistics thus increases compared to the independence case. The appropriate course of action is to account for the correlations in the model, such as by using random effects.

- **Omitted variables.** The residual dispersion in the model increases if important variables are omitted from the analysis. The standard estimate of ϕ based on residuals increases as a result. The appropriate action is not to multiply the variance function with a scalar estimated from the data. The appropriate action is to correct the model.
- **Misspecified distribution.** Data may appear overdispersed because they do not follow the assumed distribution. A typical example is the zero-inflation in count processes. Many data show an excess amount of zeros compared to, say, the Poisson distribution. One reason might be that the zeros in the data are generated by two processes, one that produces zeros with probability π , and a Poisson process. The resulting data are a zero-inflated mixture of two processes. Assuming a single Poisson process and increasing the variance of all observations proportionately is not the needed correction.

14.4.4 Estimation Methods in PROC GLIMMIX

The GLIMMIX procedure implements the PL and REPL approaches of Wolfinger and O'Connell (1993) as well as the PQL and the marginal quasi-likelihood (MQL) approach of Breslow and Clayton (1993).

The four basic estimation methods for generalized linear mixed models are controlled by the METHOD= option of the PROC GLIMMIX statement. The four methods are METHOD=RSPL (the default), RMPL, MSPL, and MMPL. The abbreviations are deciphered as follows. The last two letters indicate that estimation is based on a pseudo-likelihood in the sense that a linearization is applied, and a likelihood-based technique is applied to a pseudo-model. The first letter of the METHOD= identifier determines whether the log-likelihood is formed as a **residual** (restricted) log-likelihood or a **maximum** log-likelihood. The second letter identifies the locus of the expansion. In a **Subject-specific** expansion, the linearization is carried out about current estimates of β and u . This is the approach taken by Wolfinger and O'Connell (1993). In a **Marginal** expansion, the linearization is carried out about a current estimate of β and the $E[u] = \mathbf{0}$. Table 14.3 contrasts the methods. Breslow and Clayton (1993) term this the marginal quasi-likelihood (MQL) approach.

Table 14.3 Estimation Methods of the GLIMMIX Procedure (METHOD= Option)

Estimation Principle for Pseudo-Model	Expansion Locus in Linearization	
	$\hat{\beta}, \hat{u}$	$\hat{\beta}, E(u) = \mathbf{0}$
Restricted Maximum Likelihood	RSPL	RMPL
Maximum Likelihood	MSPL	MMPL

Notes: If you estimate ϕ always, then RSPL equals REPL and MSPL equals PL of Wolfinger and O'Connell (1993). If you do not estimate ϕ , then MMPL equals MQL of Breslow and Clayton (1993).

To overcome the problems of omitting the scale parameter ϕ for distributions where $a(\phi)$ is not equal to 1 and to overcome the potential problem of masking a model breakdown by adding a default scale parameter, the GLIMMIX procedure approaches the issue as follows. If the conditional distribution of $Y|u$ does not contain a free scale parameter, ϕ is constrained to equal 1 by default. This applies, for example, to the Bernoulli, binomial, geometric, exponential, and Poisson distribution. If the conditional distribution contains a scale parameter, then ϕ is estimated by default. This applies, for example, to the normal, inverse Gaussian, gamma, beta,

and negative binomial distribution. Through statements and options you can add the scale parameter or constrain it, as needed.

The next two sections consider specific examples of generalized linear mixed models.

14.5 Example: Binomial Data in a Multi-center Clinical Trial

Section 14.2.1 described an experiment involving two treatments conducted at eight randomly sampled clinics. The data, which appeared in Beitler and Landis (1985), are given as Data Set 14.5, “Clinics,” in Appendix 2, “Data Sets.” At the j^{th} CLINIC ($j=1,2,\dots,8$), n_{ij} subjects were assigned to the i^{th} treatment (TRT) ($I=1,2$). At the ij^{th} treatment-clinic combination, y_{ij} subjects responded favorably and the remaining $n_{ij} - y_{ij}$ did not. The number of patients for the ij^{th} CLINIC by TRT combinations is denoted by NIJ in the data set; the number of favorable responses y_{ij} is denoted by FAV. The response variable of interest was the proportion of favorable responses, $p_{ij} = Y_{ij}/n_{ij}$, or FAV/NIJ.

The distribution of the Y_{ij} is binomial(n_{ij}, π_{ij}), where π_{ij} is the probability of a subject at the ij^{th} CLINIC by TRT having a favorable response. Thus, p_{ij} is distributed binomial/ n_{ij} and the conditional likelihood given the j^{th} clinic has the binomial/ n form as given in Section 14.3. It has the features shown in Table 14.4.

Table 14.4 Features of Conditional Likelihood

Expression	Function	Value
Conditional mean	μ_{ij}	π_{ij}
Natural parameter	$\theta(\mu_{ij})$	$\log[\pi_{ij}/(1-\pi_{ij})]$
Variance function	$V(\mu_{ij})$	$\pi_{ij}(1-\pi_{ij})$
Scale parameter	$a(\phi_{ij})$	$1/n_{ij}$

14.5.1 Analysis of Binomial Data Using Logit (Canonical) Link

Model

Using the canonical link, a generalized linear mixed model for these data is

$$\eta_{ij} = \log[\pi_{ij}/(1-\pi_{ij})] = \mu + \tau_i + c_j + (\pi c)_{ij}$$

where

μ is the intercept

τ_i is the effect of the i^{th} TRT

c_j is the effect of the j^{th} CLINIC

$(\pi c)_{ij}$ is the ij^{th} TRT \times CLINIC interaction effect

Because the clinics are a random sample, the CLINIC and TRT \times CLINIC effects are random. The assumptions are as follows:

- The vector of clinic effects, $\mathbf{c} = [c_1, \dots, c_8]'$, is distributed $N(\mathbf{0}, \sigma_c^2 \mathbf{I})$.
- The vector of TRT \times CLINIC effects, $(\tau\mathbf{c}) = [(\tau\mathbf{c})_{11}, \dots, (\tau\mathbf{c})_{28}]'$, is distributed $N(\mathbf{0}, \sigma_{TC}^2 \mathbf{I})$.

Thus, the random effects vector is $\mathbf{u}' = [\mathbf{c}' (\tau\mathbf{c})']$. Its variance-covariance matrix, $\mathbf{G} = \text{Var}[\mathbf{u}]$, is

$$\mathbf{G} = \begin{bmatrix} \sigma_c^2 \mathbf{I} & 0 \\ 0 & \sigma_{TC}^2 \mathbf{I} \end{bmatrix}$$

The generalized mixed model equations for this model are

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{y}^* \\ \mathbf{Z}'\mathbf{W}\mathbf{y}^* \end{bmatrix}$$

where

$$\boldsymbol{\beta} = [\mu, \tau_1, \tau_2]'$$

$$\mathbf{W} = \mathbf{D}'(\mathbf{R}_\mu^{-1/2} \mathbf{A} \mathbf{R}_\mu^{-1/2})^{-1} \mathbf{D}$$

$$\mathbf{D} = [\partial \mathbf{u} / \partial \boldsymbol{\eta}] = \text{diag}[\pi_{ij}(1-\pi_{ij})]$$

$$\mathbf{R}_\mu = \text{diag}[\pi_{ij}(1-\pi_{ij})]$$

$$\mathbf{A} = \text{diag}[1/n_{ij}]$$

$$\mathbf{y}^* = \hat{\eta} + (\mathbf{y} - \hat{\mu})\mathbf{D}^{-1}$$

$$\hat{\eta} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$$

$$\hat{\pi} = \exp(\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}) / [1 + \exp(\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}})]$$

You obtain $\hat{\pi}$ from the inverse of the link function.

You can fit this model with the GLIMMIX procedure by using the following program statements.

Program

```
proc glimmix data=clinics;
  class clinic trt;
  model fav/nij = trt / solution ddfm=satterth;
  random clinic trt*clinic;
run;
```

The syntax is very similar to what you would use to fit a mixed model with the MIXED procedure. The important difference in this example is the use of the *events/trials* syntax to specify the binomial proportion in the MODEL statement. The numerator variable is the number of events or “successes” (Y_{ij}), and the denominator is the total number of trials (n_{ij}) for a given fixed-by-random effect combination. Here, FAV gives the number of favorable responses and NIJ gives the total number of subjects observed at the ij^{th} clinic-treatment combination. Thus the

response variable, the sample proportion, is FAV/NIJ. The events/trials syntax is also used in PROC GENMOD and PROC LOGISTIC. If you use this syntax, PROC GLIMMIX defaults to fitting a model for binomial data and chooses the logit link as the default link.

The DDFM=SATTERTH option in the MODEL statement requests that the denominator degrees of freedom for tests of fixed effects be computed by the Satterthwaite method.

Results

Output 14.1 Binomial Example—LOGIT Link—Selected Output

Model Information	
Data Set	WORK.A
Response Variable (Events)	fav
Response Variable (Trials)	nij
Response Distribution	Binomial
Link Function	Logit
Varia0nce Function	Default
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Satterthwaite

Dimensions	
G-side Cov. Parameters	2
Columns in X	3
Columns in Z	24
Subjects (Blocks in V)	1
Max Obs per Subject	16

Iteration History					
Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	5	49.290327801	0.98031717	9.214E-6
1	0	4	50.144322891	0.22087434	0.000017
2	0	3	50.234558383	0.01524299	4.089E-6
3	0	2	50.239087668	0.00064807	4.641E-6
4	0	1	50.239258606	0.00002281	3.837E-6
5	0	0	50.239264016	0.00000000	3.807E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	50.24
Generalized Chi-Square	13.52
Gener. Chi-Square / DF	0.97

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
clinic	2.0063	1.2687
clinic*trt	0.05949	0.2031

Solutions for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-0.4161	0.5600	7.648	-0.74	0.4797
trt	cntl	-0.7469	0.3320	4.62	-2.25	0.0787
trt	drug	0

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	4.62	5.06	0.0787

Interpretation

The “Model Information” table provides information about how the GLIMMIX procedure forms the model and about the method of estimation. The default estimation method for generalized linear mixed models is RSPL, a pseudo-likelihood method based on REML estimation in a linearized model.

The “Dimensions” table informs you about the dimension of the parameter space and about the size of important matrices. PROC GLIMMIX terms random effects that are specified through the Z matrix as “G-side effects.” Consequently, there are two G-side covariance parameters in this model, the variance of the random CLINIC and TRT × CLINIC effects.

The “Iteration History” table provides information about the optimization process. In a generalized linear mixed model where estimation is based on pseudo-data, the optimization process is doubly iterative. At each stage, a linear mixed model is fit, itself an iterative process. For this analysis, the GLIMMIX procedure forms six linear pseudo-models. The optimization for the first model requires five iterations, the optimization for the second model requires four iterations, and so forth. After optimization 5 the estimates have stabilized and convergence has been achieved. Repeating the process of updating the pseudo-data would now lead to the same data set and identical estimates in the next optimization.

For models fit with pseudo-likelihood methods, PROC GLIMMIX reports in the “Fit Statistics” table the $-2 \text{ Res Log Pseudo-Likelihood}$ and two chi-square statistics. The Generalized Chi-Square statistic corresponds to $\mathbf{r}'(\mathbf{V}^*)^{-1}\mathbf{r}$ in the final pseudo-model. The matrix \mathbf{V}^* is the marginal variance matrix in this model. The scale parameter ϕ is scaled from \mathbf{V} in this calculation so that the ratio of the generalized χ^2 statistic and its degrees of freedom is the customary estimate of the scale parameter ϕ . Notice that if this were a normal linear mixed model, the estimate of ϕ would be the residual variance estimate.

In many generalized linear mixed models, the observed conditional variance can differ from what is expected under the baseline distribution in much the same way in which overdispersion is possible in generalized linear models (see Section 14.4.3). In this example, the conditional variance should be determined by the binomial distribution, i.e., $\text{Var}[Y_{ij}|\mathbf{u}] = [\pi_{ij}(1-\pi_{ij})]/n_{ij}$. In other words, the dispersion scale parameter ϕ is assumed to be 1, because the binomial distribution does not have a scale parameter. The Generalized Chi-Square statistic divided by its degrees of freedom estimates the residual variation based on the final pseudo-data. It is often incorrectly assumed that values larger than 1.0 are indicative of a misspecified conditional distribution and require the addition of extra scale parameters. For an example of overdispersion in the conditional distribution and how to diagnose it, see Section 14.6.

To explicitly estimate an additional scale parameter with the GLIMMIX procedure, you can add the following statement to the PROC GLIMMIX statements:

```
random _residual_;
```

The estimates of the two variance components are shown in the “Covariance Parameter Estimates” table. The estimate of the CLINIC variance σ_C^2 is 2.0063, and the estimate of the TRT \times CLINIC variance σ_{TC}^2 is 0.05949. These variance components are interpreted like any other variance component in a mixed model except that they are estimated in terms of the scale determined by the link function. Here, we are estimating the variance of CLINIC effects and TRT \times CLINIC effects measured on the “logit scale.”

When the link function is a model of the mean that has a real-world basis, you can interpret the variance components with respect to that model (for example, see the alternative analysis of these data using the **probit** model given below). When you use the *canonical link* and do not frame it in real-world terms, the meaning of the variance components is more elusive. The random effects act linearly on η , and thus indirectly on the conditional mean through the model given by the inverse link, $g^{-1}(\eta)$.

From the “Solutions for Fixed Effects” table we see that the estimate of $\mu + \tau_2$ is -0.4161 , and the estimate of $\tau_1 - \tau_2$ is -0.7469 . The p -value of 0.0787 for the first treatment entry in the table thus tests $H_0: \tau_1 = \tau_2$ as described in Section 14.4. The corresponding F -test is given in the “Type III Tests of Fixed Effects” table. Here, $F = 5.06$ with a p -value of 0.0787. There is weak evidence (depending on the α -level you deem appropriate) of a difference in the probability of a favorable output between the CNTL and TRT drugs.

Program

The following program requests least-squares means for the treatment levels and uses the ESTIMATE statement to produce a comparison of the two treatments as well as parameter estimates for treatment CNTL in clinic 6.

```

proc glimmix data=clinics;
  class clinic trt;
  model fav/nij = trt / solution ddfm=satterth
    oddsratio;
  random clinic trt*clinic;
  lsmeans trt / ilink;
  estimate 'trt diff' trt 1 -1 / exp cl;
  estimate 'trt 1 clinic 1 BLUP' intercept 1 trt 1 0
    | clinic 1 0 / ilink;
  estimate 'trt 1 clinic 6 BLUP' intercept 1 trt 1 0
    | clinic 0 0 0 0 1 0 / ilink;
  ods select LSMeans Estimates;
run;

```

PROC GLIMMIX performs these post-processing tasks on the linked scale—that is, the logit scale in this example. Often you are interested in results on the scale of the data, which is the probability scale here. For models with the logit link, you may also be interested in expressing inferences in terms of odds ratios. The ODDSRATIO option in the MODEL statement computes odds ratios for the TRT effect in the model.

In models with logit link you can also use the EXP option in the ESTIMATE (or the LSMESTIMATE) statement to compute odds or odds ratios. The first ESTIMATE statement computes a difference on the logit scale, and hence you get a log odds ratio. The EXP option exponentiates the estimate and its confidence limits. This leads to an odds ratio that compares the odds of an outcome between the two treatments.

The ILINK option allows you to apply the inverse link function to the estimate on the linear scale. Output 14.2 shows the results of the least-squares means and ESTIMATE analysis in this model.

Results

Output 14.2 Binomial Example—LOGIT Link—Odds Ratio, Estimate, and Least-Squares Mean Results

Odds Ratio Estimates					
trt	_trt	Estimate	DF	95% Confidence Limits	
cntl	drug	0.474	4.62	0.198	1.137

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
trt diff	-0.7469	0.3320	4.62	-2.25	0.0787	0.05	-1.6218	0.1280
trt 1 clinic 1 BLUP	-1.2875	0.3545	2.589	-3.63	0.0457	0.05	-2.5243	-0.05080
trt 1 clinic 6 BLUP	-2.7832	0.7400	14	-3.76	0.0021	0.05	-4.3703	-1.1961

Estimates							
Label	Mean	Standard Error Mean	Lower Mean	Upper Mean	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
trt diff					0.4738	0.1976	1.1366
trt 1 clinic 1 BLUP	0.2163	0.06009	0.07418	0.4873	.	.	.
trt 1 clinic 6 BLUP	0.05824	0.04059	0.01249	0.2322	.	.	.

trt Least Squares Means							
trt	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
cntl	-1.1630	0.5650	7.849	-2.06	0.0742	0.2381	0.1025
drug	-0.4161	0.5600	7.648	-0.74	0.4797	0.3975	0.1341

Interpretation

The “Odds Ratio Estimates” table gives the odds ratio for a comparison of the two treatments. The odds of a favorable response in the control group are less than half the odds of a favorable response in the group receiving the drug. Odds ratios computed in the “Odds Ratio Estimates” table of PROC GLIMMIX are obtained from differencing least-squares means of classification effects. In this simple example, the least-squares mean differences that correspond to log odds ratios are identical to simple parameter differences of the parameters (see below). The 95% confidence limits of the odds ratio covers 1.0, indicating that this discrepancy of the estimated odds ratio from the case of equal odds is not significant at the $\alpha = 0.05$ level. This coincides with the p -value of 0.0787 in the “Solutions for Fixed Effects” table in Output 14.1.

The degrees of freedom in the “Odds Ratio Estimates” table (and elsewhere in the PROC GLIMMIX output) reflects the Satterthwaite method (DDFM=SATTERTH in the MODEL statement). In generalized linear models it is often desirable to perform chi-square-based inferences instead of t - or F -based inferences. You can accomplish this easily by changing the degrees-of-freedom method in the MODEL statement to DDFM=NONE.

The “Estimates” table collects the results of the ESTIMATE statements and provides the following:

- **Estimates** computed by applying the generalized mixed model solutions to the various predictable functions. Estimates are defined on the **linked** scale. ‘Mean’ uses the inverse link to express the estimate on the original scale. Here, ‘Estimates’ are logits, and ‘Means’ are predicted probabilities, *except where the predictable function is a difference*. An example below shows how the estimate and the inversely linked estimate (‘Mean’) are obtained and how to interpret the ‘ilink’ed estimate.
- **Standard error.** The GLIMMIX procedure computes the standard errors using the formula $(\mathbf{L}'\mathbf{CL})^{-1}$ given in Section 14.4.

- **T-test** ('DDF', 'T', 'Pr > |T|'). The *t*-tests result from $t=(\text{estimate})/(\text{std error})$. DDF are determined using the Satterthwaite method.
- **Confidence limits**. The columns labeled 'Lower' and 'Upper' contain confidence limits for the 'Estimate'. They apply to the linked (=logit) scale. The confidence limits on the mean scale are given as 'Lower Mean' and 'Upper Mean'. They use the original scale (in this case, the probability scale). Confidence limits for the exponentiated estimates ratios are labeled as 'Exponentiated Lower' and 'Exponentiated Upper'.

The least-squares means in the "trt Least Squares Means" table are obtained using estimable functions for treatment means—e.g., $\mu + \tau_i$ for these data. The least-squares mean for the CNTL treatment group is -1.163 and the least-squares mean for the DRUG group is -0.4161 . These values apply to the logit scale. Inversely linking these values yields a probability of favorable response of 0.2381 in the CNTL group and 0.3975 in the DRUG treatment group. The drug appears to increase the probability of a favorable response, but that increase is not significant at the 0.05 significance level. The *p*-value for $H_0: \tau_1 = \tau_2$ is 0.0787 as seen earlier in the "Solutions for Fixed Effects" table and the "Estimates" table. The same *p*-value would be obtained if you add the DIFF option to the LSMEANS statement.

Example: How Estimates and Means Are Obtained

This example shows how the solution vectors for the fixed and random effects are used to obtain the logit 'Estimate' and the sample proportion 'Mean' for these data. Output 14.3 repeats the solution vectors for fixed and random effects.

Output 14.3 Solution Vectors for Binomial Data, Logit Model

Solutions for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-0.4161	0.5600	7.648	-0.74	0.4797
trt	cntl	-0.7469	0.3320	4.62	-2.25	0.0787
trt	drug	0

Solution for Random Effects							
Effect	trt	clinic	Estimate	Std Err Pred	DF	t Value	Pr > t
clinic		1	-0.1246	0.5966	8.091	-0.21	0.8398
clinic		2	1.7828	0.6159	8.947	2.89	0.0179
clinic		3	0.9453	0.6198	9.311	1.53	0.1605
clinic		4	-1.3158	0.7102	13.98	-1.85	0.0851
clinic		5	-0.6166	0.6700	11.68	-0.92	0.3760
clinic		6	-1.6202	0.8083	14	-2.00	0.0648
clinic		7	-0.7298	0.7806	14	-0.93	0.3657
clinic		8	1.6790	0.7345	14	2.29	0.0384
clinic*trt	cntl	1	0.08600	0.2296	1	0.37	0.7718
clinic*trt	drug	1	-0.08969	0.2294	1	-0.39	0.7627

Solution for Random Effects							
Effect	trt	clinic	Estimate	Std Err Pred	DF	t Value	Pr > t
clinic*trt	cntl	2	0.04971	0.2323	1	0.21	0.8658
clinic*trt	drug	2	0.003155	0.2329	1	0.01	0.9914
clinic*trt	cntl	3	-0.06843	0.2326	1	-0.29	0.8178
clinic*trt	drug	3	0.09646	0.2326	1	0.41	0.7497
clinic*trt	cntl	4	-0.01749	0.2388	1	-0.07	0.9535
clinic*trt	drug	4	-0.02153	0.2382	1	-0.09	0.9426
clinic*trt	cntl	5	-0.09451	0.2378	1	-0.40	0.7591
clinic*trt	drug	5	0.07622	0.2368	1	0.32	0.8017
clinic*trt	cntl	6	-0.03299	0.2412	1	-0.14	0.9135
clinic*trt	drug	6	-0.01505	0.2404	1	-0.06	0.9602
clinic*trt	cntl	7	-0.01000	0.2401	1	-0.04	0.9735
clinic*trt	drug	7	-0.01164	0.2402	1	-0.05	0.9692
clinic*trt	cntl	8	0.08772	0.2389	1	0.37	0.7760
clinic*trt	drug	8	-0.03793	0.2394	1	-0.16	0.9000

Interpretation

As in conventional mixed models, the solutions themselves have no intrinsic meaning because they are based on a generalized inverse. However, **linear predictors**—that is, estimates of predictable functions $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$ obtained from the solution vectors of $\boldsymbol{\beta}$ and \mathbf{u} —do have meaning.

You can use the elements of the solution vector to obtain best linear unbiased predictors. For example, consider the ESTIMATE statement labeled “trt 1 clinic 1 BLUP.” This is the **broad** inference space BLUP for the CNTL treatment at clinic 1. It is defined as

$$\text{BLUP}(\text{trt } 1, \text{clinic } 1) = \mu + \tau_1 + c_1$$

That is,

$$\mathbf{K}' = [1 \ 1 \ 0] \text{ and } \mathbf{M}' = [1 \ 0 \dots 0]$$

Using the elements of the solution vector,

$$\text{BLUP}(\text{trt } 1, \text{clinic } 1) = -0.4161 - 0.7469 - 0.1246 = -1.2876$$

The linear predictor is computed on the *logit* scale. You can convert it to a predicted probability by applying the inverse link function. Here

$$\hat{\pi}_{11} = 1 / [1 + \exp\{-1.2876\}] = 0.2163$$

You can calculate BLUPs for all TRT \times CLINIC combinations using the appropriate μ , τ_i , and c_j estimates.

You can calculate the narrow inference space BLUP by adding (τ_{ij}). For example, the narrow space BLUP for treatment 1, clinic 1 is $\mu + \tau_1 + c_1 + (\tau_{ij})$.

You can compute least-squares means for TRT using the estimable function $\mu + \tau_i$. For example, for the CNTL treatment, the linear predictor, the estimate of $\eta_i = \mu + \tau_i$ is $-0.4161 - 0.7469 = -1.1630$. The corresponding mean is $g^{-1}(\eta)$, which in this case is

$$\text{MEAN} = \exp\{-1.1630\} / [1 + \exp\{-1.1630\}] = 0.2381$$

The standard errors for the MEAN predictions are computed based on the standard errors on the logit scale by applying the **Delta method**, which involves a Taylor series approximation of $g^{-1}(\eta)$. See Bishop, Fienberg, and Holland (1975) for details. The important point is that standard errors obtained by the Delta method are approximations. The formula is

$$s.e.[\mathbf{g}(\eta)] = \sqrt{\left[\frac{\partial \mathbf{g}^{-1}(\eta)}{\partial \eta} \right]^2 \mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}}$$

For example, the TRT least-squares mean has a standard error of 0.5650. This is computed from

$$\sqrt{\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}}$$

Note that $\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K} = \mathbf{L}'\mathbf{C}\mathbf{L}$ when the vector \mathbf{m} contains all zeros, as it does here. The derivative, $\partial[g^{-1}(\eta)]/\partial\eta$, was obtained in forming the \mathbf{D} matrix used in the estimating equation. For the logit link, $\partial[g^{-1}(\eta)]/\partial\eta = \pi(1-\pi)$, which is estimated by $\text{MEAN}(1-\text{MEAN})$. The resulting standard error for the probability of favorable response in the CNTL treatment group is

$$\pi(1-\pi)\sqrt{\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}} = 0.2381 \times (1 - 0.2381) \times 0.5650 = 0.102$$

Applying a similar set of calculations, the standard error for TRT DRUG is computed as 0.134. These standard errors are given in the last column of the “trt Least Squares Means” table in Output 14.2.

Caution: How to Correctly Determine a Treatment Difference

You should exercise caution when requesting inverse linking in ESTIMATE (or LSMESTIMATE) statements. The results will be meaningful only when it makes sense to apply the inverse link function to the estimate. For example, for CNTL and DRUG, the difference between the predicted probabilities is

$$\hat{\pi}_{\text{DRUG}} - \hat{\pi}_{\text{CNTL}} = 0.3975 - 0.2381 = 0.1594$$

On the other hand, if you work with the estimable function for a difference (given in Output 14.2 as TRT DIFF) using the model parameters, then $\tau_1 - \tau_2 = -0.7469$. Applying the inverse link to $\tau_1 - \tau_2$ yields

$$\exp\{-0.7469\} / [1 + \exp\{-0.7469\}] = 0.3215$$

and is likely not useful. The reason for this discrepancy is that the inverse link is nonlinear. The difference of the estimated probabilities does not equal the inversely linked difference of the linear predictors.

You can use standard errors for the difference for the logit scale, but the transformation to the difference between predicted *probabilities* is not straightforward. Furthermore, probabilities between groups are typically not compared by taking differences. The odds ratio or log odds ratio is a much more meaningful statistic. From Output 14.2, the odds ratio for the treatment difference is 0.474 with a 95% confidence interval of (0.198, 1.137). This means that the individuals in the DRUG group have twice the odds of a favorable response compared to those persons in the CNTL group. However, since the confidence bounds contain the odds ratio value of 1, there is not a significant difference between the two treatments (at the 0.05 significance level).

14.5.2 Obtaining Predicted Values

With the GLIMMIX procedure, it is easy to obtain output statistics that are computed for every observation in the data set. You accomplish this with the OUTPUT statement. Similar statements are available in other SAS/STAT procedures, such as the GLM, REG, LOGISTIC, and GENMOD procedures. Because a generalized linear mixed model contains random effects and a link function, the GLIMMIX procedure makes available some enhanced features for computing output statistics. For example, most output statistics are available in four “flavors” depending on whether the computation involves the random effect solutions (BLUPs) and whether the result is reported on the scale of the link scale or the data scale.

Program

```
proc glimmix data=clinics;
  class clinic trt;
  model fav/nij = trt / solution ddfm=satterth;
  random clinic trt*clinic;
  output out=gmxout pred(ilink noblup) = mupa
            pred(ilink blup) = mu
            pred(noilink noblup) = linppa
            pred(noilink blup ) = linp;
run;
proc print data=gmxout;
run;
```

The four types of each output statistic are requested by adding the ILINK / NOILINK and BLUP / NOBLUP options to the keyword for the output statistic. The variable MUPA in the output data set, for example, contains the predicted values based on only the fixed effects, mapped onto the probability scale. That is,

$$\text{MUPA}_i = \frac{1}{1 + \exp\{-\mathbf{x}'_i \hat{\boldsymbol{\beta}}\}}$$

Table 14.5 shows the possible combinations for the PRED keyword.

Table 14.5 Predicted Values with OUTPUT Statement of PROC GLIMMIX

	BLUP	NOBLUP
ILINK	$\frac{1}{1 + \exp\{-\mathbf{x}' \hat{\boldsymbol{\beta}} - \mathbf{z}' \hat{\boldsymbol{u}}\}}$	$\frac{1}{1 + \exp\{-\mathbf{x}' \hat{\boldsymbol{\beta}}\}}$
NOILINK	$\mathbf{x}' \hat{\boldsymbol{\beta}} + \mathbf{z}' \hat{\boldsymbol{u}}$	$\mathbf{x}' \hat{\boldsymbol{\beta}}$

Results

Output 14.4 Binomial Example—LOGIT Link—Predictions

Obs	clinic	trt	fav	unfav	nij	mupa	mu	linppa	linp
1	1	drug	11.0000	25	36	0.39745	0.34743	-0.41610	-0.63035
2	1	cntl	10.0000	27	37	0.23813	0.23120	-1.16297	-1.20153
3	2	drug	16.0000	4	20	0.39745	0.79735	-0.41610	1.36981
4	2	cntl	22.0000	10	32	0.23813	0.66139	-1.16297	0.66949
5	3	drug	14.0000	5	19	0.39745	0.65151	-0.41610	0.62567
6	3	cntl	7.0000	12	19	0.23813	0.42896	-1.16297	-0.28610
7	4	drug	2.0000	14	16	0.39745	0.14761	-0.41610	-1.75345
8	4	cntl	1.0000	16	17	0.23813	0.07612	-1.16297	-2.49628
9	5	drug	6.0000	11	17	0.39745	0.27758	-0.41610	-0.95650
10	5	cntl	0.0083	12	12	0.23813	0.13307	-1.16297	-1.87410
11	6	drug	1.0000	10	11	0.39745	0.11391	-0.41610	-2.05140
12	6	cntl	0.0100	10	10	0.23813	0.05645	-1.16297	-2.81621
13	7	drug	1.0000	4	5	0.39745	0.23912	-0.41610	-1.15753
14	7	cntl	1.0000	8	9	0.23813	0.12980	-1.16297	-1.90277
15	8	drug	4.0000	2	6	0.39745	0.77293	-0.41610	1.22494
16	8	cntl	6.0000	1	7	0.23813	0.64651	-1.16297	0.60373

Interpretation

The variables MUPA and LINPPA take on only two values, depending on whether an observation belongs to the CNTL or the DRUG group. The linear predictors for the two groups are, respectively, -0.4161 and $-0.4161 - 0.7469 = -1.163$. The associated probabilities are $MUPA_{CNTL} = 1/1+\exp\{0.4161\} = 0.39745$ and $MUPA_{DRUG} = 1/1+\exp\{1.1623\} = 0.2381$.

For the variables MU and LINP, the η term also includes the BLUP for the particular observation. For example, the first observation is associated with the DRUG treatment in clinic 1. The relevant random effects solutions for this level combination can be found in Output 14.3 as $c_1 = -0.1246$, $(\tau c_{21}) = -0.08969$. Putting these BLUPs together with the fixed effects solutions yields $LINP_1 = -0.4161 - 0.1246 - 0.0897 = -0.6304$ and $MU_1 = 1/1+\exp\{0.6304\} = 0.3474$.

14.5.3 Alternative Link Functions

Previous sections have considered only canonical link functions. In many generalized linear models, the link function may be suggested by something other than the natural parameter. Often, the link function represents a physical or biological model of the mean. For example, binomial responses are often assumed to be observable manifestations of unobservable underlying, continuous processes. Imagine there is a continuous (normally distributed) variable. When it is below a certain **threshold**, the observable response is unfavorable. When the unobservable process is at or above the threshold, a favorable response is observed. The probability of a favorable response can be determined from the normal c.d.f.

Model Using Probit Link

Assume that $\mathbf{X}\beta + \mathbf{Z}\mathbf{u}$ is a generalized linear mixed model of the underlying process. Then the probability, π , is

$$\pi = \Phi(\mathbf{X}\beta + \mathbf{Z}\mathbf{u})$$

Thus, the link function is

$$\eta = \Phi^{-1}(\mu)$$

where $\Phi(\eta)$ is the normal c.d.f. evaluated at η . The link function, $\eta = \Phi^{-1}(\mu)$, is called the **probit link**. Probit analysis is based on a generalized linear model using the probit link.

You can fit a generalized linear mixed model for the treatment-clinic example using the probit link function instead of the logit link. The model, $\mu + \tau_i + c_j + (\tau c)_{ij}$, remains the same as in the previous analysis. The variance function and scale parameter, used in \mathbf{R}_μ and \mathbf{A} , also remains the same. \mathbf{D} and $\boldsymbol{\eta}$ are adjusted accordingly to reflecting the new link function.

Program for Analysis Using Probit Link

You can obtain an analysis of the probit mixed model using the following GLIMMIX statements:

```
proc glimmix data=clinics;
  class clinic trt;
  model fav/nij = trt / solution ddfm=satterth
    link=probit;
  random clinic trt*clinic / solution;
  lsmeans trt / ilink;
  estimate 'trt diff' trt 1 -1 / cl;
  estimate 'trt 1 clinic 1 BLUP' intercept 1 trt 1 0
    | clinic 1 0 / ilink;
  estimate 'trt 1 clinic 6 BLUP' intercept 1 trt 1 0
    | clinic 0 0 0 0 1 0 / ilink;
run;
```

The only change from the previous analysis is the LINK=PROBIT option in the MODEL statement. Because of the events/trials syntax, PROC GLIMMIX defaults to the binomial distribution. The LINK=PROBIT option changes the link from the logit default.

Results from this analysis are given in Outputs 14.5 and 14.6.

Results

Output 14.5 Binomial Example—PROBIT Link

Fit Statistics	
-2 Res Log Pseudo-Likelihood	34.97
Generalized Chi-Square	13.58
Gener. Chi-Square / DF	0.97

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
clinic	0.7194	0.4447
clinic*trt	0.01886	0.06863

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	4.958	5.48	0.0668

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
trt diff	-0.4487	0.1917	4.958	-2.34	0.0668	0.05	-0.9428	0.04532
trt 1 clinic 1 BLUP	-0.7837	0.2062	2.603	-3.80	0.0408	0.05	-1.5003	-0.06713
trt 1 clinic 6 BLUP	-1.6743	0.3990	14	-4.20	0.0009	0.05	-2.5301	-0.8185

Estimates				
Label	Mean	Standard Error Mean	Lower Mean	Upper Mean
trt diff				
trt 1 clinic 1 BLUP	0.2166	0.06050	0.06677	0.4732
trt 1 clinic 6 BLUP	0.04703	0.03919	0.005701	0.2065

trt Least Squares Means							
Trt	Estimate	Standard Error	DF	t Value	Pr > t	Mean	Standard Error Mean
Cntl	-0.6999	0.3343	7.912	-2.09	0.0700	0.2420	0.1044
Drug	-0.2511	0.3326	7.789	-0.75	0.4725	0.4009	0.1286

Interpretation

The major features of the analysis are similar to the analysis using the logit link. The main differences reflect the different link function. The main points are as follows:

- The variance component estimates of σ_c^2 (0.7194) and σ_{tc}^2 (0.01886) are similar, relatively, to those obtained using the logit link. The difference in magnitude results from the use of a different link. In the probit model, the variance components have an interpretation. They reflect the variance of the CLINIC and TRT \times CLINIC effects on the underlying, normally distributed threshold process described above.

- The extra-dispersion parameter estimate is 0.97. This is similar to the estimate of ϕ for the logit link. There is no evidence of underdispersion for the probit model.
- The F -statistic to test the equality of treatment effects is 5.48. This is similar to the F -value obtained using the logit link. In this case, you would reach the same conclusion about treatments. Clearly, given the *right* set of data, it is possible to reach different conclusions. McCullagh and Nelder (1989) discuss procedures you can use to check the adequacy of the model assumptions you have made, including your choice of link function. We strongly recommend that you use their procedures when fitting any generalized linear model.
- The estimates and least-squares means are expressed on the probit scale. These are obtained in the same way as shown in Section 14.5.1, except that you use the solutions fitting the probit model. These are given in Output 14.6.

Results—Fixed and Random Effect Solutions for Probit Model

Output 14.6 Solution for Probit Model

Solutions for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-0.2511	0.3326	7.789	-0.75	0.4725
trt	cntl	-0.4487	0.1917	4.958	-2.34	0.0668
trt	drug	0

Solution for Random Effects							
Effect	trt	clinic	Estimate	Std Err Pred	DF	t Value	Pr > t
clinic		1	-0.08384	0.3541	8.327	-0.24	0.8185
clinic		2	1.0848	0.3650	9.198	2.97	0.0153
clinic		3	0.5728	0.3705	9.824	1.55	0.1537
clinic		4	-0.7748	0.4058	13.11	-1.91	0.0784
clinic		5	-0.3947	0.3948	11.91	-1.00	0.3373
clinic		6	-0.9745	0.4555	14	-2.14	0.0505
clinic		7	-0.4408	0.4531	14	-0.97	0.3472
clinic		8	1.0109	0.4378	14	2.31	0.0367
clinic*trt	cntl	1	0.04728	0.1298	1	0.36	0.7776
clinic*trt	drug	1	-0.04948	0.1298	1	-0.38	0.7681
clinic*trt	cntl	2	0.02721	0.1312	1	0.21	0.8698
clinic*trt	drug	2	0.001236	0.1315	1	0.01	0.9940
clinic*trt	cntl	3	-0.03791	0.1317	1	-0.29	0.8216
clinic*trt	drug	3	0.05293	0.1317	1	0.40	0.7568
clinic*trt	cntl	4	-0.00689	0.1341	1	-0.05	0.9673

Solution for Random Effects							
Effect	trt	clinic	Estimate	Std Err Pred	DF	t Value	Pr > t
clinic*trt	drug	4	-0.01343	0.1339	1	-0.10	0.9364
clinic*trt	cntl	5	-0.05315	0.1339	1	-0.40	0.7594
clinic*trt	drug	5	0.04280	0.1335	1	0.32	0.8024
clinic*trt	cntl	6	-0.01839	0.1355	1	-0.14	0.9141
clinic*trt	drug	6	-0.00716	0.1352	1	-0.05	0.9663
clinic*trt	cntl	7	-0.00476	0.1352	1	-0.04	0.9776
clinic*trt	drug	7	-0.00680	0.1353	1	-0.05	0.9680
clinic*trt	cntl	8	0.04661	0.1348	1	0.35	0.7881
clinic*trt	drug	8	-0.02010	0.1350	1	-0.15	0.9059

Interpretation

The broad space BLUP of the CNTL treatment for clinic 1, $\mu + \tau_1 + c_1$, has a linear predictor of

$$\hat{\eta}_{11} = -0.2511 - 0.4487 - 0.0838 = -0.7836$$

Applying the inverse link,

$$\hat{\pi}_{11} = \Phi(-0.7836) = 0.2166$$

You can obtain other BLUPs in a similar fashion. Again, the only difference between the procedure used here and the one used in the logit model is the inverse link function. As in the case of the logit model, the inverse link is not meaningful when applied to differences.

14.6 Example: Count Data in a Split-Plot Design

Section 14.2.2 described a split-plot experiment with different management methods as the whole-plot treatment factor and different seed mixes as the split-plot treatment factor. The whole plots were arranged in randomized complete blocks. Botanical composition was measured, expressed as the number of plants of a given species per split-plot experimental unit.

The data for this example are given as Data Set 14.6, “Seed Mix,” in Appendix 2, “Data Sets.” The variables are TRT (management method—7 types), BLK (4 blocks), MIX (seed mix—4 types), and COUNT (the number of plants of the species of interest).

14.6.1 Standard Analysis of Poisson GLMM

Model

Model 14.2 can be modified to accommodate the classical assumption that COUNT has a Poisson distribution. The generalized linear mixed model for these data is

$$\eta_{ijk} = \log(\lambda_{ijk}) = \mu + r_i + \tau_j + (r\tau)_{ij} + \delta_k + (\tau\delta)_{ik}$$

where

- λ_{ijk} is the conditional mean count given the random effects
- μ is the intercept
- r_i is the BLK effect
- τ_j is the TRT effect
- $(r\tau)_{ij}$ is the BLK \times TRT (whole-plot error) effect
- δ_k is the MIX effect
- $(\tau\delta)_{ik}$ is the TRT \times MIX interaction

Assumptions for the model are as follows:

- BLK and BLK \times TRT are random effects
- $r_i \sim iid N(0, \sigma_R^2)$
- $(r\tau)_{ij} \sim iid N(0, \sigma_{RT}^2)$

As in Section 14.5, you need to determine the various terms needed to work with the model assuming a Poisson distribution of the data. These are summarized in Table 14.6.

Table 14.6 Components of the Poisson Likelihood

Expression	Function	Value
Conditional mean	μ_{ijk}	λ_{ijk}
Natural parameter	$\theta(\mu_{ijk})$	$\log(\lambda_{ijk})$
Variance function	$V(\mu_{ijk})$	λ_{ijk}
Scale parameter	$a(\phi_{ijk})$	1
Inverse link	$g^{-1}(\eta_{ijk})$	$\exp(\eta_{ijk})$

Program to Analyze Count Data

You can obtain an analysis of this model with the GLIMMIX procedure using the following program:

```
proc glimmix data=mix;
  class trt blk mix;
  model y = trt mix trt*mix / dist=poisson;
  random blk blk*trt;
run;
```

The DIST=POISSON option in the MODEL statement instructs the GLIMMIX procedure that the data come from a Poisson distribution—conditional on the random BLK and BLK \times TRT effects. The default link for the Poisson distribution in PROC GLIMMIX is the log link. The results of the analysis are given in Output 14.7.

Results

Output 14.7 Example with Counts—Default Analysis—Selected Results

Iteration History					
Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	4	464.22346984	2.00000000	52.53776
1	0	4	570.57465492	0.42176341	57.62868
2	0	2	590.8991972	0.02468174	57.77231
3	0	1	591.7658988	0.00007935	57.76544
4	0	1	591.76851777	0.00000026	57.76541
5	0	0	591.76851876	0.00000000	57.76541

Convergence criterion (PCONV=1.11022E-8) satisfied.

Estimated G matrix is not positive definite.

Fit Statistics		
-2 Res Log Pseudo-Likelihood		591.77
Generalized Chi-Square		604.28
Gener. Chi-Square / DF		7.19

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
blk	6.69E-20	.
trt*blk	0.1493	0.05059

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	6	18	3.55	0.0170
mix	3	63	14.80	<.0001
trt*mix	18	63	10.24	<.0001

Interpretation

Most aspects of interpretation are similar to those discussed in the binomial example in Section 14.5. The two main features of Output 14.7 are that the variance matrix of the random effects (the **G** matrix) is not positive definite and that the ratio of the Generalized Chi-Square statistic and its degrees of freedom is considerably larger than 1 (see the “Fit Statistics” table). The message about the **G** matrix not being positive definite is caused by the estimate of the variance of the BLK effect being (practically) zero. You can go ahead and interpret the output, or you can refit the model by dropping BLK.

The large estimate of the residual dispersion (7.19) is not necessarily indicative of an overdispersion problem that needs to be corrected through the conditional distribution or the variance function. The value of 7.19 in the “Fit Statistics” table represents the estimate of the residual variance in the final pseudo-model. We will see in Section 14.6.2 that the addition of an extra scale parameter can increase this residual variance.

Some insight into whether an overdispersion adjustment is needed for the conditional distribution can be gained by examining the Pearson-type residuals in PROC GLIMMIX. The following program computes sample statistics for these residuals using the BLUPs—that is, mimicking the residuals in the conditional distribution.

Program

```
proc glimmix data=mix;
  class trt blk mix;
  model y = trt mix trt*mix / dist=poisson s;
  random blk*trt;
  output out=gmxout pearson(blup)=res;
run;
proc means data=gmxout mean std; var res;
run;
```

Result

Output 14.8 Standard Deviation of Conditional Pearson Residuals

Analysis Variable : res Pearson Residual	
Mean	Std Dev
-0.0065440	2.2952295

Interpretation

The Pearson statistic for the conditional distribution is $2.295^2 = 5.26$, smaller than the estimated residual dispersion in the marginal distribution (7.19), but considerably larger than 1. An adjustment of the model is necessary. The important question, of course, is which model component to adjust. The covariance model may be incorrect, one or more important fixed effects covariates may have been omitted, the data may not be Poisson distributed, etc. The simple adjustment is to include a multiplicative overdispersion factor on the variance function. The following analysis takes this approach.

14.6.2 Analysis of Poisson GLMM with Overdispersion

Program

```

ods html;
ods graphics on;
proc glimmix data=mix;
  class trt blk mix;
  model y = trt mix trt*mix / dist=poisson;
  random blk blk*trt;
  random _residual_;
  lsmeans trt / plots=diffplot adjust=tukey;
  lsmeans trt / plots=anomplot adjust=nelson;
run;
ods graphics off;
ods html close;

```

The overdispersion parameter is included in the analysis with the RANDOM _RESIDUAL_ statement; the _RESIDUAL_ keyword of the RANDOM statement instructs the GLIMMIX procedure that this is an R-side random effect—that is, a random component of the **R** matrix.

The LSMEANS statement requests the least-squares means for the TRT effect and differences of the least-squares means. The computation of the differences is automatically triggered by the request of least-squares means graphics. The first LSMEANS statement requests a least-squares mean by least-squares mean scatter plot (**Diffogram**). If the plot is based on arithmetic means, it is also known as a mean-mean scatter plot (Hsu 1996, Hsu and Peruggia 1994). The $(7 \times 6)/2 = 21$ pairwise comparisons are adjusted for multiplicity with Tukey's method. The Diffogram is shown in Figure 14.1. The second LSMEANS statement requests an **analysis of means** (ANOM) in which each treatment is compared against the average of all treatments (Ott, 1967; Nelson, 1982, 1991, 1993). This leads to seven comparisons, which are graphically displayed in Figure 14.2. Using Nelson's adjustment these comparisons are also adjusted for multiplicity.

Results

Output 14.9 Analysis of a Poisson GLMM with Overdispersion

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

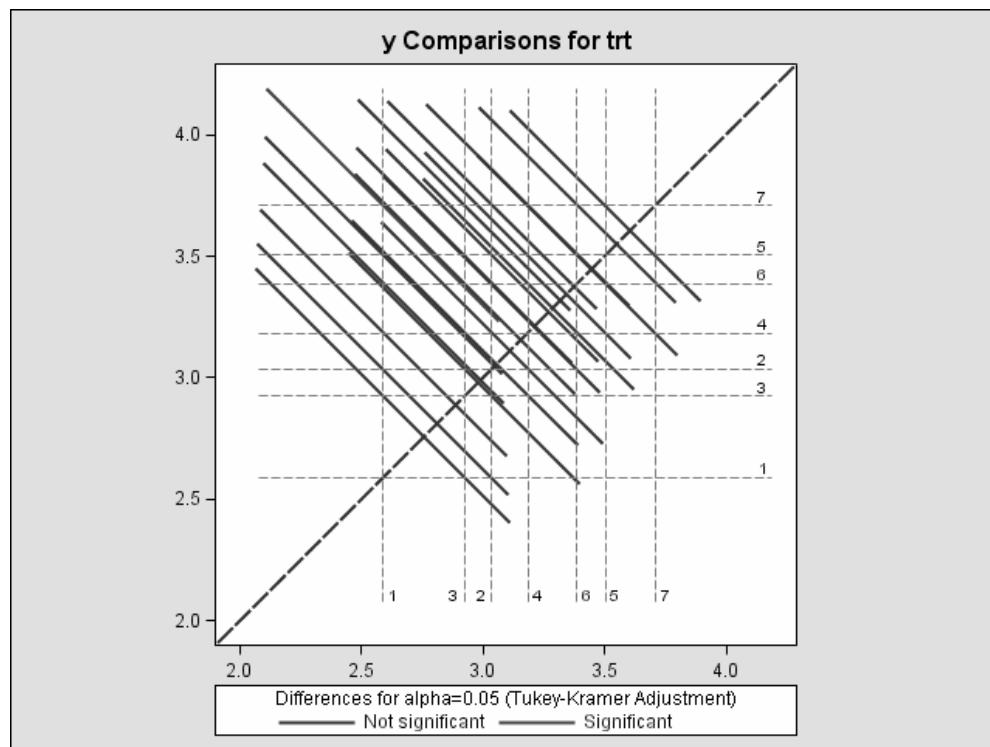
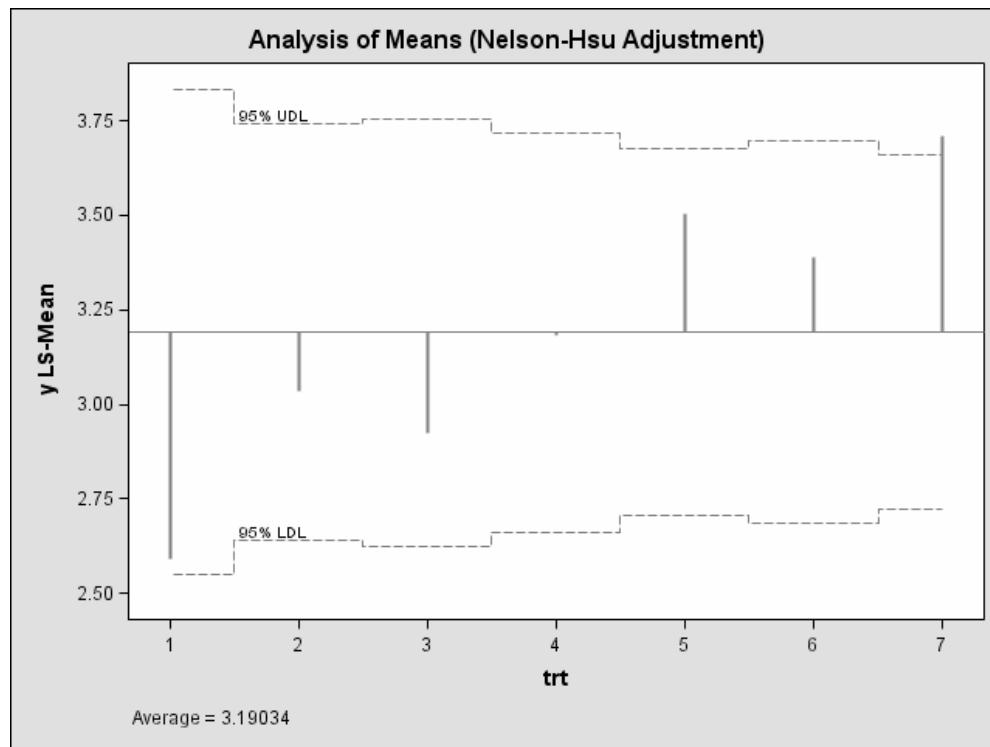
Estimated G matrix is not positive definite.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	204.50
Generalized Chi-Square	768.18
Gener. Chi-Square / DF	9.14

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
blk	0	.
trt*blk	0.04763	0.04399
Residual (VC)	9.1450	1.6211

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	6	18	3.65	0.0150
mix	3	63	1.62	0.1940
trt*mix	18	63	1.12	0.3558

trt Least Squares Means					
trt	Estimate	Standard Error	DF	t Value	Pr > t
1	2.5894	0.2394	18	10.81	<.0001
2	3.0363	0.2008	18	15.12	<.0001
3	2.9237	0.2068	18	14.13	<.0001
4	3.1836	0.1905	18	16.71	<.0001
5	3.5038	0.1716	18	20.42	<.0001
6	3.3866	0.1803	18	18.79	<.0001
7	3.7092	0.1637	18	22.66	<.0001

Figure 14.1 Diffogram for TRT Effect**Figure 14.2** Analysis of Means for TRT Effect

Interpretation

The “Optimization Information Table” shows that there is a residual variance in the estimation process. It is designated as an R-side parameter, since it modifies the **R** matrix of the generalized linear mixed model. It is also profiled from the analysis, because an estimate of ϕ can be obtained in closed form. The two parameters in the optimization are the variance components for the BLK and the BLK \times TRT effects.

As in the earlier analysis, the variance matrix of the random effects is not positive definite, and the variance of the BLK effect is zero (“Covariance Parameter Estimates” table). It is interesting to compare the estimates of the covariance parameters to those obtained in Section 14.6.1. There it was found that a model without overdispersion parameter suggests an estimate of 7.19. This does not mean that adding an overdispersion parameter yields an estimate of 7.19, because a different marginal distribution results in the current model. The presence of ϕ in the model changes the magnitude of the TRT \times BLK variance (from 0.1493 to 0.0476). The estimate of ϕ is 9.145 and is shown as ‘Residual (VC)’ in the “Covariance Parameter Estimates” table. Because it is computed as $\mathbf{r}'(\mathbf{V}^*)^{-1}\mathbf{r}/df$, it also equals the ratio of the generalized χ^2 statistic and its degrees of freedom in the “Fit Statistics” table.

This phenomenon is fairly common in modeling generalized linear mixed models. Accounting for overdispersion through a free scale parameter ϕ reduces the variability of the G-side random effects. The random effects also create overdispersion relative to a Poisson model with only TRT, MIX, and TRT \times MIX as fixed effects. The two mechanisms to account for overdispersion—random effects and changing the variance function—are in competition for the variability in the data. It only appears that adding the scale parameter ϕ made matters worse, because the estimate of 9.145 is larger than the estimate in the model with ϕ held fixed at 1. But the variance of the Pearson residuals in the current model is 0.69 (compared to 5.26 in the previous model).

The *F*-test for the fixed effects in the “Type III Test of Fixed Effects” table shows a significant treatment effect, but no effects of MIX or an interaction. Compared to the analysis in Section 14.6.1, the test for the TRT effect appears to be affected minimally by the inclusion of ϕ in the model. Conclusions about the MIX effect and the MIX \times TRT interaction changed dramatically, however. This is further indication that adding an overdispersion parameter in a generalized linear mixed model does *not* simply change standard errors in a proportional manner, as is the case in generalized linear models. The involvement of ϕ in the marginal variance matrix, which determines the precision of the fixed effects estimates, is much more complicated. If you compare the standard errors of the fixed effects solutions in this model and that from Section 14.6.1, you will find that the standard errors of the coefficients associated with the TRT effect have changed much less compared to coefficients involving the MIX effect.

Figures 14.1 and 14.2 summarize the least-squares means analysis performing all pairwise differences and an analysis of means with multiplicity adjustments. The Diffogram in Figure 14.1 displays a line for each comparison. The axes of the plot represent the scale of the least-squares means. The lines are centered at the intersection of the two means. The length of the line reflects the confidence limits for the LS-mean difference; this length is adjusted for the rotation and possibly adjusted for multiplicity (as in this example). If two least-squares means are significantly different, the line will not cross the 45-degree reference line of the plot. Interestingly, if the comparisons are adjusted for multiplicity, only the first and last levels of TRT are significantly different; only the line centered at the intersections of levels 1 and 7 fails to cross the 45-degree reference line. These two levels are the most extreme with least-squares means (on the log scale) of 2.5894 and 3.7092.

The analysis of means in Figure 14.2 displays tests about least-squares means differently. In an analysis of means, the TRT levels are not compared against each other, but against an overall average. In this case the overall average TRT effect (on the log scale) is 3.19034 (Figure 14.2). The dashed horizontal step plots in the analysis of means graph represent the upper and lower decision limits ('UDL', 'LDL'). If a vertical line crosses a decision limit, the corresponding level is significantly different from the average. This is the case only for TRT 7. It is now also obvious why the comparison of the first and seventh treatment was significant in the Diffogram. They are most extreme on opposite sides of the average.

Failing to Account for Whole-Plot Error

The main advantages of the generalized linear mixed model are that you can account for the distribution of the count data using the appropriate generalized linear model and that you can account for the split-plot nature of the experiment by fitting the BLK \times TRT random effect and by computing F -statistics and standard errors accordingly. Conventional generalized linear model programs (e.g., PROC GENMOD) do not provide for random effects, and conventional mixed models do not provide for nonnormal errors.

You can obtain a conventional generalized linear model analysis to evaluate the effect of failing to account for whole-plot error (BLK \times TRT) by deleting one RANDOM statement from the GLIMMIX program:

```
proc glimmix data=mix;
  class trt blk mix;
  model y = trt mix trt*mix / dist=poisson;
  lsmeans trt;
  random _residual_;
run;
```

The results are given in Output 14.10.

Output 14.10 Generalized Linear Model Analysis Ignoring Whole-Plot Error

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	6	84	4.75	0.0003
mix	3	84	1.40	0.2475
trt*mix	18	84	0.97	0.5001

trt Least Squares Means					
trt	Estimate	Standard Error	DF	t Value	Pr > t
1	2.5899	0.2288	84	11.32	<.0001
2	3.0454	0.1805	84	16.87	<.0001
3	2.9271	0.1885	84	15.53	<.0001
4	3.1889	0.1674	84	19.04	<.0001
5	3.5104	0.1419	84	24.75	<.0001

trt Least Squares Means					
trt	Estimate	Standard Error	DF	t Value	Pr > t
6	3.4017	0.1533	84	22.19	<.0001
7	3.7123	0.1309	84	28.36	<.0001

Interpretation

You can see that the resulting F -statistics are different. The whole-plot effect (TRT for these data) is particularly affected. In this case, the conclusion is not affected. You conclude that TRT effects are significant in both analyses. It is easy to see, however, that for the *right data* you can easily reach erroneous conclusions if you do not account for whole-plot error.

14.7 Summary

This chapter showed how to use the GLIMMIX procedure to analyze generalized linear mixed models. The generalized linear mixed model is an extension of the generalized linear model and the linear mixed model to accommodate nonnormal distributions with normal random effects. Section 14.2 presented two examples illustrating the need for such models. Section 14.3 gave an introduction to generalized linear models—fixed effects nonnormal error linear models. Section 14.4 showed how random effects are incorporated to produce generalized linear mixed models. Sections 14.5 and 14.6 presented analyses of the two examples discussed in Section 14.2 using the GLIMMIX procedure. Section 14.5 is a binomial example; Section 14.6 is a Poisson example. Sections 14.5 and 14.6 focused on the required SAS programs and interpretation of the output.



Nonlinear Mixed Models

15.1	Introduction	568
15.2	Background on PROC NL MIXED	569
15.3	Example: Logistic Growth Curve Model.....	571
15.3.1	PROC NL MIXED Analysis	571
15.3.2	%NLINMIX Macro Analysis—Method 1	575
15.3.3	%NLINMIX Macro Analysis—Method 2	581
15.3.4	%NLINMIX Macro Analysis—Method 3, With and Without Variance Weighting	581
15.3.5	Choosing the Exponent for the Weight Matrix That Leads to the Best Fit	586
15.3.6	Comparing the Logistic Model Fitting Methods	586
15.4	Example: Nested Nonlinear Random Effects Models.....	587
15.5	Example: Zero-Inflated Poisson and Hurdle Poisson Models	589
15.6	Example: Joint Survival and Longitudinal Model.....	595
15.7	Example: One-Compartment Pharmacokinetic Model	607
15.8	Comparison of PROC NL MIXED and the %NLINMIX Macro ..	623
15.9	Three General Fitting Methods Available in the %NLINMIX Macro ..	625
15.10	Troubleshooting Nonlinear Mixed Model Fitting.....	629
15.10.1	General Considerations	629
15.10.2	PROC NL MIXED Tips	630
15.10.3	%NLINMIX Macro Tips	634
15.11	Summary.....	634

15.1 Introduction

In Chapters 1 through 13, the models are all linear, with the response variable modeled directly by linear combinations of fixed and random effects. The resulting models have a common form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (15.1)$$

In Chapter 14, linear combinations of fixed and random effects are used to model the **link function** of the mean of a nonnormal response variable. The resulting generalized linear mixed model is

$$E[\mathbf{Y} | \mathbf{u}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) \quad (15.2)$$

However, not all data can be adequately characterized by linear or generalized linear models. Many applications require **nonlinear models**—that is, models whose parameters enter the model individually and nonlinearly.

Traditional nonlinear models have the general form

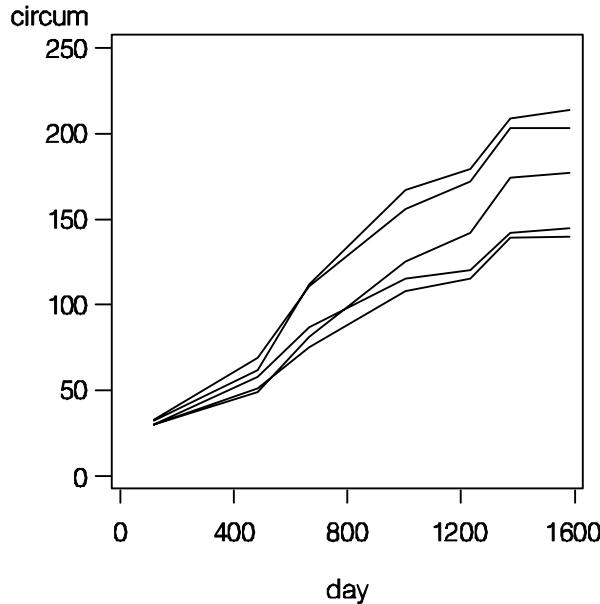
$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{e} \quad (15.3)$$

where f is a nonlinear function of known constants (\mathbf{X}) and unknown parameters ($\boldsymbol{\beta}$) and the errors are additive. Such nonlinear models can still be considered as **fixed effects models**. This chapter shows you how to extend these models to accommodate **random effects** as well.

For example, suppose your experimental data are in the form of repeated measurements on response and explanatory variables from several subjects, and you would like to fit a model that simultaneously accounts for the overall nonlinear mean structure as well as the variability between and within subjects. This situation calls for a **nonlinear mixed model**.

To be more specific, consider the orange tree data from Draper and Smith (1981, p. 524) and Lindstrom and Bates (1990). Figure 15.1 plots these data, which are trunk circumferences of five orange trees measured at seven time points. The curves display a mildly nonlinear S-shaped growth trend as well as a marked increase in variability over time.

One modeling approach to such data is to fit a higher-order polynomial trend along with a heterogeneous variance-covariance structure like those discussed in Chapter 9. While this is a potentially viable linear model, scientific experience indicates that the logistic growth model $E[Y] = \beta_1 / (1 + \beta_2 e^{\beta_3 x})$ is more interpretable and appropriate. Furthermore, you want to account for both inter-subject variability and intra-subject correlation and heterogeneity.

Figure 15.1 Orange Tree Data

The primary questions that arise are **How should I construct an appropriate nonlinear mixed model?** and **How should I fit it to the data?** Many answers to these questions have been discussed at length in the literature. The books by Vonesh and Chinchilli (1997) and Davidian and Giltinan (1995) provide good overviews as well as general theoretical developments and examples of nonlinear mixed models. Rather than discuss all of the possibilities, this chapter focuses on two primary software modules supported by SAS: the **NLMIXED** procedure and the **%NLINMIX macro**. The former is a relatively recent addition to SAS/STAT and should be your first choice for fitting nonlinear mixed models. The latter has a longer history, and is based on iteratively fitting linearized models using PROC MIXED.

This chapter focuses primarily on examples of how to use PROC NLMIXED, although some of the same examples are also fitted with the %NLINMIX macro. Later sections provide a comparison between the two programs as well as the theoretical details behind the macro. The following section provides a brief background on PROC NLMIXED. Please refer to the PROC NLMIXED documentation in the *SAS/STAT User's Guide* for more details about the specific syntax and computational algorithms behind it. The remainder of this chapter covers a set of relevant examples, as well as further details about the %NLINMIX macro and the NLMIXED procedure and some troubleshooting tips.

15.2 Background on PROC NLMIXED

PROC NLMIXED fits nonlinear mixed models by numerically maximizing an approximation to the *marginal likelihood*—that is, the likelihood integrated over the random effects. Different integral approximations are available, the primary one being adaptive Gaussian quadrature (Pinheiro and Bates 1995). This approximation uses the empirical Bayes estimates of the

random effects (analogous to empirical BLUPs in linear mixed models) as the central point for the quadrature, and updates them for every iteration. This approach can be much more efficient than quadrature over a static grid, and in fact, using just a single adaptive quadrature point (also known as Laplace's approximation) you can typically obtain very satisfactory results. A less accurate but often more stable approximation is to expand the likelihood function in a Taylor series around zero; this is the basis for the well-known first-order method of Beal and Sheiner (1982, 1988) and Sheiner and Beal (1985), available via METHOD=FIRO in PROC NLMIXED. Expanding around the empirical Bayes estimates is discussed in more detail by Beal and Sheiner (1992), Wolfinger (1993b), Vonesh (1992, 1996), Vonesh and Chinchilli (1997), and Wolfinger and Lin (1997).

The resulting marginal likelihood can be maximized using a variety of alternative optimization techniques; the default is a dual quasi-Newton algorithm. PROC NLMIXED maximizes the marginal likelihood function directly using numerical methods, whereas the %NLINMIX macro (discussed in later sections) iteratively solves a set of estimating equations. Successful convergence of the optimization problem results in maximum likelihood (not REML) parameter estimates along with their approximate standard errors based on the second derivative matrix of the likelihood function. PROC NLMIXED enables you to use the estimated model to construct predictions of arbitrary functions using empirical Bayes estimates of the random effects. You can also estimate arbitrary functions of the nonrandom parameters, and PROC NLMIXED computes their approximate standard errors using the delta method. Nonlinear mixed models have important applications in a wide variety of applications, and are a very effective way to model correlated data with a nonlinear relationship between independent and dependent variables. One area where they are used extensively is in pharmacokinetics. Yuh et al. (1994) provide an extensive bibliography on nonlinear mixed models and their use in pharmacokinetics, and Roe (1997) provides a wide-ranging comparison of many popular techniques.

The models fit by PROC NLMIXED can be viewed as generalizations of the random coefficient models fit by PROC MIXED. This generalization allows the random coefficients to enter the model nonlinearly, whereas they enter linearly in PROC MIXED. With PROC MIXED you can perform both maximum likelihood and restricted maximum likelihood (REML) estimation, whereas PROC NLMIXED implements only maximum likelihood. This is because the analog to the REML method in PROC NLMIXED would involve a high-dimensional integral over all of the fixed effects parameters, and this integral is typically not available in closed form. Finally, PROC MIXED assumes the data to be normally distributed, whereas PROC NLMIXED enables you to analyze data that are normal, binomial, or Poisson or that have any likelihood programmable with SAS statements.

PROC NLMIXED does not implement the same estimation techniques available with the %NLINMIX macro or the GLIMMIX procedure. The latter are based on the estimation methods of Lindstrom and Bates (1990), Breslow and Clayton (1993), and Wolfinger and O'Connell (1993), and they iteratively fit a set of generalized estimating equations (refer to Chapter 14 and Wolfinger 1997). In contrast, PROC NLMIXED directly maximizes an approximate integrated likelihood. For further details, see the PROC NLMIXED documentation in the *SAS/STAT User's Guide*.

15.3 Example: Logistic Growth Curve Model

As a first example, consider the orange tree data described in the Section 15.2. The following program creates the data set listed as Data Set 15.3, “Orange Trees,” in Appendix 2, “Data Sets.”

```
data tree;
  input tree time x y;
  datalines;
  ...datalines...
run;
```

Consider the following logistic random coefficient model for the j^{th} ($j = 1, \dots, 7$) observation on the i^{th} ($i = 1, \dots, 5$) tree:

$$Y_{ij} = \frac{\beta_1 + u_{i1}}{1 + (\beta_2 + u_{i2}) \exp\{-x_{ij}/\beta_3\}} + e_{ij}$$

This is a classic model for S-shaped growth patterns. The three fixed effect parameters to be estimated are β_1 (a population upper asymptote), β_2 (a shape parameter), and β_3 (a time scale parameter). The two random effect parameters u_{i1} and u_{i2} , are assumed to be the i^{th} independent realization from a multivariate distribution with mean zero and 2×2 unstructured covariance matrix \mathbf{G} . The random effects allow each tree to have their own subject-specific asymptote and shape centered at β_1 and β_2 , respectively. The residual errors e_{ij} are assumed to be independent and identically distributed random variables with mean zero and variance σ^2 .

15.3.1 PROC NL MIXED Analysis

The following code fits the preceding model:

```
proc nlmixed data=tree method=firo;
  parms b1=190 b2=10 b3=100 s2u1=500 c12=0 s2u2=10 s2e=50;
  bounds s2u1 >= 0, s2u2 >= 0, s2e > 0;
  num = b1+u1;
  ex = exp(-day/b3);
  den = 1 + (b2+u2)*ex;
  model circum ~ normal(num/den,s2e);
  random u1 u2 ~ normal([0,0],[s2u1,c12,s2u2]) subject=tree;
run;
```

The PROC NL MIXED statement invokes the procedure. The METHOD=FIRO option requests the first-order method, popularized by Beal and Sheiner (1982). This optimizes a first-order Taylor series approximation to the marginal likelihood function. The PARMS statement specifies starting values. These values were obtained in this case by consideration of the functional form of the model and visual inspection of the plot. See Section 15.11 on troubleshooting for more discussion of starting values. The BOUNDS statement specifies boundary constraints on the variance component estimates.

The next three lines of code use SAS language to specify the logistic model in terms of a numerator and denominator.

The MODEL statement specifies the probability model for the response variable, conditional on the random effects. Here the logistic mean model is declared along with normal errors with variance S2E.

The RANDOM statement specifies the normal distribution for the two normal random coefficients U1 and U2. The SUBJECT=TREE specification is important, as it indicates distinct realizations of the random effects, just like the SUBJECT= option in the RANDOM statement of PROC MIXED.

Results

PROC NLMIXED produces the results in Output 15.1.

Output 15.1 Output for PROC NLMIXED Fit to the Orange Tree Data

Specifications	
Data Set	WORK.TREE
Dependent Variable	circum
Distribution for Dependent Variable	Normal
Random Effects	u1 u2
Distribution for Random Effects	Normal
Subject Variable	tree
Optimization Technique	Dual Quasi-Newton
Integration Method	First Order

Dimensions	
Observations Used	35
Observations Not Used	0
Total Observations	35
Subjects	5
Max Obs Per Subject	7
Parameters	7

Parameters							
b1	b2	b3	s2u1	c12	s2u2	s2e	NegLogLike
190	10	100	500	0	10	50	594.79679

Iteration History						
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope	
1	32	275.908547	318.8882	4.119515	-89.8649	
2	48	169.3a77104	106.5314	0.621827	-92.8272	
3	64	165.870599	3.506504	0.230479	-3.88406	

Iteration History						
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope	
4	80	165.380252	0.490347	0.212342	-0.07912	
5	104	155.508669	9.871583	0.153711	-0.71632	
6	113	155.356297	0.152372	0.149838	-0.15877	
7	121	155.228667	0.12763	0.165682	-0.07145	
8	129	155.153969	0.074698	0.175899	-0.10725	
9	145	154.954349	0.19962	0.171706	-0.14883	
10	169	151.565865	3.388484	0.840418	-0.23765	
11	177	148.03735	3.528515	0.093466	-18.5448	
12	194	147.03713	1.00022	0.197118	-0.47338	
13	203	146.562833	0.474297	0.100963	-0.46132	
14	220	146.483571	0.079261	0.118934	-0.24607	
15	229	146.431424	0.052147	0.047293	-0.12021	
16	238	146.41842	0.013005	0.046216	-0.00674	
17	254	146.297871	0.120548	0.051887	-0.00711	
18	262	146.10487	0.193001	0.050176	-0.21916	
19	278	145.686567	0.418303	0.088157	-0.20355	
20	302	143.543882	2.142685	0.420782	-0.57624	
21	311	143.252318	0.291564	0.077769	-1.59937	
22	319	142.796248	0.45607	0.04558	-0.36972	
23	328	142.662806	0.133442	0.054311	-0.14701	
24	344	142.211396	0.45141	0.158202	-0.26168	
25	360	138.472157	3.739239	1.465597	-1.23232	
26	401	133.041523	5.430634	1.899527	-12.1726	
27	418	132.489305	0.552218	2.332948	-10.1543	
28	426	132.129672	0.359634	0.710042	-8.43645	
29	434	132.020963	0.108709	0.723868	-0.62187	
30	442	131.904226	0.116737	0.145356	-0.18925	
31	451	131.901252	0.002973	0.048437	-0.04671	
32	467	131.89142	0.009832	0.05117	-0.02091	
33	476	131.887298	0.004122	0.097005	-0.00267	
34	492	131.873197	0.0141	0.047127	-0.00523	
35	508	131.80287	0.070328	0.180915	-0.01412	
36	517	131.791033	0.011836	0.106256	-0.02434	
37	526	131.788633	0.0024	0.035644	-0.00347	
38	535	131.788023	0.000611	0.01589	-0.00057	
39	544	131.78796	0.000063	0.005644	-0.00009	
40	560	131.786949	0.001011	0.031107	-0.00002	

Iteration History						
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope	
41	568	131.786322	0.000626	0.041738	-0.0011	
42	576	131.785633	0.000689	0.013311	-0.00081	
43	592	131.775703	0.009931	0.122392	-0.00062	
44	600	131.759626	0.016076	0.028473	-0.01621	
45	616	131.715536	0.04409	0.293412	-0.01679	
46	632	131.200534	0.515002	0.328214	-0.05599	
47	649	130.866543	0.333991	0.124898	-0.15784	
48	658	130.714572	0.151972	0.160974	-0.12525	
49	667	130.672379	0.042193	0.160503	-0.06624	
50	676	130.655731	0.016647	0.027031	-0.02488	
51	685	130.652234	0.003498	0.032544	-0.0055	
52	694	130.651308	0.000925	0.011282	-0.00127	
53	703	130.651081	0.000227	0.0067	-0.00018	
54	712	130.651024	0.000057	0.000537	-0.00008	
55	721	130.651023	6.276E-7	0.000231	-9.88E-7	

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	261.3
AIC (smaller is better)	275.3
AICC (smaller is better)	279.5
BIC (smaller is better)	272.6

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
b1	190.88	16.9244	3	11.28	0.0015	0.05	137.02	244.75	0.00003
b2	7.9748	0.9063	3	8.80	0.0031	0.05	5.0905	10.8591	-0.00012
b3	347.34	25.6214	3	13.56	0.0009	0.05	265.80	428.88	-7.81E-6
s2u1	1231.58	815.24	3	1.51	0.2280	0.05	-1362.88	3826.03	3.727E-7
c12	25.7615	27.2662	3	0.94	0.4145	0.05	-61.0117	112.53	-0.00001
s2u2	0.8121	1.3460	3	0.60	0.5888	0.05	-3.4716	5.0958	0.000231
s2e	55.4778	15.6913	3	3.54	0.0385	0.05	5.5412	105.41	-6.43E-6

Interpretation

The “Specifications” table echoes inputs provided to the procedure, the “Dimensions” table lists counts of data used, and the “Parameters” table lists the parameters to be estimated and their starting values. It is a good habit to check these tables for each run to make sure the specifications they contain match those you expect for the data set and model.

The “Iteration History” table provides a step-by-step output of the numerical optimization of the first-order method objective function. Fifty-five iterations are required to achieve convergence, due in part to starting values that were somewhat far from their optimal values. The GCONV convergence criterion was met, indicating small gradients relative to approximate second derivatives.

The “Fit Statistics” table is useful for comparing different model fits. As an exercise, you may want to try some other models and compare AICC values. (AICC includes a small-sample correction to the standard AIC statistic.) Interesting models to try would be simplifications of this one that dropped one or both of the random effects.

Finally, the “Parameter Estimates” table provides the first-order maximum likelihood estimates along with the usual associated statistics. Note that the parameters S2U1, S2U2, and S2E are all variance components, and so the t -statistics are potentially unreliable because of skewed sampling distributions. Nonetheless, the relatively small t -statistic for S2U2 provides some evidence that U2 (the random coefficient for B2) is not adding much to the model and could potentially be removed, along, of course, with C12.

15.3.2 %NLINMIX Macro Analysis—Method 1

We now illustrate how to fit a similar model using the %NLINMIX macro. Since you typically want to use the NLMIXED procedure to fit nonlinear mixed models, you may skip the remainder of this example unless you are trying to familiarize yourself with the %NLINMIX macro and how it compares to PROC NLMIXED.

There are three basic methods implemented here using the %NLINMIX macro, and these are described in detail in a later section. To implement %NLINMIX macro method 1, the mean function is approximated with a first-order multivariate Taylor series expansion about $u_{i1} = u_{i2} = 0$, producing the approximate model

$$Y_{ij} = \frac{\beta_1}{1 + \beta_2 e^{-x_{ij}/\beta_3}} + \frac{1}{1 + \beta_2 e^{-x_{ij}/\beta_3}} u_{i1} + \frac{-\beta_1 e^{-x_{ij}/\beta_3}}{(1 + \beta_2 e^{-x_{ij}/\beta_3})^2} u_{i2} + e_{ij}$$

Note that the approximation is exact in u_{i1} because it enters the model linearly. The %NLINMIX program to implement method 1 for this model follows.

Program for %NLINMIX Macro—Method 1

```
%nlinmix(data=tree,
  procopt=method=ml,
  model=%str(
    num   = b1+u1;
    e     = exp(-day/b3);
    den   = 1 + (b2+u2)*e;
    predv = num/den;
  ),
  parms=%str(b1=190 b2=10 b3=100),
  stmts=%str(
    class tree;
    model pseudo_circum = d_b1 d_b2 d_b3 / noint notest
      solution cl;
    random d_u1 d_u2 / subject=tree type=un solution cl;
  ),
  expand=zero
)
run;
```

A description of the syntax of the macro used here follows.

The DATA= argument specifies the SAS data set to use for the analysis, which in this case is TREE. The PROCOPT= argument specified that METHOD=ML is to be used instead of the default REML. This is done to facilitate comparison with the previous PROC NLMIXED results.

The MODEL= statement is an important part of the code as it defines the nonlinear mean function. DATA step code is used to specify the model. Because there are multiple statements ending with semicolons, the entire argument is enclosed with the %STR() macro. The variables B1, B2, and B3 represent the fixed effects parameters in this model specification, and they correspond to β_1 , β_2 , and β_3 in the logistic model. U1 and U2 are the random effects parameters corresponding to u_{i1} and u_{i2} . The auxiliary variables NUM, E, and DEN make the specification easier, and the final predicted value must be assigned to the PREDV variable.

The fixed effects parameters and their starting values are listed in the PARMS= argument. Our starting values for this problem are $\beta_1 = 190$, $\beta_2 = 10$, and $\beta_3 = 100$. As with most nonlinear optimization problems, good starting values are important to ensure convergence of the algorithm. They should be selected from prior knowledge or a preliminary nonlinear least-squares analysis.

The STMTS= argument specifies PROC MIXED statements to be executed for each iteration. They have a particular form. The response variable in the MODEL statement must be of the form “pseudo_y”, where y is the name of the response variable in the input data set. In this case y is “circum”. You then list each of the fixed effects parameter names, each name preceded by “d_”, to indicate differentiation as described in the previous sections. You must also specify the NOINT NOTEST SOLUTION and CL options in the MODEL statement. The RANDOM statement lists the random effects parameters, also preceded by “d_” to indicate derivatives. You must also specify the SUBJECT= argument. TYPE=UN indicates a 2×2 covariance matrix between D_U1 and D_U2, and the SOLUTION and CL options request EBLUPs.

The EXPAND=ZERO argument requests a method 1 analysis, which employs a first-order Taylor series expansion around $u_{i1} = 0$ and $u_{i2} = 0$. %NLINMIX therefore sets U1 and U2 equal to zero throughout the iterations for this analysis.

Results for NLINMIX Macro—Method 1

The macro produces results in both the SAS log and SAS output. Output 15.2 shows the SAS log.

Output 15.2 Log for NLINMIX Macro Method 1 and the Orange Tree Data

```

The NLINMIX Macro

Data Set : tree
Response : circum
Fixed-Effect Parameters : b1 b2 b3
Random-Effect Parameters : u1 u2
Expansion Point : zero

Calling PROC NLIN to initialize.
b1 = b1
b2 = b2
b3 = b3
Iteratively calling PROC MIXED.
PROC MIXED call 0

iteration = 0
convergence criterion = .
b1=192.68712406 b2=7.8566083412 b3=353.53159177 COVP1=1254.5457539
COVP2=25.320282922
COVP3=0.7682136432 COVP4=55.650967995
PROC MIXED call 1

iteration = 1
convergence criterion = 0.0000126663
b1=192.68756706 b2=7.8565657925 b3=353.53356035 COVP1=1254.5531506
COVP2=25.320185239
COVP3=0.768200977 COVP4=55.651060675
PROC MIXED call 2

iteration = 2
convergence criterion = 9.2178887E-8
b1=192.68759078 b2=7.8565638392 b3=353.5336601 COVP1=1254.5532663
COVP2=25.320187573
COVP3=0.7682010478 COVP4=55.651065805
PROC MIXED call 3

iteration = 3
convergence criterion = 4.8900634E-9
b1=192.68759201 b2=7.8565637377 b3=353.53366526 COVP1=1254.5532724
COVP2=25.320187697
COVP3=0.7682010515 COVP4=55.651066077
NLINMIX convergence criteria met.

NOTE: Numeric values have been converted to character values at the
      places given by:
      (Line):(Column).
      13:71
NOTE: There were 35 observations read from the data set
      WORK._NLINMIX.
NOTE: The data set WORK._NLINMIX has 35 observations and 24
      variables.

```

```

NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
PROC MIXED call 4

NOTE: Convergence criteria met.
NOTE: The data set WORK._SOLNR has 10 observations and 10
      variables.
NOTE: The data set WORK._COV has 4 observations and 3 variables.
NOTE: The data set WORK._SOLN has 3 observations and 9 variables.
NOTE: The data set WORK._FIT has 4 observations and 2 variables.
NOTE: PROCEDURE MIXED used (Total process time):
      real time           0.31 seconds
      cpu time            0.11 seconds

```

Interpretation for NLINMIX Macro—Method 1

The macro first prints some basic information about your model specification as well as some relevant dimensions. It then calls PROC NLIN one time to obtain initial estimates of the fixed effects parameters. %NLINMIX uses these, along with 0's for the random effects parameters, to create the first pseudo-data set according to the details of the previous section.

The iterative phase begins after the initial call to PROC MIXED, and three iterations are required in order to drop below the default convergence criterion of 1E-8.

Several data sets are available for your use after the macro terminates. WORK._NLINMIX, the primary working data set, has all of the original data plus parameter estimates, derivatives, and the pseudo-data values in a variable named PSEUDO_CIRCUM. The data set WORK._SOLN contains the fixed effects parameter estimates along with their estimated standard errors and approximate confidence limits. WORK._COV and WORK._SOLNR contain the same statistics for the covariance parameters and the EBLUPs, respectively. Finally, WORK._FIT has the basic model-fitting information, which can be useful in comparing different models.

%NLINMIX also prints the results from the final PROC MIXED call to SAS output. The results are shown in Output 15.3.

Output 15.3 Results for %NLINMIX Macro—Method 1 on the Orange Tree Data

Model Information	
Data Set	WORK._NLINMIX
Dependent Variable	pseudo_circum
Covariance Structure	Unstructured
Subject Effect	tree
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
tree	5	1 2 3 4 5

Dimensions	
Covariance Parameters	4
Columns in X	3
Columns in Z Per Subject	2
Subjects	5
Max Obs Per Subject	7

Number of Observations	
Number of Observations Read	35
Number of Observations Used	35
Number of Observations Not Used	0

Parameter Search						
CovP1	CovP2	CovP3	CovP4	Variance	Log Like	-2 Log Like
1254.55	25.3202	0.7682	55.6511	55.6511	-130.6807	261.3615

Iteration History			
Iteration	Evaluations	-2 Log Like	Criterion
1	1	261.36147124	0.00000000

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	tree	1254.55
UN(2,1)	tree	25.3202
UN(2,2)	tree	0.7682
Residual		55.6511

Fit Statistics	
-2 Log Likelihood	261.4
AIC (smaller is better)	275.4
AICC (smaller is better)	279.5
BIC (smaller is better)	272.6

PARMS Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
3	0.00	1.0000

Solution for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
d_b1	192.69	17.1074	24	11.26	<.0001	0.05	157.38	228.00
d_b2	7.8566	0.8695	24	9.04	<.0001	0.05	6.0620	9.6511
d_b3	353.53	26.0043	24	13.60	<.0001	0.05	299.86	407.20

Solution for Random Effects									
Effect	tree	Estimate	Std Err Pred	DF	t Value	Pr > t	Alpha	Lower	Upper
d_u1	1	-34.9228	16.5344	24	-2.11	0.0453	0.05	-69.0481	-0.7975
d_u2	1	-0.9270	0.5990	24	-1.55	0.1348	0.05	-2.1633	0.3092
d_u1	2	34.1932	16.5344	24	2.07	0.0496	0.05	0.06791	68.3186
d_u2	2	0.4798	0.5990	24	0.80	0.4310	0.05	-0.7565	1.7160
d_u1	3	-41.4448	16.5344	24	-2.51	0.0194	0.05	-75.5702	-7.3195
d_u2	3	-0.7759	0.5990	24	-1.30	0.2075	0.05	-2.0121	0.4603
d_u1	4	44.9252	16.5344	24	2.72	0.0120	0.05	10.7999	79.0505
d_u2	4	0.8558	0.5990	24	1.43	0.1660	0.05	-0.3804	2.0920
d_u1	5	-2.7508	16.5344	24	-0.17	0.8693	0.05	-36.8761	31.3745
d_u2	5	0.3673	0.5990	24	0.61	0.5455	0.05	-0.8689	1.6036

Interpretation for %NLINMIX Macro—Method 1 (continued)

The “Class Level Information” table in Output 15.3 provides the levels of the TREE variable, as specified in the CLASS statement. The “Parameter Search” table is printed because the covariance parameter values from the previous iteration are passed to PROC MIXED as starting values. The “Iteration History” table reveals that only one step is required to attain convergence from this initial value, as might be expected after performing previous iterations.

The data from the next several tables, beginning with the “Covariance Parameter Estimates” table, are available as output data sets, as mentioned at the end of the SAS log listing above. The estimates for the parameters in \mathbf{G} equal (1254, 25.3, 0.77) and that for the residual variance is 55.7. As mentioned by Lindstrom and Bates (1990), it appears that nearly all of the intra-subject variability occurs in the u_{11} parameters. This is also apparent in the magnitudes of the EBLUPs displayed in the final table of output.

The estimates for β_1 , β_2 , and β_3 are 192.69, 7.85, and 353.53, respectively, as shown in the “Solution for Fixed Effects” table. These are the estimated coefficients for the derivative variables D_B1, D_B2, and D_B3 computed by the macro using finite differences.

15.3.3 %NLINMIX Macro Analysis—Method 2

We now fit the model by using a linearization around the EBLUPs (method 2), a method described by Lindstrom and Bates (1990). You can use the following program to invoke method 2. The only difference from the method 1 program is the EXPAND=EBLUP macro option, rather than EXPAND=ZERO.

Program for %NLINMIX Macro—Method 2

```
%nlinmix(data=tree,
  procopt=method=ml,
  model=%str(
    num   = b1+u1;
    e     = exp(-day/b3);
    den   = 1 + (b2+u2)*e;
    predv = num/den;
  ),
  parms=%str(b1=190 b2=10 b3=100),
  stmts=%str(
    class tree;
    model pseudo_circum = d_b1 d_b2 d_b3 /
      noint notest solution cl;
    random d_u1 d_u2 / subject=tree type=un solution cl;
  ),
  expand=eblup
)
run;
```

The results are not shown. The macro converges in 11 iterations and produces fixed effect parameters estimates 190.29, 7.95, and 343.4., which are between those from PROC NLMIXED and NLINMIX macro—method 1 (see Table 15.6 below).

15.3.4 %NLINMIX Macro Analysis—Method 3, With and Without Variance Weighting

As an example of a method 3 analysis, consider a matrix $\text{Var}[\mathbf{e}] = \mathbf{R}(\boldsymbol{\theta})$ with compound symmetry structure, roughly analogous to the role played by u_1 in the previous models. The following code fits this model.

Program for %NLINMIX Macro—Method 3

```
%nlinmix(data=tree,
  procopt=method=ml,
  model=%str(
    num    = b1;
    e      = exp(-day/b3);
    den    = 1 + b2*e;
    predv = num/den;
  ),
  parms=%str(b1=190 b2=10 b3=100),
  stmts=%str(
    class tree;
    model pseudo_circum = d_b1 d_b2 d_b3 /
      noint notest solution cl;
    repeated / subject=tree type=cs;
  )
)
run;
```

Note that U1 and U2 have been removed as random effects, and that a REPEATED statement is used instead of a RANDOM statement. Results are not shown, but are similar to those for methods 1 and 2 (see Table 15.6 below).

Program for NLINMIX Macro—Method 3 with Variance Weighting

To further expand on this example, suppose you are concerned that the variability of the residual errors is heterogeneous, and more specifically, varies as a function of the mean value. One way to model this is to introduce a diagonal weight matrix $\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})$; its elements are a power of the mean function. The assumed variance-covariance matrix is then

$$\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})^{-1/2} \mathbf{R}_i(\mathbf{0}) \mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})^{-1/2}.$$

Let's first try a power equal to 2, which is a constant coefficient-of-variation model. The %NLINMIX program is as follows:

```
%nlinmix(data=tree,
  procopt=method=ml,
  model=%str(
    num    = b1;
    e      = exp(-day/b3);
    den    = 1 + b2*e;
    predv = num/den;
  ),
  weight = %str(
    wghtv = 1/predv**2;
  ),
  parms=%str(b1=190 b2=10 b3=100),
  stmts=%str(
    class tree;
    model pseudo_circum = d_b1 d_b2 d_b3 /
      noint notest solution cl;
    repeated / subject=tree type=cs;
    weight wghtv;
  )
)
run;
```

Note that METHOD=ML is still used here instead of the default METHOD=REML to facilitate a later comparison between different power models. This program is the same as the previous

method 3 program, except for the addition of the WEIGHT= argument and the WEIGHT statement within the STMTS= block. These two set up $\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})$ to contain values equal to the reciprocal of the square of the mean.

Note: You can also fit this kind of model using previous methods, but for random effects models, you must decide if the weights are to depend upon the random effects or not. While making them depend on the random effects may be a little easier to code, the model can become less stable because of the presence of the random effects in the variance function. You can remove dependence of the weights on the random effects by creating a variable that equals the mean value without any random effects, as shown for the method 3 code above.

Results for %NLINMIX Macro—Method 3 with Variance Weighting

The SAS log from this program is shown in Output 15.4.

Output 15.4 SAS Log for %NLINMIX Macro—Method 3 with Variance Weighting

The NLINMIX Macro

Data Set	:	tree
Response	:	circum
Fixed-Effect Parameters	:	b1 b2 b3

```

b1 = b1
b2 = b2
b3 = b3
Iteratively calling PROC MIXED.
PROC MIXED call 0

iteration = 0
convergence criterion = .
b1=192.68712406 b2=7.8566083412 b3=353.53159177
PROC MIXED call 1

iteration = 1
convergence criterion = 0.0049023284
b1=198.65009255 b2=7.5995029327 b3=374.86619483
PROC MIXED call 2

iteration = 2
convergence criterion = 0.000016595
b1=198.83107259 b2=7.6369417294 b3=374.73757435
PROC MIXED call 3

iteration = 3
convergence criterion = 6.9334977E-7
b1=198.83338423 b2=7.6370684665 b3=374.74256143
PROC MIXED call 4

iteration = 4
convergence criterion = 3.6129124E-8
b1=198.83328855 b2=7.6370683967 b3=374.7423016
PROC MIXED call 5

iteration = 5
convergence criterion = 1.8824486E-9
b1=198.83329354 b2=7.6370684005 b3=374.74231514
NLINMIX convergence criteria met.

```

NOTE: Numeric values have been converted to character values at the places given by:
 (Line):(Column).
 11:71

NOTE: There were 35 observations read from the data set WORK._NLINMIX.

NOTE: The data set WORK._NLINMIX has 35 observations and 20 variables.

NOTE: DATA statement used (Total process time):
 real time 0.01 seconds
 cpu time 0.02 seconds

PROC MIXED call 6

NOTE: Convergence criteria met.

NOTE: The data set WORK._SOLN has 3 observations and 9 variables.

NOTE: The data set WORK._FIT has 4 observations and 2 variables.

NOTE: PROCEDURE MIXED used (Total process time):
 real time 0.11 seconds
 cpu time 0.09 seconds

The macro iterates to convergence in five steps. The final PROC MIXED results from the macro are shown in Output 15.5.

Output 15.5 Results for NLINMIX Macro—Method 3 with Variance Weighting

Model Information	
Data Set	WORK._NLINMIX
Dependent Variable	pseudo_circum
Weight Variable	wghtv
Covariance Structure	Compound Symmetry
Subject Effect	tree
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information		
Class	Levels	Values
tree	5	1 2 3 4 5

Dimensions		
Covariance Parameters		2
Columns in X		3
Columns in Z		0

Dimensions	
Subjects	5
Max Obs Per Subject	7

Number of Observations	
Number of Observations Read	35
Number of Observations Used	35
Number of Observations Not Used	0

Iteration History				
Iteration	Evaluations	-2 Log Like	Criterion	
0	1	295.22988690		
1	1	268.94511837	0.00000000	

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
CS	tree	0.01838
Residual		0.008545

Fit Statistics		
-2 Log Likelihood	268.9	
AIC (smaller is better)	278.9	
AICC (smaller is better)	281.0	
BIC (smaller is better)	277.0	

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	26.28	<.0001

Solution for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
d_b1	198.83	15.9038	27	12.50	<.0001	0.05	166.20	231.47
d_b2	7.6371	0.4546	27	16.80	<.0001	0.05	6.7044	8.5697
d_b3	374.74	23.2821	27	16.10	<.0001	0.05	326.97	422.51

Interpretation for NLINMIX Macro—Method 3 with Variance Weighting

The estimate of the common correlation assumed in $\mathbf{R}(\boldsymbol{\theta})$ is obtained from values in the “Covariance Parameter Estimates” table: $0.01838/(0.01838+0.008545) = 0.68$. The fixed effects parameter estimates are 198.83, 7.637, and 374.74, respectively. Note that these are a bit different compared to those in previous methods, illustrating the effect of the constant-coefficient-of-variance weighting.

15.3.5 Choosing the Exponent for the Weight Matrix That Leads to the Best Fit

An interesting exercise is to fit models with different powers to see which one minimizes the approximate $-2 \log \text{Likelihood}$. For this model with a power of 2, its value is 268.9, as shown in the preceding “Fit Statistics” table. Values for other models are listed in the following data set:

```
data pom;
  input power m211;
  datalines;
-4      275.4207
-3.5    269.8219
-3      266.9001
-2.75   266.4461
-2.5    266.6548
-2.25   267.5017
-2      268.9451
-1      279.5383
  0      295.2636
run;
```

The M2LL values were obtained by running the macro again for each power. You can fit a quadratic curve through these points with the following program:

```
proc glm data=pom;
  model m211 = power power*power;
run;
```

This program produces an approximate maximum likelihood power estimate of $-22.0336041 / 4.1773302 / 2 = -2.637$.

15.3.6 Comparing the Logistic Model Fitting Methods

Table 15.1 summarizes the previous five analyses by presenting estimates for β_1 , β_2 , and β_3 and their approximate standard errors. Note that the point estimates are similar from the first four approaches, whereas the variance weighting performed with %NLINMIX method 3 does produce changes. As for standard errors, those from %NLINMIX method 3 without weighting are a fair bit larger than those from the random effects models. This is akin to subject-specific versus population-averaged results obtained with the %GLIMMIX macro. The variance weighting for method 3 reduces the standard errors substantially.

Table 15.1 Comparison of Estimates of β_1 , β_2 , and β_3 and Their Estimated Standard Errors from Different Logistic Model Fits to the Orange Tree Data

Method	$\hat{\beta}_1 \pm SE$	$\hat{\beta}_2 \pm SE$	$\hat{\beta}_3 \pm SE$
PROC NL MIXED (first-order)	190.9 ± 16.9	7.97 ± 0.91	347.3 ± 25.6
NLINMIX 1	192.7 ± 17.1	7.86 ± 0.87	353.5 ± 26.0
NLINMIX 2	190.3 ± 16.5	7.95 ± 0.90	343.4 ± 24.2
NLINMIX 3	192.3 ± 16.3	7.92 ± 2.26	352.4 ± 55.7
NLINMIX 3 (with weight)	198.8 ± 15.9	7.64 ± 0.45	374.7 ± 23.3

15.4 Example: Nested Nonlinear Random Effects Models

One critical restriction of PROC NL MIXED is that it allows only one RANDOM statement with one SUBJECT= variable. Users often misinterpret this to mean that only one level of experimental units can be fit. This is not strictly the case, as it is possible to code for additional levels by constructing the variance-covariance matrix appropriately.

Program

Consider the following code to simulate data from a nested model:

```
%let na = 100;
%let nb = 5;
%let nr = 2;
data nested;
  do A = 1 to &na;
    err1 = 3*rannor(339205);
    do B = 1 to &nb;
      err2 = 2*rannor(0);
      do rep = 1 to &nr;
        err3 = 1*rannor(0);
        resp = 10 + err1 + err2 + err3;
        output;
      end;
    end;
  end;
run;

/*---PROC NL MIXED requires that the data are grouped---*/
/*---by subject. These data are already properly ---*/
/*---grouped. The PROC SORT step is simply added here ---*/
/*---for completeness. ---*/
proc sort data=nested;
  by A;
run;
```

```

proc nlmixed data=nested method=firo;
  array aeffect { 1 };
  array beffect {&nb };
  vara      = exp(2*logsigma);
  varb_a    = exp(2*logsigb_a);
  mean      = intercept + aeffect{1} + beffect{B};
  model resp ~ normal (mean,s2);
  random aeffect1 beffect1 beffect2 beffect3 beffect4 beffect5 ~
    normal( [0,0,0,0,0,0],
            [vara,
             0, varb_a,
             0, 0, varb_a,
             0, 0, 0, varb_a,
             0, 0, 0, 0, varb_a,
             0, 0, 0, 0, 0, varb_a] ) subject=A;
  estimate 'Var[A]'     vara;
  estimate 'Var[B(A)]' varb_a;
run;

/*---equivalent proc mixed code---*/
proc mixed data=nested method=ml;
  class a b;
  model resp = / s;
  random a b(a);
run;

```

Note in the PROC NLMIXED code how the AEFFECT and BEFFECT arrays are used to construct the appropriate value for the MEAN variable. You can use the ARRAY statement in PROC NLMIXED similarly to the way you use it in SAS DATA step programming. The key part of the code is the construction of the nested covariance structure using one RANDOM statement. Of course, for this linear model, the PROC MIXED code is much more parsimonious, but this illustrates how you could go about fitting certain nonlinear models with nested random effects. Selected results from this code are shown in Output 15.6. Note that the estimates from PROC NLMIXED and PROC MIXED (with METHOD=ML) are identical.

Results

Output 15.6 Results for Nested Model

The NLMIXED Procedure

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
logsigma	1.1082	0.07723	94	14.35	<.0001	0.05	0.9549	1.2615	0.000043
logsigb_a	0.6540	0.04019	94	16.27	<.0001	0.05	0.5742	0.7338	-0.00017
intercept	10.1587	0.3164	94	32.11	<.0001	0.05	9.5304	10.7869	-0.00003
s2	0.9673	0.06118	94	15.81	<.0001	0.05	0.8458	1.0888	0.000024

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Var[A]	9.1740	1.4169	94	6.47	<.0001	0.05	6.3607	11.9874
Var[B(A)]	3.6988	0.2973	94	12.44	<.0001	0.05	3.1084	4.2891

The Mixed Procedure	
Covariance Parameter Estimates	
Cov Parm	Estimate
A	9.1740
B(A)	3.6988
Residual	0.9673

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	10.1587	0.3164	99	32.11	<.0001

15.5 Example: Zero-Inflated Poisson and Hurdle Poisson Models

Count data encountered in practice often have an excess of zeros beyond that expected from the standard Poisson distribution. Zero-inflation and hurdle models are two ways to effectively handle this kind of data; refer to Lambert (1992), Cameron and Trivedi (1998), and Bohning et al. (1999). Consider the following simulated data, which are plotted in Figure 15.2.

Program

```

/* ----- */
/* A zero-inflated Poisson (ZIP) model */
/* Z ~ Poisson(lambda) */
/* Pr(Count=0) = pi + (1-pi)*Pr(Z=0) */
/* Pr(Count=j) = (1-pi)*Pr(Z=j), j > 0 */
/* lambda is a function of covariates */
/* and a random effect */
/* ----- */

%let pi = 0.27;
data zip;
  do s = 1 to 100;
    u = rannor(556712);
    do i = 1 to 20;
      x = int(ranuni(0)*100);
      y = int(rannor(0)*100);
      if (ranuni(0) < &pi) then do;
        count = 0;
      end;
      else do;
        count = x;
      end;
    end;
  end;
run;

```

```

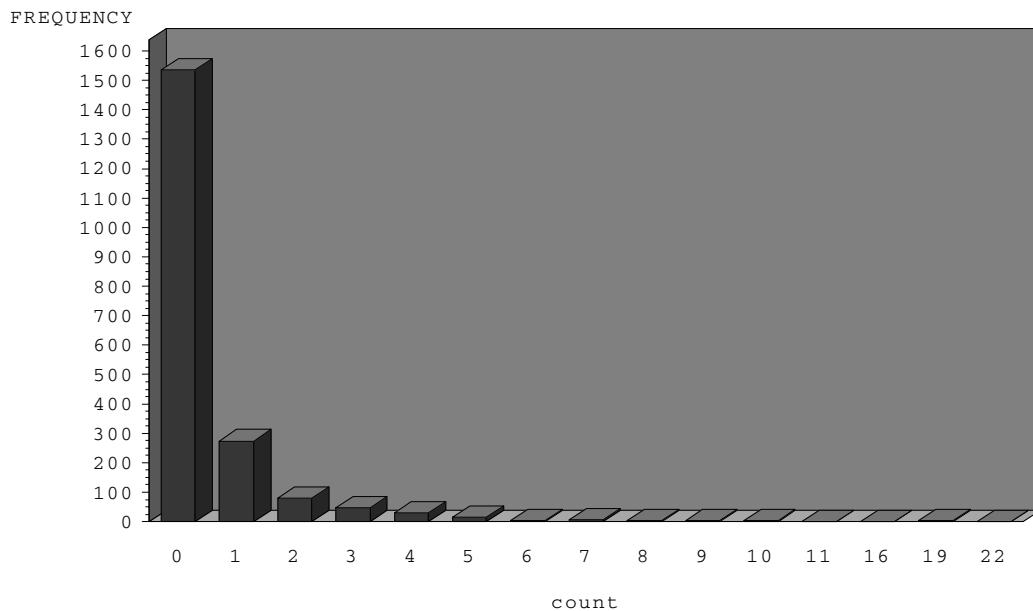
        lambda = .;
end; else do;
    lambda = exp(-2 + 0.01*x + 0.01*y + u);
    count = ranpoi(0,lambda);
end;
output;
end;
drop i u lambda;
run;

proc gchart data=zip;
    vbar3d count / discrete;
run;

```

Results

Figure 15.2 Zero-Inflated Poisson Data with a Random Effect



Interpretation

These data equal zero with probability $\pi = 1/(1+e) = 0.27$, and are otherwise Poisson distributed with mean parameter λ modeled in the standard way as a function of a linear predictor in covariates X and Y plus a simple random effect that induces correlation within observations sharing the same value of variable S. We fit three models to the data:

1. Poisson model with a random effect
2. negative binomial model with a random effect
3. zero-inflated Poisson model with a random effect (the true model)

The first two models do not directly account for the excess zeros. They are generalized linear mixed models that you can alternatively fit with PROC GLIMMIX. Note for the negative binomial, the maximum count of 22 is hard-coded. The code below fits the models, and selected tables from the fits are shown in Outputs 15.7, 15.8, and 15.9.

Program

```

/* ----- */
/* Poisson model with a random effect */
/* ----- */
proc nlmixed data=zip;
parameters b0=0 b1=0 b2=0;
lambda = exp(b0 + b1*x + b2*y + u);
model count ~ poisson(lambda);
random u ~ normal(0,s2u) subject=s;
run;

/* ----- */
/* Negative binomial model with a random effect */
/* ----- */
proc nlmixed data=zip;
parameters b0=0 b1=0 b2=0;
eta = b0 + b1*x + b2*y + u;
p = exp(eta)/(1+exp(eta));
model count ~ negbin(22,p);
random u ~ normal(0,s2u) subject=s;
run;

/*-----*/
/* ZIP model with a random effect */
/*-----*/
proc nlmixed data=zip;
parameters b0=0 b1=0 b2=0
a0=0 s2u=1;
/* linear predictor for the inflation probability */
linpinfl = a0;
/* infprob = inflation probability for zeros */
/* = logistic transform of the linear predictor */
infprob = 1/(1+exp(-linpinfl));
/* Poisson mean */
lambda = exp(b0 + b1*x + b2*y + u);
/* Build the ZIP log likelihood */
if count=0 then
l1 = log(infprob + (1-infprob)*exp(-lambda));
else l1 = log((1-infprob)) + count*log(lambda) -
lgamma(count+1) - lambda;
model count ~ general(l1);
random u ~ normal(0,s2u) subject=s;
estimate "inflation probability" infprob;
run;

```

Results

Output 15.7 Results of Poisson Mixed Model on Zero-Inflated Data

Fit Statistics	
-2 Log Likelihood	2958.6
AIC (smaller is better)	2966.6
AICC (smaller is better)	2966.6
BIC (smaller is better)	2977.0

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
b0	-2.3920	0.1438	99	-16.64	<.0001	0.05	-2.6772	-2.1067	-0.00117
b1	0.01174	0.001197	99	9.80	<.0001	0.05	0.009360	0.01411	-0.09366
b2	0.009925	0.000360	99	27.56	<.0001	0.05	0.009211	0.01064	-0.30186
s2u	1.1393	0.2132	99	5.34	<.0001	0.05	0.7163	1.5623	-0.00047

Output 15.8 Results of Negative Binomial Mixed Model on Zero-Inflated Data

Fit Statistics	
-2 Log Likelihood	2932.4
AIC (smaller is better)	2940.4
AICC (smaller is better)	2940.4
BIC (smaller is better)	2950.8

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
b0	5.4653	0.1446	99	37.80	<.0001	0.05	5.1784	5.7522	0.000255
b1	-0.01165	0.001254	99	-9.29	<.0001	0.05	-0.01414	-0.00916	0.01476
b2	-0.00982	0.000376	99	-26.09	<.0001	0.05	-0.01056	-0.00907	0.027356
s2u	1.1117	0.2081	99	5.34	<.0001	0.05	0.6989	1.5245	0.000173

Output 15.9 Results of ZIP Mixed Model on Zero-Inflated Data

Fit Statistics	
-2 Log Likelihood	2803.6
AIC (smaller is better)	2813.6
AICC (smaller is better)	2813.7
BIC (smaller is better)	2826.7

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
b0	-1.9979	0.1530	99	-13.06	<.0001	0.05	-2.3014	-1.6944	-0.00224
b1	0.01011	0.001299	99	7.78	<.0001	0.05	0.007535	0.01269	-0.15649
b2	0.01016	0.000394	99	25.78	<.0001	0.05	0.009378	0.01094	-0.0434

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a0	-1.0934	0.1594	99	-6.86	<.0001	0.05	-1.4097	-0.7771	-0.00034
s2u	1.0828	0.2095	99	5.17	<.0001	0.05	0.6671	1.4985	-0.00145

Additional Estimates									
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	
inflation probability	0.2510	0.02997	99	8.38	<.0001	0.05	0.1915	0.3104	

Interpretation

Based on Outputs 15.7, 15.8, and 15.9, all three models do a good job of estimating the coefficients of X and Y and have a small upward bias in their estimates of the variance of the random effect S2U; the true values for these parameters are 0.01, 0.01, and 1, respectively. As expected for the Poisson model, the Intercept estimate is biased downward from the true value of -2.

Interestingly, the negative binomial mixed model fits quite a bit better than the Poisson mixed model according to all fit criteria (likelihood ratio test, AIC, AICC, and BIC). The true ZIP model improves on the negative binomial fit to an even larger extent and is clearly the best-fitting model of the three. For the ZIP model, all parameter estimates are close to their true values, including the estimate of the inflation probability (estimated value 0.25 ± 0.03 , true value = 0.27).

An alternative to the ZIP model is the hurdle model, which uses two different Poisson variables to handle the excess zeros. Consider the following simulated data and PROC NLMIXED code, which produce the results in Output 15.10.

Program

```

/* -----
/* A Poisson hurdle mixed model */
/*
/* Z0 ~ Poisson(lambda0) */
/* Z1 ~ Poisson(lambda1) */
/*
/* Pr(Count=0) = Pr(Z0=0) */
/*
/* Pr(Count=j) = -----
/*           1 - Pr(Z1=0)
/*           for j > 0
/* -----
*/
data hurdle;
do s = 1 to 100;
  u = rannor(9213740);
  do i = 1 to 20;
    x = int(ranuni(0)*100);
    y = int(rannor(0)*100);
    count = ranpoi(0, exp(-2 + 0.01*x + 0.01*y + u) );
  end;
end;

```

```

if (count > 0) then do;
  count1 = 0;
  do while (count1 = 0);
    count1 = ranpoi(0,exp(-1 + 0.01*x + 0.01*y + u));
  end;
  count = count1;
end;
output;
end;
drop i u;
run;

proc nlmixed data=hurdle;
parameters a0=-1 a1=0 a2=0
            b0=-1 b1=0 b2=0
            s2u=1;
linpzero = a0 + a1*x + a2*y + u;
linpnozero = b0 + b1*x + b2*y + u;
muzero = exp(linpzero);
munozero = exp(linpnozero);
f10 = exp(-muzero);
f20 = exp(-munozero);
logpnozero = log(1-f10) - log(1-f20) - munozero -
             lgamma(count+1) + count*log(munozero);
if count = 0 then l1 = -muzero;
else l1 = logpnozero;
model count ~ general(l1);
random u ~ normal(0,s2u) subject=s;
run;

```

Results

Output 15.10 Results for Hurdle Poisson Mixed Model

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
a0	-1.8229	0.1403	99	-12.99	<.0001	0.05	-2.1013	-1.5444	-0.00023
a1	0.008264	0.001565	99	5.28	<.0001	0.05	0.005159	0.01137	-0.00824
a2	0.01023	0.000555	99	18.44	<.0001	0.05	0.009130	0.01133	-0.02238
b0	-0.9445	0.1247	99	-7.58	<.0001	0.05	-1.1918	-0.6971	0.000244
b1	0.01021	0.000779	99	13.10	<.0001	0.05	0.008660	0.01175	0.018054
b2	0.009689	0.000268	99	36.22	<.0001	0.05	0.009158	0.01022	0.018143
s2u	0.9453	0.1696	99	5.57	<.0001	0.05	0.6088	1.2817	0.000021

Interpretation

The parameter estimates are all near their true, simulated values.

15.6 Example: Joint Survival and Longitudinal Model

For this example we consider data containing both time-to-event and longitudinal responses. We use the CD4 data from Carlin and Louis (1996) and Guo and Carlin (2003); data sets are available at <http://www.biostat.umn.edu/~brad/data.html>.

Program

You can input the data to SAS using the following code. Output 15.11 shows the first 20 observations.

```

data alldata;
  input t2death death randgrp1 gender1 prevoil stratum1 y1-y5;
  array yy{5} y1-y5;
  array t{5} (0,2,6,12,18);
  patient = _n_;
  do i = 1 to 5;
    obstime = t{i};
    if (yy{i} ne .) then do;
      cd4 = sqrt(yy{i});
      output;
    end;
  end;
  drop i y1-y5 t1-t5;
  datalines;
...datalines...
;

proc sort data=alldata;
  by patient obstime;
data alldata;
  set alldata;
  by patient;
  last = last.patient;
run;

title1 'The first 20 observations of the joint data';
proc print data=alldata(obs=20);
run;

```

Results

Output 15.11 PROC PRINT Output for Joint Survival and Longitudinal Data

The first 20 observations of the joint data

Obs	t2death	death	randgrp1	gender1	prevoil	stratum1	patient	obstime	cd4	last
1	16.97	0	0	1	1	-1	1	0	10.6771	0
2	16.97	0	0	1	1	-1	1	6	8.4261	0
3	16.97	0	0	1	1	-1	1	12	9.4340	1
4	19.00	0	1	1	-1	-1	2	0	6.3246	0
5	19.00	0	1	1	-1	-1	2	6	8.1240	0
6	19.00	0	1	1	-1	-1	2	12	4.5826	0
7	19.00	0	1	1	-1	-1	2	18	5.0000	1
8	18.53	1	1	-1	1	-1	3	0	3.4641	0

Obs	t2death	death	randgrp1	gender1	prevoi1	stratum1	patient	obstime	cd4	last
9	18.53	1	1	-1	1	-1	3	2	3.6056	0
10	18.53	1	1	-1	1	-1	3	6	6.1644	1
11	12.70	0	0	1	1	1	4	0	3.8730	0
12	12.70	0	0	1	1	1	4	2	4.5826	0
13	12.70	0	0	1	1	1	4	6	2.6458	0
14	12.70	0	0	1	1	1	4	12	1.7321	1
15	15.13	0	1	1	1	1	5	0	7.2801	0
16	15.13	0	1	1	1	1	5	2	8.6023	0
17	15.13	0	1	1	1	1	5	6	8.6023	0
18	15.13	0	1	1	1	1	5	12	6.7082	1
19	1.90	1	0	-1	1	1	6	0	4.5826	1
20	14.33	0	0	1	1	-1	7	0	6.7823	0

Interpretation

T2DEATH is time to death and DEATH indicates a death event, with a value of 0 indicating that T2DEATH has been right-censored. RANDGRP1 is the treatment group, GENDER1 indexes sex, and PREVOI1 and STRATUM1 are covariates. The PATIENT variable identifies the subject, and CD4 contains the patients' longitudinal measurement of CD4 counts, measured at values of OBSTIME. LAST indicates the final measurement for each patient.

Program

Before detailing a joint model for CD4 and T2DEATH, we consider how you would go about modeling each variable separately. For CD4, we fit a random coefficients model and use TYPE=FA0(2) to implement a Cholesky parameterization of the **G** matrix in order to ensure nonnegative definiteness of its estimate. Note METHOD=ML is used for comparability with PROC NLMIXED below. This code produces the results in Output 15.12.

```

title1 'Longitudinal Model - MIXED';
proc mixed data=alldata noclprint covtest method=ml;
  class patient;
  model cd4 = obstime obstime*randgrp1 gender1
            prevoi1 stratum1/ s;
  random intercept obstime / subject=patient type=fa0(2);
  ods output covparms=cp;
  ods output solutionf=solf;
run;

```

Results

Output 15.12 Results of PROC MIXED Fit to CD4 Longitudinal Data

Longitudinal Model - MIXED

The Mixed Procedure

Model Information	
Data Set	WORK.ALADATA
Dependent Variable	cd4
Covariance Structure	Factor Analytic
Subject Effect	patient
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions	
Covariance Parameters	4
Columns in X	6
Columns in Z Per Subject	2
Subjects	467
Max Obs Per Subject	5

Number of Observations	
Number of Observations Read	1408
Number of Observations Used	1408
Number of Observations Not Used	0

Iteration History			
Iteration	Evaluations	-2 Log Like	Criterion
0	1	8187.90196231	
1	4	7024.64499493	0.01110168
2	2	7013.30285093	0.00283397
3	1	7006.44514698	0.00007588
4	1	7006.27404444	0.00000008
5	1	7006.27387441	0.00000000

Convergence criteria met.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
FA(1,1)	patient	3.9889	0.1467	27.19	<.0001
FA(2,1)	patient	-0.03259	0.01507	-2.16	0.0306
FA(2,2)	patient	0.1658	0.01703	9.74	<.0001
Residual		3.0715	0.1713	17.93	<.0001

Fit Statistics	
-2 Log Likelihood	7006.3
AIC (smaller is better)	7026.3
AICC (smaller is better)	7026.4
BIC (smaller is better)	7067.7

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
3	1181.63	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	8.0129	0.3511	463	22.82	<.0001
obstime	-0.1668	0.02038	404	-8.19	<.0001
obstime*randgrp1	0.02998	0.02891	535	1.04	0.3003
gender1	-0.1582	0.3249	535	-0.49	0.6265
prevoi1	-2.3152	0.2382	535	-9.72	<.0001
stratum1	-0.1309	0.2352	535	-0.56	0.5780

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
obstime	1	404	67.00	<.0001
obstime*randgrp1	1	535	1.08	0.3003
gender1	1	535	0.24	0.6265
prevoi1	1	535	94.50	<.0001
stratum1	1	535	0.31	0.5780

Interpretation

Note that OBSTIME1 and PREVOI1 are strongly significant; the other effects are not.

Program

To fit the survival outcome T2DEATH, we use PROC LIFEREG as follows, producing the results in Output 15.13.

```
title1 'Survival Model - LIFEREG';
proc lifereg data=alldata(where=(last=1));
  model t2death*death(0)=randgrp1 gender1 prevoi1 stratum1
    / dist=exponential;
  ods output ParameterEstimates=survpe(
    rename=(Parameter=SurvEffect));
run;
```

Results

Output 15.13 Results of PROC LIFEREG Fit to T2DEATH Survival Data

Survival Model - LIFEREG

The LIFEREG Procedure

Model Information	
Data Set	WORK.ALADATA
Dependent Variable	Log(t2death)
Censoring Variable	death
Censoring Value(s)	0
Number of Observations	467
Noncensored Values	188
Right Censored Values	279
Left Censored Values	0
Interval Censored Values	0
Name of Distribution	Exponential
Log Likelihood	-442.5331124

Number of Observations Read	467
Number of Observations Used	467

Algorithm converged.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
randgrp1	1	2.0197	0.1553
gender1	1	1.9100	0.1670
prevoi1	1	29.9625	<.0001
stratum1	1	1.0241	0.3115

Analysis of Parameter Estimates						
Parameter	DF	Estimate	Standard Error	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	3.7020	0.1619	3.3847 4.0194	522.85	<.0001
randgrp1	1	-0.2080	0.1464	-0.4949 0.0789	2.02	0.1553
gender1	1	0.1694	0.1226	-0.0708 0.4096	1.91	0.1670
prevoi1	1	-0.6195	0.1132	-0.8413 -0.3977	29.96	<.0001
stratum1	1	-0.0824	0.0815	-0.2421 0.0772	1.02	0.3115
Scale	0	1.0000	0.0000	1.0000 1.0000		
Weibull Shape	0	1.0000	0.0000	1.0000 1.0000		

Lagrange Multiplier Statistics		
Parameter	Chi-Square	Pr > ChiSq
Scale	49.5704	<.0001

Note that PREVOI1 is also significant in this model, whereas the other effects are not.

Program

Finally, you can use the following code to construct a joint model of both CD4 and T2DEATH. Please refer to the comments embedded within the code for specific explanations of the workflow. Note that the connection between CD4 and T2DEATH is accomplished via the random effects parameters U0 and U1. These are both present in the constructed LINPSURV and LINPLONG variables. The random effects U0 and U1 have mean zero, and the parameters R1 and R2 rescale their variances for LINPSURV in order to account for the different scales of T2DEATH and CD4. The LINPSURV section of the code also illustrates how you can fit parametric frailty models with PROC NLMIXED. The code produces the results in Output 15.14.

```

/*
/*-----*/
/*      Combine the estimates from the separate models.      */
/*      You can use this data set for comparison with the    */
/*      estimates in the joint model and to provide          */
/*      starting values in the joint model.                  */
/*-----*/
data SeparateEstimates; set solf cp survpe;
length Parameter $5;
select ;
when (Effect      = 'Intercept'           ) Parameter = 'b10';
when (Effect      = 'obstime'              ) Parameter = 'b11';
when (Effect      = 'obstime*randgrp1') Parameter = 'b12';
when (Effect      = 'gender1'              ) Parameter = 'b13';
when (Effect      = 'prevoi1'              ) Parameter = 'b14';
when (Effect      = 'stratum1'             ) Parameter = 'b15';
when (CovParm     = 'FA(1,1)'             ) Parameter = 'a11';
when (CovParm     = 'FA(2,1)'             ) Parameter = 'a12';
when (CovParm     = 'FA(2,2)'             ) Parameter = 'a22';
when (CovParm     = 'Residual'            ) Parameter = 's2' ;
when (SurvEffect= 'Intercept'           ) Parameter = 'bs0';
when (SurvEffect= 'randgrp1'            ) Parameter = 'bs1';
when (SurvEffect= 'gender1'             ) Parameter = 'bs2';
when (SurvEffect= 'prevoi1'             ) Parameter = 'bs3';
when (SurvEffect= 'stratum1'            ) Parameter = 'bs4';
otherwise Parameter = '';
end;
if Parameter ne '';
keep Parameter Estimate StdErr;
rename Estimate=SeparateEstimate StdErr=SeparateStdErr;
run;

/*
-----*/
/*---Model XI in Guo and Carlin (2003) ---*/
/* Fit by maximum likelihood with NL MIXED.      */
/* Marginal log likelihood is computed by Gaussian */
/* quadrature.                                     */
/*-----*/
title1 'Joint Model - NL MIXED';
proc nlmixed data=alldata;

/*---All parameters not assigned starting values ---*/
/*---explicitly are assigned the default value (1)---*/
parameters r1=0.13;

/*---Compute log likelihood contribution of the ---*/
/*---survival data part when the last observation---*/
/*---of a subject is reached.                      */
/*---*/

/*---NOTE: This parameterization yields estimates ---*/
/*---      equivalent to those in LIFEREG.        ---*/
if (last) then do;
linpsurv = bs0 + bs1*randgrp1 +
            bs2*gender1 +
            bs3*prevoi1 +
            bs4*stratum1 + r1*u0 + r2*u1;
alpha   = exp(-linpsurv);
G_t     = exp(-alpha*t2death);
g       = alpha*G_t;
llsurv = (death=1)*log(g) + (death=0)*log(G_t);
end; else llsurv=0;

/*---Cholesky parameterization of the random effects ---*/
/*---variance matrix.                                */

```

```

      /*---This ensures that the variance-covariance matrix---*/
      /*---of the random effects is nonnegative definite. ---*/
      v11 = a11*a11;
      v12 = a11*a12;
      v22 = a12*a12 + a22*a22;

      /*---Compute the contribution of the longitudinal---*/
      /*---part. Every observation in the data set ---*/
      /*---makes a contribution. Notice that conditional---*/
      /*---on the random effects we have independent ---*/
      /*---Gaussian contributions. ---*/
      linplong = (b10 + u0)          +
                 (b11 + u1)*obstime   +
                 b12*obstime*randgrp1 +
                 b13*gender1          +
                 b14*prevoil          +
                 b15*stratum1;

      resid = (cd4-linplong);
      if (abs(resid) > 1.3E100) or (s2 < 1e-12) then do;
         lllong = -1e20;
      end; else do;
         lllong = -0.5*(1.837876 + resid**2 / s2 + log(s2));
      end;

      /*---Any numeric variable in the data set can be used---*/
      /*---as the response in the MODEL statement. It has ---*/
      /*---no bearing on the results. ---*/
      model last ~ general(lllong + ll surv);
      random u0 u1 ~ normal([0, 0],[v11,v12,v22]) subject=patient;

      /*---Compute median of the patient-specific ---*/
      /*---survival distributions. ---*/
      predict (1/alpha)* log(2) out=median;

      /*---Compute the variances and covariance of the random---*/
      /*---effects and obtain their std. errors by the delta ---*/
      /*---method. ---*/
      estimate 'Var[U0]'    v11;
      estimate 'Cov[U0,U1]' v12;
      estimate 'Var[U1]'    v22;

      ods output ParameterEstimates=JointEstimates;
      run;

      /*-----*/
      /*---Compare estimates in joint and separate models. ---*/
      /*-----*/
      proc sort data=SeparateEstimates; by Parameter;
      proc sort data=JointEstimates; by Parameter;
      data Estimates; merge SeparateEstimates JointEstimates; by
      Parameter;
      run;

      title1 'Estimates and Standard Errors from Separate and Joint
      Analysis';
      proc print data=Estimates label;
      label SeparateEstimate = 'MLE in separate models'
            SeparateStdErr   = 'Asymptotic Std. Error of MLE in
            separate models'

```

```

      Estimate      = 'MLE in joint model'
      StandardError = 'Asymptotic Std. Error of MLE in joint
                        model';
      var SeparateEstimate SeparateStdErr
          Estimate      StandardError;
          id Parameter;
run;

/*-----*/
/*---Extract predicted patient-specific median      ---*/
/*-----*/
title1 'Patient Specific Modes and Medians for Patients 450 and
454';
title2 'With Standard Errors and Confidence Intervals';
proc print data=median(where=( (patient=450) or (patient=454) ) and
(last=1));
    var patient pred stderrpred lower upper;
run;

```

Results

Output 15.14 Results of Joint Fit to CD4 and T2DEATH

Joint Model - NLMIXED

The NLMIXED Procedure

Specifications	
Data Set	WORK.ALldata
Dependent Variable	last
Distribution for Dependent Variable	General
Random Effects	u0 u1
Distribution for Random Effects	Normal
Subject Variable	patient
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	1408
Observations Not Used	0
Total Observations	1408
Subjects	467
Max Obs Per Subject	5
Parameters	17
Quadrature Points	1

Parameters																	
r1	bs0	bs1	bs2	bs3	bs4	r2	a11	a12	a22	bl0	bl1	bl2	bl3	bl4	bl5	s2	NegLogLike
0.13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20567.16	

Iteration History						
Iter		Calls	NegLogLike	Diff	MaxGrad	Slope
1		3	7275.281	13291.88	1688.291	-4205784
2		5	5846.56004	1428.721	513.3119	-516.448
3		7	5154.59734	691.9627	683.3737	-465.014
4		8	4972.4203	182.177	220.7717	-388.16
5		11	4784.13246	188.2878	489.8002	-219.963
6		12	4605.5938	178.5387	237.5976	-447.908
7		15	4549.67781	55.916	327.6178	-180.756
8		17	4525.37066	24.30715	235.9264	-131.657
9		18	4490.61512	34.75553	121.6993	-106.709
10		20	4481.65434	8.960788	173.7602	-31.632
11		22	4474.69724	6.957101	71.12838	-15.5419
12		23	4464.57392	10.12332	108.2756	-8.41254
13		25	4435.68406	28.88986	90.03952	-15.0859
14		26	4418.53129	17.15277	291.5043	-20.8023
15		27	4393.09471	25.43658	98.9816	-41.388
16		29	4375.44435	17.65036	46.68809	-22.951
17		31	4367.30831	8.136041	41.60579	-11.1667
18		33	4364.4233	2.885006	54.82507	-2.42661
19		35	4353.8619	10.5614	109.89	-3.48345
20		36	4349.18942	4.67248	143.8184	-10.2733
21		37	4344.29231	4.897104	27.1274	-8.6083
22		39	4342.05523	2.237089	19.85248	-1.51811
23		41	4330.09465	11.96057	78.45915	-2.61138
24		43	4325.72135	4.373303	32.56913	-6.02955
25		45	4323.961	1.760353	25.9384	-1.23241
26		47	4318.58764	5.373358	66.41182	-2.78818
27		48	4309.99812	8.589521	48.04017	-10.0093
28		50	4307.11754	2.880574	24.96392	-4.60301
29		51	4303.89771	3.219835	55.27574	-1.78307
30		53	4293.83098	10.06673	102.4151	-4.6509
31		55	4290.67229	3.158687	14.13309	-4.3957
32		57	4289.77719	0.895109	35.861	-0.90052
33		59	4285.02063	4.756558	48.80174	-1.17547

Iteration History						
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope	
34	61	4282.54118	2.479447	21.87464	-3.3123	
35	63	4282.00702	0.534161	6.182027	-0.86295	
36	65	4281.92113	0.085893	5.239593	-0.11599	
37	67	4281.68323	0.237894	21.69237	-0.04775	
38	69	4281.07846	0.604769	3.241209	-0.31084	
39	71	4281.02894	0.049521	3.335275	-0.06639	
40	73	4280.76966	0.259283	13.87388	-0.02651	
41	74	4280.52521	0.244455	6.067981	-0.24615	
42	76	4280.49756	0.027643	0.930731	-0.05032	
43	77	4280.48206	0.015497	3.432418	-0.00386	
44	79	4280.44538	0.036687	1.81302	-0.02043	
45	81	4280.4373	0.00808	0.846975	-0.01209	
46	83	4280.42602	0.011276	3.1078	-0.00083	
47	85	4280.24641	0.179608	1.680393	-0.02164	
48	87	4280.19897	0.047446	1.925969	-0.0843	
49	89	4280.19101	0.007954	1.7452	-0.00427	
50	91	4280.1263	0.064718	2.800298	-0.00935	
51	93	4280.11817	0.008123	0.280778	-0.01397	
52	95	4280.11813	0.000041	0.031838	-0.00008	
53	97	4280.11813	2.443E-7	0.001393	-5.09E-7	

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	8560.2
AIC (smaller is better)	8594.2
AICC (smaller is better)	8594.7
BIC (smaller is better)	8664.7

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
r1	0.2068	0.03098	465	6.67	<.0001	0.05	0.1459	0.2676	0.000303
bs0	4.0158	0.1995	465	20.13	<.0001	0.05	3.6238	4.4077	0.000064
bs1	-0.2666	0.1555	465	-1.71	0.0870	0.05	-0.5722	0.03891	6.292E-7
bs2	0.1472	0.1423	465	1.03	0.3015	0.05	-0.1325	0.4269	0.000147

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
bs3	-0.7712	0.1304	465	-5.91	<.0001	0.05	-1.0275	-0.5150	-0.00005
bs4	-0.08467	0.09560	465	-0.89	0.3762	0.05	-0.2725	0.1032	-0.00025
r2	2.5083	0.9163	465	2.74	0.0064	0.05	0.7076	4.3089	-0.00003
a11	3.9899	0.1465	465	27.23	<.0001	0.05	3.7020	4.2778	-0.00011
a12	-0.02273	0.01522	465	-1.49	0.1360	0.05	-0.05263	0.007174	-0.00048
a22	-0.1706	0.01720	465	-9.92	<.0001	0.05	-0.2044	-0.1368	0.001326
bl0	7.9999	0.3525	465	22.69	<.0001	0.05	7.3072	8.6927	-0.00004
bl1	-0.1879	0.02143	465	-8.77	<.0001	0.05	-0.2300	-0.1458	-0.00139
bl2	0.02183	0.02925	465	0.75	0.4558	0.05	-0.03565	0.07931	-0.00114
bl3	-0.1254	0.3268	465	-0.38	0.7015	0.05	-0.7676	0.5168	-0.00002
bl4	-2.3483	0.2400	465	-9.78	<.0001	0.05	-2.8201	-1.8766	0.000054
bl5	-0.1372	0.2368	465	-0.58	0.5627	0.05	-0.6026	0.3282	0.000184
s2	3.0370	0.1681	465	18.07	<.0001	0.05	2.7068	3.3673	-0.00008

Additional Estimates									
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	
Var[U0]	15.9190	1.1692	465	13.62	<.0001	0.05	13.6215	18.2165	
Cov[U0,U1]	-0.09067	0.06169	465	-1.47	0.1423	0.05	-0.2119	0.03054	
Var[U1]	0.02961	0.006032	465	4.91	<.0001	0.05	0.01776	0.04147	

Estimates and Standard Errors from Separate and Joint Analysis

Parameter	MLE in separate models	Asymptotic Std. Error of MLE in separate models	MLE in joint model	Asymptotic Std. Error of MLE in joint model
a11	3.9889	0.1467	3.9899	0.1465
a12	-0.03259	0.01507	-0.02273	0.01522
a22	0.1658	0.01703	-0.1706	0.01720
bl0	8.0129	0.3511	7.9999	0.3525
bl1	-0.1668	0.02038	-0.1879	0.02143
bl2	0.02998	0.02891	0.02183	0.02925
bl3	-0.1582	0.3249	-0.1254	0.3268
bl4	-2.3152	0.2382	-2.3483	0.2400
bl5	-0.1309	0.2352	-0.1372	0.2368
bs0	3.7020	0.1619	4.0158	0.1995

Parameter	MLE in separate models	Asymptotic Std. Error of MLE in separate models	MLE in joint model	Asymptotic Std. Error of MLE in joint model
bs1	-0.2080	0.1464	-0.2666	0.1555
bs2	0.1694	0.1226	0.1472	0.1423
bs3	-0.6195	0.1132	-0.7712	0.1304
bs4	-0.08245	0.08147	-0.08467	0.09560
r1	.	.	0.2068	0.03098
r2	.	.	2.5083	0.9163
s2	3.0715	0.1713	3.0370	0.1681

*Patient Specific Modes and Medians for Patients 450 and 454
With Standard Errors and Confidence Intervals*

Obs	patient	Pred	StdErrPred	Lower	Upper
1360	450	71.5799	26.7589	18.9965	124.163
1372	454	24.0501	6.8419	10.6052	37.495

Interpretation

The model has 17 parameters and requires 53 iterations to converge using adaptive Gaussian quadrature with one point (equivalent to a Laplace approximation) to compute the marginal likelihood.

Note in particular the output table comparing the parameter estimates and approximate standard errors from the separate and joint analyses. One main conclusion for this example is that the joint model estimates are nearly all identical to those from the separate analyses. One exception is the decrease in the significant survival coefficient for PREVOI1 (parameter BS3 in the table), which decreased from -0.62 to -0.77 over a standard error of around 0.12. The similarity of the covariance parameter estimates to those from PROC MIXED alone reveals that the survival data contribute virtually nothing to the estimation of the covariance structure of these data.

15.7 Example: One-Compartment Pharmacokinetic Model

For this example we consider the phenobarbital data from Grasela and Donn (1985), which are also analyzed by Davidian and Gallant (1993). These are routine clinical data collected from 59 newborn infants treated with phenobarbital during the first 16 days after birth.

Program

The data appear as Data Set 15.7, “Phenobarbital,” in Appendix 2, “Data Sets,” and are processed with the following statements:

```
data pheno;
  input indiv time dose weight apgar conc;
  retain cursub .;
  if cursub ne indiv then do;
    newsub = 1;
    cursub = indiv;
  end;
  else newsub = 0;
  if (apgar < 5) then apgarlow = 1;
  else apgarlow = 0;
  tlag = lag(time);
  if (newsub=1) then tlag = 0;
  drop apgar cursub;
  datalines;
...datalines...
run;
```

Table 15.2 describes the variables in the PHENO data set. Note that each time point represents either a dose administration or a concentration measurement.

Table 15.2 Description of Variables in the PHENO Data Set

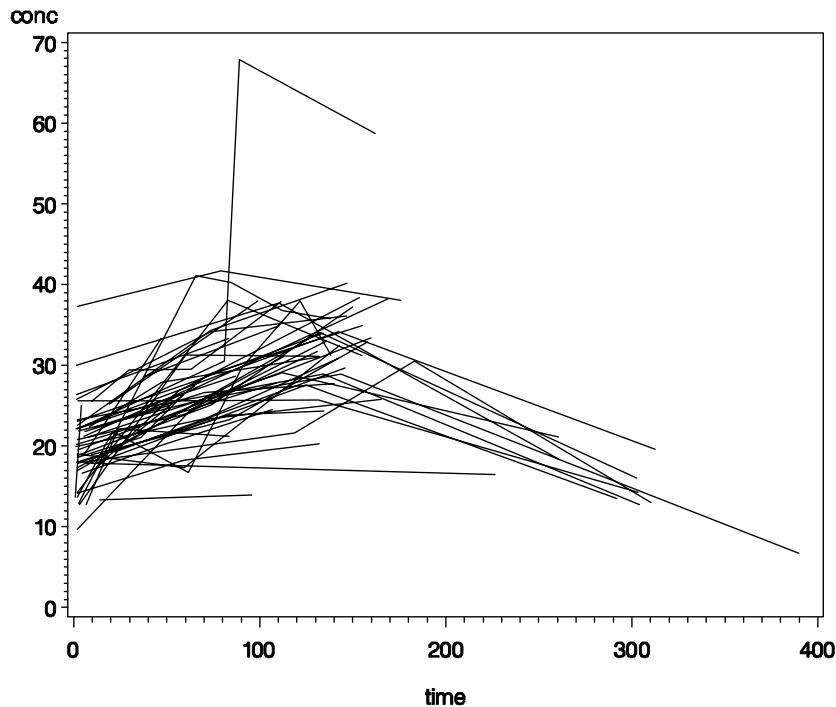
Variable	Description
INDIV	subject
NEWSUB	equals 1 if the current observation is from a new subject, 0 otherwise
TLAG	the previous time value (if NEWSUB=0) or 0.0 (if NEWSUB=1)
DOSE	dose amount (in µg/kg)
WEIGHT	birth weight (in kg)
APGARLOW	equals 1 if 5-minute Apgar score is less than 5, 0 otherwise
CONC	plasma concentration of phenobarbital (in µg/kg)

You can use the following program to reproduce Figure 15.3, which graphs the subject profiles:

```
symbol i=join r=100;
proc gplot data=pheno;
  plot conc*time=indiv / nolegend;
run;
```

Results

Figure 15.3 Phenobarbital Data



Interpretation

Subject 50 is a potential outlier because of its larger values, but it is retained in the analysis for comparison purposes.

Pharmacokinetic Model

The proposed pharmacokinetic model for mean phenobarbital concentration is a one-compartment open model with intravenous bolus administration and first-order elimination. The mean concentration for subject i at time t due to a single dose d_{ij} given at time $t_{ij} < t$ is modeled as

$$(d_{ij}/v_i) \exp[-(t-t_{ij})c_i/v_i]$$

Here c_i is the total clearance of phenobarbital in (liter/hour)/kg, and v_i is the apparent volume of distribution of phenobarbital in liter/kg. The specific models for c_i and v_i in terms of fixed and random effects parameters are as follows:

$$\begin{aligned} c_i &= \beta_1 w_i \exp(u_{1i}) \\ v_i &= \beta_2 w_i (1 + \beta_3 \delta_i) \exp(u_{2i}) \end{aligned}$$

where

β_1 , β_2 , and β_3 are the fixed effects parameters to be estimated

u_{1i} and u_{2i} are the i^{th} subject's random effects having mean 0 and variance-covariance matrix to be estimated

w_i is the i^{th} subject's birth weight

δ equals 1 if the i^{th} subject's 5-minute Apgar score is less than 5 and 0 otherwise

Because each subject receives multiple doses, the concentration at any particular time is computed as the sum of the contributions from all doses received prior to that time. For computational purposes, it is convenient to write this sum recursively as

$$f(d_{ij}, t_{ij}, c_i, v_i) = d_{ij}/v_i + f(d_{ij-1}, t_{ij-1}, c_i, v_i) \exp[-(t_{ij-1})c_i/v_i]$$

where $f(d_{ij}, t_{ij}, c_i, v_i)$ is the concentration for subject i immediately after receiving dose d_{ij} at time t_{ij} . This expression also holds for times at which no dose is given if one simply sets $d_{ij} = 0$ for these times. This fact is exploited in the coding specifications below.

Letting Y_{ik} denote the k^{th} observed concentration for the i^{th} subject, the complete stochastic model is

$$Y_{ik} = f(d_{ik}, t_{ik}, c_i, v_i) + e_{ik}$$

where the d_{ik} and t_{ik} are the subset of the d_{ij} and t_{ij} for which there are observed measurements. Note here that c_i and v_i are functions of β_1 , β_2 , and β_3 , u_{1i} , u_{2i} . Finally, e_{ik} are unobserved Gaussian errors with mean 0 and variance $\sigma^2[f(d_{ik}, t_{ik}, c_i, v_i)]^2$, implying that the data have a constant coefficient of variation.

Program

The PROC NLMIXED code for the previously described model is as follows. We specify QPOINTS=1 in order to fit a single adaptive Gaussian quadrature point, which is equivalent to a Laplace approximation to the marginal likelihood. Starting values were chosen by trial and error, and β_1 is divided by 100 to improve numerical stability. For the second and subsequent observations for each subject, the ZLAG function is used to model the current value of PRED as a function of the previous value of PRED in time. Note the data must be sorted by INDIV and TIME for this to work correctly. You should always use ZLAG instead of LAG when programming in PROC NLMIXED. The code produces the results in Output 15.15.

```

proc nlmixed data=pheno qpoints=1;
  * starting values;
  parms beta1=0.5 beta2=1 beta3=0.15
    s2u1=0.05 s2u2=0.03 cu12=0.015 s2e=0.01;

  * bounds;
  bounds beta1>=0, beta2>=0, s2u1>=0, s2u2>=0, s2e>0;

  * compute key terms in the model;
  clear = beta1/100*weight*exp(u1);
  vol   = beta2*weight*(1+beta3*apgarlow)*exp(u2);
  eterm = exp(-(time-tlag)*clear/vol);

  * compute predicted value depending on whether this is a
    new subject or not;
  if (newsub = 1) then pred = dose/vol;
  else pred = dose/vol + zlag(pred)*eterm;

```

```

* variance goes up with the square of the predicted value,
  a constant coefficient of variation model;
var = s2e*(pred**2);

* model specification;
model conc ~ normal(pred,var);

* random effects specification;
random u1 u2 ~ normal([0,0],[s2u1,cu12,s2u2])
  subject=indiv out=ebayes;

* compute an additional estimate of the correlation;
estimate 'corr(u1,u2)' cu12/sqrt(s2u1*s2u2);

* output subject-specific predictions of clearance and volume;
predict clear out=clear;
predict vol out=vol;
run;

```

Results

Output 15.15 Results for PROC NL MIXED Applied to Phenobarbital Data

Specifications	
Data Set	WORK.PHENO
Dependent Variable	conc
Distribution for Dependent Variable	Normal
Random Effects	u1 u2
Distribution for Random Effects	Normal
Subject Variable	indiv
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	155
Observations Not Used	589
Total Observations	744
Subjects	59
Max Obs Per Subject	6
Parameters	7
Quadrature Points	1

Parameters							
beta1	beta2	beta3	s2u1	s2u2	cu12	s2e	NegLogLike
0.5	1	0.15	0.05	0.03	0.015	0.01	437.759491

Iteration History						
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope	
1	5	436.624676	1.134816	249.0696	-10645.5	
2	9	435.107501	1.517175	225.3187	-371.212	
3	11	434.015793	1.091707	121.1016	-48.9049	
4	14	433.996886	0.018907	105.5664	-10.0008	
5	16	433.941294	0.055592	90.84407	-2.21844	
6	18	433.80331	0.137984	80.67595	-1.12135	
7	20	433.764465	0.038845	89.06237	-1.32549	
8	21	433.708496	0.055968	19.46526	-0.10437	
9	23	433.705581	0.002916	5.191272	-0.00864	
10	25	433.704789	0.000791	3.106978	-0.0016	
11	27	433.704749	0.00004	0.296499	-0.00007	
12	29	433.704746	3.366E-6	0.115745	-5.74E-6	
13	31	433.704746	6.569E-8	0.001814	-9.36E-8	

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	867.4
AIC (smaller is better)	881.4
AICC (smaller is better)	882.2
BIC (smaller is better)	896.0

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta1	0.4738	0.02202	57	21.52	<.0001	0.05	0.4297	0.5179	-0.00072
beta2	0.9756	0.02549	57	38.28	<.0001	0.05	0.9246	1.0267	0.000067
beta3	0.1380	0.07437	57	1.86	0.0687	0.05	-0.01094	0.2869	0.000049
s2u1	0.03707	0.01765	57	2.10	0.0402	0.05	0.001721	0.07243	0.000423
s2u2	0.02171	0.006737	57	3.22	0.0021	0.05	0.008221	0.03520	-0.00062
cu12	0.01903	0.009439	57	2.02	0.0485	0.05	0.000133	0.03793	0.000755
s2e	0.01320	0.002152	57	6.13	<.0001	0.05	0.008889	0.01751	0.001814

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
corr(u1,u2)	0.6709	0.2703	57	2.48	0.0160	0.05	0.1297	1.2121

Interpretation

The procedure converges in 13 iterations. The Fit Statistics entries are useful ways to compare different model fits. The estimates of β_1 and β_2 are significantly different from zero, whereas that for β_3 is not. The correlation between the random effects also appears to be significantly larger than zero.

Program

To visualize this, we plot the pairs of empirical Bayes estimates against one another using the following code. This produces the plot in Figure 15.4.

```

proc transpose data=ebayes out=ebt;
  var estimate;
  by indiv;
  id effect;
run;

data ebt;
  set ebt;
  format u1 u2 4.1;
run;

goptions reset=all
  hsize   = 5 in
  vsize   = 5 in
  htext   = 3.0 pct
  htitle  = 3.5 pct
  vorigin = 0 in
  horigin = 0 in
  ftext   = swiss
  lfactor = 1;

axis1 order=(-0.3 to 0.5 by 0.1) value=(font=swissl height=1)
  label=(font=swissl h=1) major=(h=0.5)minor=none;

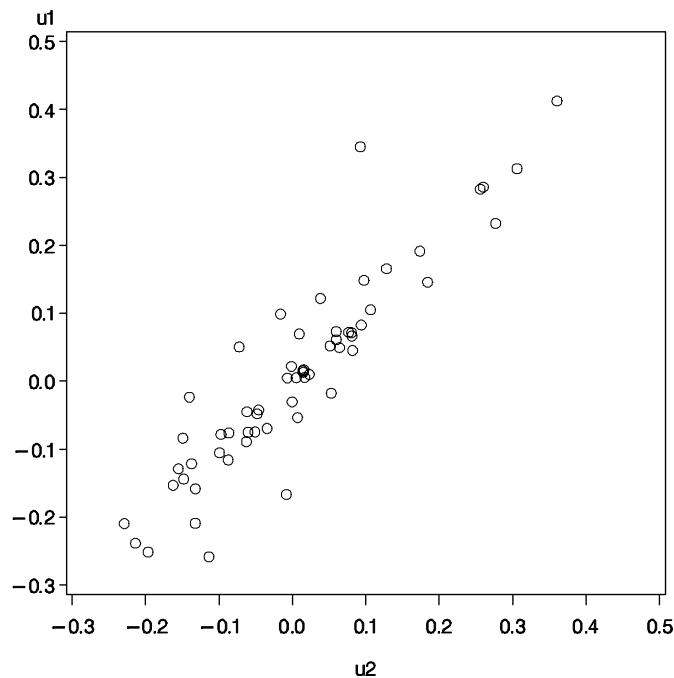
symbol v=circle;

proc gplot data=ebt;
  plot u1*u2 / haxis=axis1 vaxis=axis1;
run;

```

Results

Figure 15.4 Plot of Empirical Bayes Estimates from PROC NLMIXED of Pharmacokinetics Model Applied to Phenobarbital Data



Interpretation

Figure 15.4 shows how the clearance and volume random effects are strongly positively correlated; that is, subjects with higher clearance also have higher volume.

Program

To further visualize this correlation, the following code plots the predicted subject-specific estimates of clearance and volume against one another. It produces the plot in Figure 15.5.

```

data clear1;
  set clear;
  if time = 0;
  clearance = pred;
  format clearance 4.1;
  keep indiv clearance;
run;

data vol1;
  set vol;
  if time = 0;
  volume = pred;
  format volume 4.1;
  keep indiv volume;
run;

data cv;
  merge clear1 vol1;
  by indiv;
run;

goptions reset=all;
  hsize   = 5 in

```

```

vsize    = 5 in
htext    = 3.0 pct
htitle   = 3.5 pct
vorigin  = 0 in
horigin  = 0 in
ftext    = swiss
lfactor  = 1;

axis1 order=(0 to 5 by 1)
       value=(font=swissl height=1)
       label=(font=swissl h=1) major=(h=0.5) minor=none;

axis2 order=(0 to 0.025 by 0.005)
       value=(font=swissl height=1)
       label=(font=swissl h=1) major=(h=0.5) minor=none;

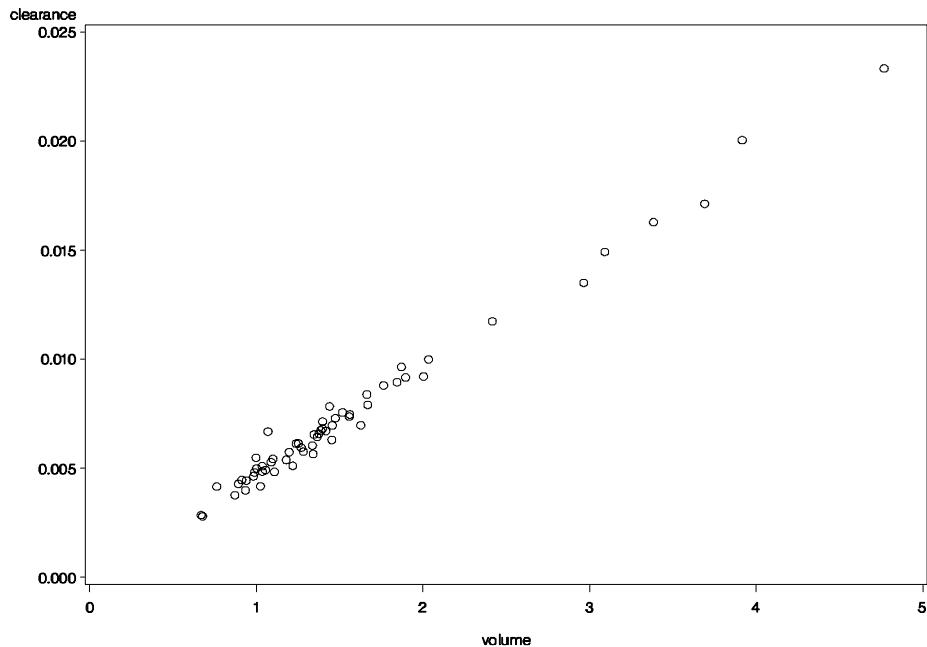
symbol v=circle;

proc gplot data=cv;
  plot clearance*volume / haxis=axis1 vaxis=axis2;
run;

```

Results

Figure 15.5 Plot of Clearance and Volume Estimates from PROC NLMIXED of Pharmacokinetics Model Applied to Phenobarbital Data



Interpretation

In Figure 15.5, note the correlation is even stronger between clearance and volume because these estimates also include the correlation between their respective fixed effects estimates.

The constant-coefficient-of-variance model for the errors fitted previously depends directly on the random effects. This can occasionally cause some instabilities in the optimization process. If you want to remove this dependence, replace the line

```
var = (sig*pred)**2;
```

with code like the following:

```

clear0 = beta1/100*weight;
vol0   = beta2*weight*(1+beta3*apgarlow);
eterm0 = exp(-(time-tlag)*clear0/vol0);
if (newsub = 1) then
    pred0 = dose/vol0;
else pred0 = dose/vol0 + zlag(pred0)*eterm0;
var    = (sig*pred0)**2;

```

Program

%NLINMIX macro code to fit a comparable model with method 2 is as follows. If you are not interested in using the macro, you can skip the remainder of this example.

```

%nlinmix(data=pheno,
modinit=%str(
array plag{11} _temporary_;
if (newsub=1) then do call = 1 to 11;
    plag[call] = 0;
end;
call = 1;
),
model=%str(
    clear = beta1/100*weight*exp(u1);
    vol   = beta2*weight*(1+beta3*apgarlow)*exp(u2);
    eterm = exp(-(time-tlag)*clear/vol);
    predv = dose/vol + plag[call]*eterm;
    plag[call] = predv;
    call = call + 1;
),
weight=%str(
    wghtv = 1/predv**2;
),
parms=%str(beta1=0.5 beta2=1 beta3=0.15),
stmts=%str(
    class indiv;
    model pseudo_conc = d_beta1 d_beta2 d_beta3 /
                    noint notest solution cl;
    random d_u1 d_u2 / subject=indiv type=un solution;
    weight wghtv;
),
expand=eblup
)
run;

```

The heart of the code is in the MODEL= specification. Note that the CLEAR, VOL, and ETERM variables are all based directly on the theoretical model. U1 and U2 are the random effects parameters, allowing population variance of clearance and volume. The computation of the required PREDV variable uses the recursion relation described previously, but some special coding is required in order to implement it correctly.

Because no derivatives are specified, %NLINMIX computes them numerically. In order for %NLINMIX to carry out the finite differencing correctly for each observation, it must retain two previous values of PRED for each of the fixed and random effect parameters. This results in 10 values to save, and these, plus the original prediction, make 11. Therefore, the array PLAG is created under the MODINIT= option with 11 elements.

The MODINIT= option is called only once at the beginning of the DATA step processing for each observation, and is available for applications like this one. The _TEMPORARY_ variables are used in PLAG because they are automatically retained and are not saved as a part of the data. The NEWSUB variable is used to initialize them to 0 for each new subject, and the CALL variable is used to index evaluations of the model. %NLINMIX uses the first element of PLAG to save the actual prediction, and the other 10 values hold results from numerical differentiation. The PLAG array is unnecessary if you specify your own derivatives with the DERIVS= option; however, you would have to use the RETAIN statement to keep track of the previous value of PRED.

The WEIGHT= argument specifies a constant coefficient of variation model, weighting by the reciprocal of the squared prediction.

The PARMS= argument specifies initial values for the fixed effects parameters.

The STMTS= block provides the appropriate PROC MIXED statements. Note that the derivatives of the fixed effects parameters are specified in the MODEL statement and those for the random effects parameters are in the RANDOM statement. By default their covariance structure is a simple variance components model; that is, U1 and U2 are assumed to be independent. EXPAND=ZERO requests the method 1 expansion around 0.

Results

The SAS log from this program is shown in Output 15.16.

Output 15.16 SAS Log for %NLINMIX Macro—Method 2 Applied to Phenobarbital Data

The NLINMIX Macro

Data Set	:	pheno
Response	:	conc
Fixed-Effect Parameters	:	beta1 beta2 beta3
Random-Effect Parameters	:	u1 u2
Expansion Point	:	eblup

```

Calling PROC NLIN to initialize.
beta1 = beta1
beta2 = beta2
beta3 = beta3
WARNING: PROC NLIN failed to converge.
Iteratively calling PROC MIXED.
PROC MIXED call 0

iteration = 0
convergence criterion = .
beta1=0.5 beta2=1 beta3=0.15 COVP1=0.0479929388 COVP2=0.0160310936
COVP3=0.0259354362
COVP4=0.0120242645
PROC MIXED call 1

iteration = 1
convergence criterion = 0.0199867091
beta1=0.4640049844 beta2=0.9690620313 beta3=0.1253363764
COVP1=0.0363928522 COVP2=0.0196966744
COVP3=0.0241398569 COVP4=0.0129790897
PROC MIXED call 2

```

```

iteration = 2
convergence criterion = 0.0063328689
beta1=0.4839916935 beta2=0.9748724732 beta3=0.1370469759
COVP1=0.0423760915 COVP2=0.0196403437
COVP3=0.0246509048 COVP4=0.0128202988
PROC MIXED call 3

iteration = 3
convergence criterion = 0.0029705656
beta1=0.4776588246 beta2=0.9784249077 beta3=0.1334854206
COVP1=0.040810045 COVP2=0.02037608461
COVP3=0.0242791249 COVP4=0.0130426739
PROC MIXED call 4

iteration = 4
convergence criterion = 0.0011578996
beta1=0.4806293902 beta2=0.9769017307 beta3=0.1347971255
COVP1=0.0417428185 COVP2=0.0202080285
COVP3=0.0243757855 COVP4=0.0129446734
PROC MIXED call 5

iteration = 5
convergence criterion = 0.0005553039
beta1=0.4794714906 beta2=0.9774992951 beta3=0.1341357994
COVP1=0.0414068278 COVP2=0.0203309416
COVP3=0.0243056947 COVP4=0.012989028
PROC MIXED call 6

iteration = 6
convergence criterion = 0.0002334218
beta1=0.4800267944 beta2=0.9772152365 beta3=0.1343883958
COVP1=0.0415843123 COVP2=0.0202903417
COVP3=0.0243295665 COVP4=0.01296888
PROC MIXED call 7

iteration = 7
convergence criterion = 0.0001066178
beta1=0.4797933726 beta2=0.9773349672 beta3=0.1342641083
COVP1=0.041518075 COVP2=0.0203114568
COVP3=0.0243169993 COVP4=0.0129774169
PROC MIXED call 8

iteration = 8
convergence criterion = 0.0000448063
beta1=0.4798999904 beta2=0.9772805717 beta3=0.1343151107
COVP1=0.0415501739 COVP2=0.0203034087
COVP3=0.0243217974 COVP4=0.012973658
PROC MIXED call 9

iteration = 9
convergence criterion = 0.0000201677
beta1=0.4798551841 beta2=0.9773035951 beta3=0.134291597
COVP1=0.0415369834 COVP2=0.020307329
COVP3=0.0243194723 COVP4=0.0129753063
PROC MIXED call 10

iteration = 10
convergence criterion = 8.6559212E-6
beta1=0.4798753518 beta2=0.9772932829 beta3=0.134301489
COVP1=0.0415430058 COVP2=0.0203057425
COVP3=0.0243204183 COVP4=0.0129745868
PROC MIXED call 11

```

```

iteration = 11
convergence criterion = 3.8394755E-6
beta1=0.4798666959 beta2=0.9772977225 beta3=0.1342970399
COVP1=0.0415404438 COVP2=0.0203064777
COVP3=0.024319982 COVP4=0.0129749024
PROC MIXED call 12

iteration = 12
convergence criterion = 1.6647442E-6
beta1=0.4798705353 beta2=0.9772957572 beta3=0.1342989508
COVP1=0.0415415871 COVP2=0.0203061688
COVP3=0.024320166 COVP4=0.0129747646
PROC MIXED call 13

iteration = 13
convergence criterion = 7.3327243E-7
beta1=0.4798688706 beta2=0.9772966104 beta3=0.1342981035
COVP1=0.0415410934 COVP2=0.020306308
COVP3=0.0243200833 COVP4=0.012974825
PROC MIXED call 14

iteration = 14
convergence criterion = 1.9062927E-7
beta1=0.4798696039 beta2=0.977296235 beta3=0.134298471
COVP1=0.0415410876 COVP2=0.0203063051
COVP3=0.0243200799 COVP4=0.0129748232
PROC MIXED call 15

iteration = 15
convergence criterion = 2.9457028E-8
beta1=0.4798694132 beta2=0.9772963303 beta3=0.1342983713
COVP1=0.0415410871 COVP2=0.0203063049
COVP3=0.0243200796 COVP4=0.012974823
PROC MIXED call 16

iteration = 16
convergence criterion = 6.7816756E-9
beta1=0.4798694427 beta2=0.9772963115 beta3=0.1342983987
COVP1=0.0415410869 COVP2=0.0203063048
COVP3=0.0243200795 COVP4=0.0129748229
NLINMIX convergence criteria met.

NOTE: Numeric values have been converted to character values at the
      places given by:
      (Line):(Column).
      21:71
NOTE: Missing values were generated as a result of performing an
      operation on missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      589 at 7:61
NOTE: There were 744 observations read from the data set
      WORK._NLINMIX.
NOTE: The data set WORK._NLINMIX has 744 observations and 31
      variables.
NOTE: DATA statement used (Total process time):
      real time          0.06 seconds
      cpu time          0.07 seconds

PROC MIXED call 17

NOTE: 589 observations are not included because of missing values.
NOTE: Convergence criteria met.

```

NOTE: The data set WORK._SOLNR has 118 observations and 7 variables.

NOTE: The data set WORK._COV has 4 observations and 3 variables.

NOTE: The data set WORK._SOLN has 3 observations and 9 variables.

NOTE: The data set WORK._FIT has 4 observations and 2 variables.

NOTE: PROCEDURE MIXED used (Total process time):

real time	0.36 seconds
cpu time	0.19 seconds

The %NLINMIX macro first echoes several specifications and then prints the dimensions of the problem. The calculations begin with an initial call to PROC NLIN; however, it fails to converge. This may be caused by inaccuracies due to numerical differentiation, but nonetheless, the final estimates of β_1 , β_2 , and β_3 serve as reasonable starting values for the PROC MIXED iterations.

The %NLINMIX macro requires 16 calls to PROC MIXED to drop below the convergence tolerance of 1E-8.

The WORK._NLINMIX data set is used as input for the sixth and final call to PROC MIXED, and four data sets are output as a result. The WORK._SOLN data set contains the fixed effects parameter estimates, and WORK._SOLNR contains the EBLUPs. WORK._COV contains the variance-covariance parameter estimates, and WORK._FIT contains model-fitting information. %NLINMIX creates these data sets using the Output Delivery System (ODS).

The output from the final call to PROC MIXED is shown in Output 15.17.

Output 15.17 Results for %NLINMIX Macro—Method 2 Applied to Phenobarbital Data

The Mixed Procedure

Model Information	
Data Set	WORK._NLINMIX
Dependent Variable	pseudo_conc
Weight Variable	wghtv
Covariance Structure	Unstructured
Subject Effect	indiv
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions	
Covariance Parameters	4
Columns in X	3
Columns in Z Per Subject	2
Subjects	59
Max Obs Per Subject	20

Number of Observations	
Number of Observations Read	744
Number of Observations Used	155
Number of Observations Not Used	589

Parameter Search						
CovP1	CovP2	CovP3	CovP4	Variance	Res Log Like	-2 Res Log Like
0.04154	0.02031	0.02432	0.01297	0.01297	-441.1632	882.3265

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
1	1	882.32646755	0.00000000

Convergence criteria met.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	indiv	0.04154
UN(2,1)	indiv	0.02031
UN(2,2)	indiv	0.02432
Residual		0.01297

Fit Statistics	
-2 Res Log Likelihood	882.3
AIC (smaller is better)	890.3
AICC (smaller is better)	890.6
BIC (smaller is better)	898.6

PARMS Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
3	0.00	1.0000

Solution for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
d_beta1	0.4799	0.02151	46	22.31	<.0001	0.05	0.4366	0.5232
d_beta2	0.9773	0.02625	46	37.23	<.0001	0.05	0.9245	1.0301
d_beta3	0.1343	0.07623	46	1.76	0.0847	0.05	-0.01914	0.2877

Interpretation

The approximate REML estimates of the covariance parameters agree very well with the Laplace-approximated maximum likelihood estimates obtained previously using PROC NLMIXED. These values and their estimated standard errors agree fairly well with the estimates reported by Grasela and Donn (1985) and Davidian and Gallant (1993).

The “Solution for Random Effects” table is omitted because of its length (118 rows). It contains 2 EBLUPs for each subject.

Program

You can create a scatter plot of the EBLUPs with the following program. This program produces the plot in Figure 15.6.

```
/*---code to plot EBLUPs---*/
proc transpose data=_solnr out=eblups;
  var estimate;
  by indiv;
  id effect;
run;

data eblups;
  set eblups;
  format d_u1 d_u2 4.1;
run;

goptions reset=all
  hsize   = 5 in
  vsize   = 5 in
  htext   = 3.0 pct
  htitle  = 3.5 pct
  vorigin = 0 in
  horigin = 0 in
  ftext   = swiss
  lfactor = 1;

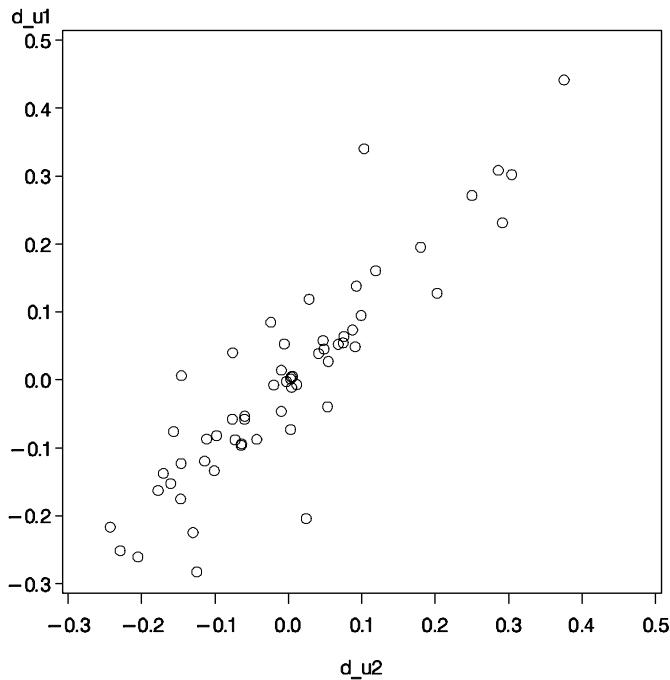
axis1 order=(-0.3 to 0.5 by 0.1)
  value=(font=swissl height=1)
  label=(font=swissl h=1)
  major=(h=0.5)
  minor=none;

symbol v=circle;

proc gplot data=eblups;
  plot d_u1*d_u2 / haxis=axis1 vaxis=axis1;
run;
```

Results

Figure 15.6 Plot of EBLUPs from %NLINMIX Macro—Method 2 Applied to Phenobarbital Data



Interpretation

Figure 15.6 is very similar to Figure 15.4, indicating strong agreement between the two fitting methods.

15.8 Comparison of PROC NLMIXED and the %NLINMIX Macro

Since publication of the first edition of this book in 1996, SAS has both updated the %NLINMIX macro and released PROC NLMIXED in SAS/STAT. The latter offers all the advantages of a SAS procedure, including BY-group processing and access to powerful nonlinear modeling syntax like that found in the NLIN and MODEL procedures; however, it was not designed to be a complete replacement for the %NLINMIX macro. Table 15.3 provides a comparison with respect to several key aspects.

Table 15.3 Comparison of the %NLINMIX Macro and PROC NLMIXED

Aspect	%NLINMIX Macro	PROC NLMIXED
Syntax	Based on SAS DATA step code and PROC MIXED statements	Customized SAS procedural grammar
Fitting methods	Linearization-based, iteratively solves a set of estimating equations by fitting pseudo-data with PROC MIXED, REML or ML	Maximizes an objective function computed by direct numerical quadrature or a linear approximation, ML only
Distribution of response variable	Normal	A variety of common distributions, plus capability to write a general likelihood function
Covariance structures	Full complement of structures via REPEATED or multiple RANDOM statements	Offers a single RANDOM statement, typically used for nonlinear random coefficient models
Starting values	Requires starting values for fixed effects parameters only	Requires starting values for all parameters

Also, the syntax for the %NLINMIX macro has been expanded to allow more flexibility in model specification. In particular, the previous RESPONSE=, SUBJECT=, and RANDOM= arguments have been replaced by the STMTS= argument that accepts PROC MIXED statements. For the syntax of PROC NLMIXED, refer to the SAS/STAT documentation. The examples below provide some additional templates beyond those discussed in the documentation.

A critical difference in the two programs is their method of fitting the nonlinear mixed model to the data. As discussed more thoroughly below, the %NLINMIX macro iteratively fits a set of linear estimating equations using PROC MIXED as a computing engine at each stage. In contrast, PROC NLMIXED constructs a marginal likelihood function and maximizes it using numerical methods.

PROC NLMIXED is more general than the %NLINMIX macro with respect to the assumed probability distribution of the response variable, conditioned on the random effects. The macro assumes normality, whereas the procedure offers a full complement of standard distributions along with the capability to “roll your own” likelihood function.

On the other hand, the macro offers a wider range of covariance structures and effectively extends all of PROC MIXED to the nonlinear case. The procedure only allows one RANDOM statement and no REPEATED statement.

So which one should you use for your nonlinear mixed modeling problem? A rule of thumb is to use PROC NLINMIXED as a first choice, and the macro only when necessary. The macro is typically needed to fit complex covariance structures, such as when you have two sets of crossed random effects. In addition, an update to the macro as described in Galecki et al. (2004) allows you to specify models in terms of ordinary differential equations, functionality especially useful for kinetic models.

15.9 Three General Fitting Methods Available in the %NLINMIX Macro

The three linearization-based methods available in the %NLINMIX macro all use the same basic algorithm, which is to iterate between the following two steps until the parameter estimates converge:

1. Create pseudo-data, usually modified residuals computed from the original data and the assumed model evaluated at current estimates of the parameters.
2. Call PROC MIXED using the pseudo-data and obtain new estimates of the parameters.

The three methods differ in how they create the pseudo-data and in their PROC MIXED specifications. Table 15.4 outlines the major differences.

Table 15.4 Three Methods Available in the NLINMIX Macro

Method	Name	Pseudo-data	PROC MIXED Statement Used
1	Approximate first-order	residuals+fixed	RANDOM
2	Approximate second-order	residuals+fixed+random	RANDOM
3	Marginal structures	residuals+fixed	REPEATED

Method 1 results in estimates similar to those from the first-order method described in Beal and Sheiner (1982) and Sheiner and Beal (1985), as it is based on a Taylor series expansion around zero, which is the expected value of the random effects. Method 2 produces the estimates of Lindstrom and Bates (1990), although using a different algorithm from the one they describe. It is based on a Taylor series expansion around the empirical BLUPs of the random effects (Vonesh and Chinchilli 1997, Wolfinger and Lin 1997). Method 3 has close ties with the population-averaged generalized estimating equation (GEE) approach of Zeger, Liang, and Albert (1988) and is based on a marginal approximation to the nonlinear mixed model.

None of the three methods is clearly superior to the others for all data analytic situations. Table 15.5 provides a comparison of some of the advantages and disadvantages of each.

Table 15.5 Advantages and Disadvantages of the Three Methods

Method	Advantages	Disadvantages
1	Both subject-specific and population-averaged inference; more robust to model misspecification; converges quickly	Approximation may be inaccurate
2	More accurate approximation, subject-specific inference; weights can depend on random effects	Sensitive to model misspecification; sometimes converges slowly or not at all
3	Population-averaged inference; wide variety of covariance structures helps prevent model misspecification; converges quickly	Can be difficult to interpret and to compute predictions

Because all three methods are implemented through calls to PROC MIXED, they accommodate unbalanced and randomly missing data quite well. Please be aware, though, that none of the methods are guaranteed to converge even when your model specification is correct. Upon achieving successful convergence, the statistics produced by the final call to PROC MIXED are asymptotically valid provided that you have correctly specified the first and second moments. By default the standard errors of estimates are based directly on the linearized model, although you can also specify PROCOPT=EMPIRICAL to obtain “empirical sandwich” estimators commonly associated with GEE (Liang and Zeger 1986). The weight matrix \mathbf{W}_i provides a nice mechanism for modeling heterogeneity, especially the type that varies with a power of the mean function (Carroll and Ruppert 1988, Davidian and Giltinan 1993). The EBLUPs from the final PROC MIXED call from methods 1 and 2 are empirical estimates of the random effects.

An option available with all three methods is the use of either maximum likelihood (ML) or restricted maximum likelihood (REML) in the PROC MIXED calls. For method 1, the use of ML is closer to Sheiner and Beal’s first-order method than is REML, but it is not exactly the same because of a difference in their estimating equations. ML criteria also appear to be better suited for comparing different models than are criteria based on REML because the approximate fixed effects design matrices depend upon the fixed effects parameter estimates and thus change from fit to fit. REML is attractive because of its bias adjustments, and also because it is based on an orthogonality between the fixed effects and variance-covariance parameters that is also present in the form of the estimating equations solved by the three methods. Of course, the differences between ML and REML are slight whenever there are only a few fixed effects parameters, which is often the case with many nonlinear mixed models. Together the three methods offer you a flexible set of nonlinear methods that can accommodate a wide range of correlation and heterogeneity scenarios.

We now describe the theoretical details behind the three methods. To fix notation and assumptions, write the j^{th} ($j = 1, \dots, t_i$) observation on the i^{th} ($i = 1, \dots, s$) subject as

$$Y_{ij} = f(\mathbf{x}_{ij}, \boldsymbol{\beta}, \mathbf{u}_i) + e_{ij}$$

where

- f is some nonlinear function
- \mathbf{x}_{ij} is a known vector of covariates
- $\boldsymbol{\beta}$ is a vector of unknown fixed effect parameters
- \mathbf{u}_i is a vector of unknown vector of random effects
- e_{ij} are unknown random errors

The \mathbf{u}_i are assumed to be independent vectors with zero mean and variance-covariance matrix \mathbf{G} . The $e_{ij}, j = 1, \dots, t_i$ form a multivariate random vector with zero mean and conditional (on \mathbf{u}_i) variance-covariance matrix

$$\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}_i)^{-1/2} \mathbf{R}_i(\boldsymbol{\theta}) \mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}_i)^{-1/2}$$

where

- \mathbf{x}_i is the vector of covariate values $x_{ij}, j = 1, \dots, t_i$
- \mathbf{W}_i is a $t_i \times t_i$ diagonal weight matrix that may depend on \mathbf{x}_i , as well as $\boldsymbol{\beta}$ and \mathbf{u}_i
- \mathbf{R}_i is a $t_i \times t_i$ variance-covariance matrix depending on a vector of unknown parameters $\boldsymbol{\theta}$

To begin distinguishing the three methods of Table 15.4, assume that for methods 1 and 2 $\mathbf{R}_i(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}_{t_i}$, where \mathbf{I}_{t_i} is the $t_i \times t_i$ identity matrix and that the \mathbf{u}_i establish correlations for these methods. In contrast, assume that in method 3 the \mathbf{u}_i are not present, and thus $\mathbf{R}_i(\boldsymbol{\theta})$ models all correlations for method 3.

All three methods are implemented by iterative calls to PROC MIXED on pseudo-data, and Table 15.6 provides the form of the pseudo-data for all three methods.

Table 15.6 Pseudo-data for the Three Nonlinear Mixed Model Methods

Method	Pseudo-data y_{ij}^*
1	$y_{ij} - f(x_{ij}, \hat{\boldsymbol{\beta}}, \mathbf{0}) + \tilde{\mathbf{x}}_{ij}(\hat{\boldsymbol{\beta}}, \mathbf{0})'\hat{\boldsymbol{\beta}}$
2	$y_{ij} - f(x_{ij}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i) + \tilde{\mathbf{x}}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i)'\hat{\boldsymbol{\beta}} + \tilde{\mathbf{z}}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i)'\hat{\mathbf{u}}_i$
3	$y_{ij} - f(x_{ij}, \hat{\boldsymbol{\beta}}) + \tilde{\mathbf{X}}_{ij}(\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\beta}}$

For method 1,

$$\tilde{\mathbf{x}}_{ij}(\hat{\boldsymbol{\beta}}, \mathbf{0}) = \frac{\partial f(x_{ij}, \boldsymbol{\beta}, \mathbf{u}_i)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}, \mathbf{0})$$

where $\hat{\boldsymbol{\beta}}$ is the current estimate of the fixed effects parameters. Method 1 is therefore carried out by a Taylor series expansion of f about $\mathbf{u}_i = 0$.

For method 2,

$$\tilde{\mathbf{x}}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i) = \frac{\partial f(x_{ij}, \boldsymbol{\beta}, \mathbf{u}_i)}{\partial \boldsymbol{\beta}'}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$$

$$\tilde{\mathbf{z}}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i) = \frac{\partial f(x_{ij}, \boldsymbol{\beta}, \mathbf{u}_i)}{\partial \mathbf{u}_i}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$$

and the expansion is now about $\mathbf{u}_i = \hat{\mathbf{u}}_i$ the current EBLUP.

For method 3, as noted above, the dependence on \mathbf{u}_i is dropped, and the pseudo-data are a mechanism for taking Gauss-Newton optimization steps. The notation $\tilde{\mathbf{x}}_{ij}$ reflects the fact that these vectors form the rows of the fixed effects design matrices for each call to PROC MIXED in all three methods.

The specification of the linear mixed model fitted at each PROC MIXED step is completed with a covariance structure. Define $\boldsymbol{\theta}$ to be the vector of unknown variance-covariance parameters, which are those in $[\mathbf{G}, \sigma^2]$ for methods 1 and 2 and those in $\mathbf{R}_i(\boldsymbol{\theta})$ for method 3. Table 15.7 displays the form of this structure for each of the three methods. For methods 1 and 2, $\tilde{\mathbf{Z}}_i$ is the matrix with rows $\tilde{\mathbf{z}}_{ij}$, and this matrix is passed to the RANDOM statement in PROC MIXED.

For method 3, the covariance structure in $\mathbf{R}_i(\boldsymbol{\theta})$ is passed to the TYPE= option of the REPEATED statement in PROC MIXED. For all three methods, the elements of the diagonal matrix \mathbf{W}_i are used as a WEIGHT variable in PROC MIXED. Note that for method 2 the weights in \mathbf{W}_i can depend on the current estimates of the random effects, but not for method 1.

Table 15.7 Covariance Structure for the Three Nonlinear Mixed Model Methods

Method	$\mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\theta})$
1	$\tilde{\mathbf{Z}}_i(\hat{\boldsymbol{\beta}}, \mathbf{0})\mathbf{G}\tilde{\mathbf{Z}}_i(\hat{\boldsymbol{\beta}}, \mathbf{0})' + \sigma^2\mathbf{W}_i^{-1}(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0})$
2	$\tilde{\mathbf{Z}}_i(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i)\mathbf{G}\tilde{\mathbf{Z}}_i(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i)' + \sigma^2\mathbf{W}_i^{-1}(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}_i)$
3	$\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})^{-1/2}\mathbf{R}_i(\boldsymbol{\theta})\mathbf{W}_i(\mathbf{x}_i, \boldsymbol{\beta})^{-1/2}$

The final estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ for all three methods solve a set of generalized estimating equations (GEE) that can be written as follows:

$$\sum_{i=1}^s \tilde{\mathbf{X}}_i' \mathbf{V}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^{-1} (\mathbf{y}_i^* - \tilde{\mathbf{X}}_i \hat{\boldsymbol{\beta}}) = \mathbf{0}$$

$$\sum_{i=1}^s \frac{\partial l(\mathbf{y}_i^*, \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \mathbf{0}$$

Here, $\tilde{\mathbf{X}}_i$ is the matrix with rows $\tilde{\mathbf{x}}_{ij}$ and l is the standard linear mixed model likelihood or restricted likelihood.

The estimates from method 1 are similar to those from the first-order method of Beal and Sheiner (1982) and Sheiner and Beal (1985). The methods are generally different, however, because Beal and Sheiner maximize the extended least squares objective function and because its estimating equations for β involve an extra derivative term whenever V_i depends on β . Method 1 is also closely related to the techniques of Hirst, Zerbe, Boyle, and Wilkening (1991) and Vonesh and Carter (1992), as well as to the marginal quasi-likelihood method of Breslow and Clayton (1993).

Method 2 produces the estimates of Lindstrom and Bates (1990), although it uses a different algorithm from the one they describe (refer to Wolfinger (1993b) for details). Method 2 is also very similar to the penalized quasi-likelihood approach for generalized linear mixed models (Breslow and Clayton, 1993), and in fact, %NLINMIX can be used to produce some of the same estimates as the %GLIMMIX macro.

When the random effects parameters enter the model linearly and their design matrix does not depend upon the fixed effects parameters, then methods 1 and 2 are equivalent and produce the maximum likelihood estimates of Gumpertz and Pantula (1992) or their restricted maximum likelihood counterparts.

Method 3 generalizes the multivariate nonlinear model of Gennings, Chinchilli, and Carter (1989) to consider parameterized covariance structures. Method 3 has close ties with the population-averaged GEE approach of Zeger, Liang, and Albert (1988).

All three of the methods simultaneously solve GEE for both mean and variance-covariance parameters similar to the GEE2 methodology of Prentice and Zhao (1991). The %NLINMIX macro iteratively creates pseudo-data and calls PROC MIXED until consecutive parameter estimates do not differ by some specified criterion, which by default is 1E-8.

15.10 Troubleshooting Nonlinear Mixed Model Fitting

15.10.1 General Considerations

Numerical fitting of nonlinear models is often much less stable than fitting of linear models, and so do not be discouraged if your attempts at fitting a nonlinear model fail at first. Nonlinear mixed models are even more challenging because the likelihood function is an integral over the random effects. All of the standard caveats for nonlinear model fitting apply (see, for example, Gallant 1987, Bates and Watts 1988, Davidian and Giltinan 1995, Vonesh and Chinchilli 1997, Seber and Wild 2003). Please also keep in mind that if your data are really noisy, or if your nonlinear model is not appropriate for the data, your chances for model fitting failures increase considerably. Ideally successful convergence will be indicated by small gradients for every parameter (preferably less than 1E-3), a positive definite Hessian matrix at the solution, and small overall values for convergence criteria like GCONV in PROC NLMIXED.

One common source of instability is parameters with widely varying scales. If the scaling of your parameters varies by more than a few orders of magnitude, the numerical stability of the optimization problem can be seriously reduced and result in computational difficulties. A simple remedy is to rescale each parameter so that its final estimated value has a magnitude near 1.

The parameterization itself can also be an issue. It often helps to reparameterize your model so that parameters are roughly independent (in terms of their second derivative matrix). Also, it

can be helpful to use boundary constraints to keep parameters away from values that produce singularities.

Another source of difficulty is starting values for the parameters. Starting values far from their optimal estimated values can produce unpredictable results. If your nonlinear model has interpretable parameters, you can sometimes obtain good guesses for the parameters simply by visually inspecting a plot of the data. Note that the %NLINMIX macro requires starting values only for the fixed effects, whereas PROC NLMIXED requires them for all parameters.

If your initial starting values do not work, try “sneaking up” on the problem by first fitting a model with no random effects in order to get an idea of where the fixed effects parameters lie. Then try a zero-expansion method (which tends to be more stable) to get an approximation to the covariance parameters. Then fit your final, full model, using output from the previous runs to determine good starting values. In combination with this, you may also want to simplify your model by dropping certain parameters in order to have some simpler models that you can then extend.

You should also be aware of the possibility of local minima in your likelihood function. Although most nonlinear mixed models—like the examples in this chapter—have a single, unique, global solution, it is certainly possible for certain combinations of models and data values to produce likelihood functions with more than one maximum. Most nonlinear optimization methods can be trapped by local minima and will provide no clear indication that you have in fact converged to a suboptimal solution. If you suspect that there may be multiple modes in your likelihood function, try a grid of starting values across the entire range of feasible parameter space and make sure the solution you are achieving is in fact a global one.

15.10.2 PROC NLMIXED Tips

Starting Values

Regarding starting values for PROC NLMIXED, it is important to realize that all unknown parameters not listed in the PARMs statement are assigned a starting value of 1. If 1 is not a reasonable value for any of these parameters, then the optimization method may fail to converge. Furthermore, if you do not specify the QPOINTS= option, then the number of quadrature points is determined adaptively at the first starting value. If this first starting value is far from the optimal value, the number of quadrature points necessary for accurate likelihood approximation may not be appropriate. To be more confident that you have an accurate likelihood approximation, you can rerun the model using starting values that are near the optimum value.

Parameterizing Variances and Covariance Matrices

In contrast to the MIXED procedure, optimization in the NLMIXED procedure involves both fixed effects parameters and covariance parameters. In the MIXED procedure, fixed effects are profiled and only covariance parameters take part in the optimization. Since parameters describing the mean and parameters describing variation and covariation are often on very different scales, the optimization can suffer unless you take care in scaling covariates and in choosing the model parameterization properly. As far as parameterizing the covariance structure is concerned, you can often improve numerical properties of the optimization by

parameterizing variances in terms of the log standard deviation. For example, the following statements fit a model involving two variances, the conditional variance of the response (S2E) and the variance of the random effects (S2U):

```
proc nlmixed;
  s2e = exp(2*logsig);
  s2u = exp(2*logsigu);
  mean = b0 + b1*x1 + b2*x2 + u;
  model y ~ normal (mean,s2e);
  random u ~ normal(0,s2u) subject=sub;
run;
```

Note, however, that S2E and S2U are determined through assignment statements in the NLMIXED code. The parameters determining the variances are LOGSIGE and LOGSIGU, the logarithms of the standard deviations, taking advantage of the simple relationship

$$\sigma^2 = \exp\{\log(\sigma^2)\} = \exp\{2\log(\sigma)\} = \exp\{2\phi\}$$

In this notation, the NLMIXED code above chooses ϕ as the parameter in the optimization. The “log-sigma” parameterization has a number of advantages: (1) it is guaranteed that σ^2 is positive, even without imposing any boundaries; (2) the parameter ϕ can range on the entire real line; you do not need boundary constraints for ϕ ; (3) taking the logarithm improves scaling of the parameters. If you want to obtain an estimate of the variance σ^2 and its standard error, you can add ESTIMATE statements to your NLMIXED code as needed. Here is an example:

```
estimate 'Var[u]' s2u;
estimate 'Var[error]' s2e;
```

A possible disadvantage of this parameterization is that the estimated value can never be exactly zero. In variance component models, for example, a zero estimate can be of importance for modeling and inference.

When your model contains more than one random effect, then you also need to worry about the conditioning and properties of their covariance matrix, Var[u]. In many cases, the covariance matrix will be unstructured, akin to TYPE=UN in PROC MIXED. For example, the following statements fit a nonlinear mixed model with two random effects.

```
proc nlmixed;
  parameters beta=0.049 gamma=-1.5
            s2u1=0.1 s2u2=0.1 cov12=0
            s2e=20;
  term = (beta+u1)*(rate**^(gamma+u2));
  mean = 100 * (term / (1 + term));
  model y ~ normal(mean,s2e);
  random u1 u2 ~ normal([0,0],[s2u1,cov12,s2u2]) subject=sub;
run;
```

The covariance matrix of $\mathbf{u} = [u_1, u_2]$ is parameterized in terms of two variances (S2U1, S2U2) and a covariance (COV12). The BOUNDS statement guarantees that the variances are nonnegative, but this does not guarantee that the resulting covariance matrix of \mathbf{u} is valid. Even if the off-diagonal element is positive, the resulting matrix may not be positive definite. For example, the matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

is not a covariance matrix (it has a positive and a negative eigenvalue). Since there are no provisions in the NLMIXED code above to prevent a situation where the estimated variance matrix of \mathbf{u} is not positive definite, the optimization can be greatly impeded. There are a number of ways to patch things up. You can, for example, parameterize the covariance matrix in terms of variances and correlations and place bounds on the correlation parameter (note the assignment to COV12 before the RANDOM statement):

```
proc nlmixed;
  parameters beta=0.049 gamma=-1.5
            s2u1=0.1 s2u2=0.1 rho=0
            s2e=20;
  bounds s2u1 > 0, s2u2 > 0, -1 < rho < 1;
  term = (beta+u1)*(rate**^(gamma+u2));
  mean = 100 * (term / (1 + term));
  model y ~ normal(mean,s2e);
  cov12 = sqrt(s2u1*s2u2)*rho;
  random u1 u2 ~ normal([0,0],[s2u1,cov12,s2u2]) subject=sub;
run;
```

This can be combined with the “log-sigma” parameterization mentioned earlier to stabilize the two variance estimates and to eliminate the need for their bounds:

```
proc nlmixed;
  parameters beta=0.049 gamma=-1.5
            logsig1=0 logsig2=0 rho=0 s2e=20;
  bounds -1 < rho < 1;
  term = (beta+u1)*(rate**^(gamma+u2));
  mean = 100 * (term / (1 + term));
  s2u1 = exp(2*logsig1);
  s2u2 = exp(2*logsig2);
  cov12 = sqrt(s2u1*s2u2)*rho;
  random u1 u2 ~ normal([0,0],[s2u1,cov12,s2u2]) subject=subject;
  model y ~ normal(mean,s2e);
run;
```

The recommended method to parameterize unstructured covariance matrices for more than one random effect is not in terms of correlations, however. This method becomes unwieldy quickly as the number of random effects increases. The recommended method uses a Cholesky parameterization of $\text{Var}[\mathbf{u}]$. This has the following advantages: (1) it is easy to code; (2) it is easy to extend to higher-dimensional cases; (3) the estimated covariance matrix is guaranteed to be at least positive semi-definite; (4) the parameters are scaled, and this often improves convergence behavior. The Cholesky parameterization you can achieve in PROC NLMIXED is the same as the TYPE=FA0(q) covariance structure in PROC MIXED (or GLIMMIX), where q equals the number of random effects, or the TYPE=CHOL covariance structure in PROC GLIMMIX. Consider the 2×2 case. Let \mathbf{T} be an upper triangular matrix such that

$$\mathbf{T}'\mathbf{T} = \begin{bmatrix} t_{11} & 0 \\ t_{12} & t_{22} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{11}t_{12} \\ t_{11}t_{12} & t_{12}^2 + t_{22}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Instead of the parameters σ_1^2 , σ_2^2 , and σ_{12} , the covariance matrix is parameterized in terms of t_{11} , t_{12} , and t_{22} .

```
proc nlmixed;
  parameters beta=0.049 gamma=-1.5
            t11=0.3 t22=0 t12=0 s2e=20;
  term  = (beta+u1)*(rate**(gamma+u2));
  mean  = 100 * (term / (1 + term));
  s2u1  = t11*t11;
  cov12 = t11*t12;
  s2u2  = t12*t12 + t22*t22;
  model pctdrywt ~ normal(mean,s2e);
  random u1 u2 ~ normal([0,0],[s2u1,cov12,s2u2]) subject=pop;
run;
```

Other Tips to Reduce Run Time and Improve Convergence

PROC NLMIXED can take a long time to run for problems with complex models, many parameters, or large input data sets. Although the optimization techniques used by PROC NLMIXED are some of the best ones available, they are not guaranteed to converge quickly for all problems. Ill-posed or misspecified models can cause the algorithms to use more extensive calculations designed to achieve convergence, and this can result in longer run times. So, first make sure that your model is specified correctly, that your parameters are scaled to be of the same order of magnitude, and that your data reasonably match the model you are contemplating. Here are a few more basic things to try when your model is not working:

- Specify the ITDETAILS option in the PROC NLMIXED statement to obtain more detailed information about when and where the problem is occurring.
- Add PUT statements to your code to output critical variables and verify they are being computed as you intend.
- Place checks and bounds in your code to prevent overflows and underflows of functions like exp() and log().
- For normal models, try METHOD=FIRO, which tends to be fairly stable. Alternatively, specify QPOINTS=1 to obtain Laplace-approximation estimates.
- Evaluate your modeling code in a SAS DATA step with the parameters set to fixed values in order to make sure the calculations are as you intend.
- Try running your model through the %NLINMIX macro, since its linearization-based iterative methods can sometimes be more stable than a full-scale numerical optimization problem.

For cases where you may suspect that there are local maxima in your likelihood function, try specifying a grid of parameter values in the PARMs statement and saving them to an output data set by adding a statement like the following:

```
ods output parameters=p;
```

Then plot your likelihood surface over parameters likely related to the multi-modalities.

For more details on troubleshooting PROC NLMIXED runs, refer to the documentation for PROC NLMIXED in the *SAS/STAT User's Guide*.

15.10.3 %NLINMIX Macro Tips

Although %NLINMIX attempts to recover from errors when possible, you may encounter cases where it exits abnormally. Because of its great flexibility, the SAS macro language is not always clear about precisely what causes execution errors. The following is a brief list of possible causes:

- There is a syntax error in the %NLINMIX model specification in the form of missing or extra commas or semicolons.
- The model specification is creating an error condition when computing derived variables.
- There are unusual patterns of missing values in the data set.
- The input data set is not sorted by subject.

If %NLINMIX fails to converge, double-check your MODEL= and DERIVS= (if you have any) specifications. Models that do not reasonably explain a data set are often difficult to fit to that data. Data with outliers can also cause problems.

If you feel that your model should be converging but it is not, the following suggestions may help:

- Rescale the data and model so that all parameters are of the same order of magnitude. This can improve the stability of the algorithms.
- If the convergence criterion appears to be descending nicely, you may only need to increase the maximum number of iterations using the MAXIT= option.
- Try different starting values, possibly those from EXPAND=ZERO if you are using EXPAND=EBLUP.
- Skip the initial PROC NLIN step by using OPTIONS=SKIPNLIN.
- If you are not specifying your own derivatives, use the TOL= option, which may help in computing them more accurately.
- If the PROC MIXED call itself is not converging, try the RIDGE= option. This option allows you to change the minimum initial ridge value applied to Hessian matrices, which are not positive definite. The default is 0.3145, and lowering it to 1E-3 or 1E-6 often works well.
- If you are using EXPAND=EBLUP, the GAUSS=, MAXSUBIT=, FRACTION=, and SUBCONV= options request %NLINMIX to take extra Gauss-Newton steps within each iteration. Using these can help you achieve and speed convergence. They are documented as a part of the header portion of the macro.

15.11 Summary

This chapter considered mixed models for which the fixed and random effects enter nonlinearly. Standard likelihood approaches are more difficult to implement in this situation than with the linear mixed models considered in previous chapters, and we discussed how to

address the difficulties using either PROC NLMIXED or the NLINMIX macro. Section 15.2 provided background on PROC NLMIXED. Sections 15.3 through 15.7 discussed examples for the following kinds of nonlinear mixed models: logistic growth curve, nested nonlinear random effects, zero-inflated and hurdle Poisson, joint survival and longitudinal, and one-compartment pharmacokinetic. Section 15.8 compared PROC NLMIXED and the NLINMIX macro, and Section 15.9 described analytical methods in the NLINMIX macro. Section 15.10 provided tips for troubleshooting nonlinear mixed model fitting.



Case Studies

16.1 Introduction	638
16.2 Response Surface Experiment in a Split-Plot Design.....	639
16.2.1 Introduction	639
16.2.2 Analysis Using PROC MIXED	640
16.3 Response Surface Experiment with Repeated Measures	643
16.3.1 Introduction	643
16.3.2 Analysis Using Compound Symmetry.....	644
16.3.3 Analysis Using Correlated Error Model.....	647
16.4 A Split-Plot Experiment with Correlated Whole Plots.....	650
16.4.1 Introduction	650
16.4.2 Analysis Using PROC GLIMMIX	652
16.4.3 Comparison with Standard Split-Plot ANOVA.....	656
16.5 A Complex Split Plot: Whole Plot Conducted as an Incomplete Latin Square	659
16.5.1 Introduction	659
16.5.2 Analysis Using PROC MIXED	661
16.5.3 Problems Encountered with Using PROC GLM	667
16.6 A Complex Strip-Split-Split-Plot Example	667
16.6.1 Introduction	667
16.6.2 Analysis Using PROC GLIMMIX	670
16.6.3 Standard Errors in Complex Designs.....	673
16.7 Unreplicated Split-Plot Design.....	674
16.7.1 Introduction	674
16.7.2 Analysis Using PROC MIXED	678

16.8 2³ Treatment Structure in a Split-Plot Design with the Three-Way Interaction as the Whole-Plot Comparison	684
16.8.1 Introduction.....	684
16.8.2 Analysis Using PROC MIXED and PROC GLIMMIX.....	686
16.8.3 Comparison with PROC GLM	690
16.9 2³ Treatment Structure in an Incomplete Block Design Structure with Balanced Confounding.....	694
16.9.1 Introduction.....	694
16.9.2 Analysis Using PROC GLIMMIX and PROC MIXED	696
16.10 Product Acceptability Study with Crossover and Repeated Measures	699
16.10.1 Introduction.....	699
16.10.2 Variable Definitions.....	702
16.10.3 Independent Error Model	702
16.10.4 Autoregressive Errors for the Time Interval Part of the Model	707
16.10.5 Autoregressive Errors for the Time Interval and Period Parts of the Model	711
16.10.6 Conclusions	716
16.11 Random Coefficients Modeling of an AIDS Trial	716
16.11.1 Introduction.....	716
16.11.2 Analysis Using PROC MIXED	717
16.12 Microarray Example	727
16.12.1 Introduction.....	727
16.12.2 Analysis with PROC GLIMMIX	729

16.1 Introduction

The case studies in this chapter provide examples of mixed model applications that go beyond the explanatory examples used to introduce the various topics. Mixed model applications have become so diverse that no textbook can cover the entire range of possibilities. However, each example has either a complex model structure or other features that draw on material from several chapters. From these case studies, it is hoped that the information will enable you to better understand and use the advanced options available in SAS mixed model procedures. For all of the case studies presented in this chapter, you could use either PROC MIXED or PROC GLIMMIX. Where there is no difference, the PROC MIXED analysis is shown. However, for some case studies, we illustrate with PROC GLIMMIX because it has features that make in-depth analysis more convenient.

Several examples use complex split-plot or repeated measures structures. Examples of this type are particularly instructive for many aspects of mixed model analysis. The examples in many of these case studies perhaps could have been used, such as in Chapters 4 and 5, except that the

models are so advanced that readers might be distracted from the chapters' main ideas. Readers familiar with the first edition will recall that most of the case studies included a comparison with PROC GLM. This was partly because PROC MIXED was new and partly because, at that time, PROC GLM was widely used for split-plot and repeated measures data. However, times have changed. Comparisons between PROC MIXED and PROC GLM in these case studies are similar to those in Chapter 4. The discussion in Chapter 4 should effectively lay to rest any illusions about PROC GLM being a suitable tool for *any* mixed model analysis. Readers who are new to the subject are referred to Chapter 4.

The case study in Section 16.2 involves a response surface design with a split-plot feature. The objective is to fit a response surface model to data from a corn milling experiment. Section 16.3 also involves a response surface design, but has a repeated measures as well as a split-plot feature. Section 16.4 involves a split-plot design with correlated whole plots. The data are from a line-source irrigation study where the levels of the whole-plot factor cannot be randomized and thus the whole-plot experimental units are correlated. Section 16.5 shows an incomplete Latin square used for the whole-plot treatments in a complex split-plot design. A strip-split-split-plot design used to evaluate a reclamation study in the Sand Hills region of Nebraska is described in Section 16.6. Section 16.7 shows how PROC MIXED can be used to analyze an unreplicated repeated measures experiment. The next two examples involve the use of 2^3 factorial treatment structures in the semiconductor industry. The case study in Section 16.8 involves a split-plot design where the three-way interaction is the whole-plot effect, and the case study in Section 16.9 involves incomplete blocks with balanced confounding. Section 16.10 consists of a case study involving product acceptability (in this case, taste testing) in a crossover design with repeated measures. The PROC MIXED code involves using more than one RANDOM statement and the REPEATED statement to fit the desired model. The analysis of multilocation experiments requires the use of mixed models. The example in Section 16.11 consists of applying a random coefficients model to data from a multicenter, randomized comparative trial of two compounds in HIV-infected patients. Finally, Section 16.12 shows a mixed model analysis of microarray data used in statistical genomics.

16.2 Response Surface Experiment in a Split-Plot Design

16.2.1 Introduction

Many industrial experiments use split-plot designs. As discussed in Chapter 4, split-plot designs are used when one or more of the factors in the experiment are “hard to vary” and the researcher restricts the randomization of the run order by grouping sets of runs together based on the hard-to-vary factor or factors. The experiment described in this section has one hard-to-vary factor.

This experiment involves milling corn. The goal is to model the effect of four treatment factors on the amount of grits that can be obtained from a one-minute run of a grinding mill. The four factors are moisture content of the corn (A), roll gap (B), screen size (C), and roller speed (D). To prepare corn for the experiment, a batch of corn (30 kg) has to be tempered to the desired moisture content. Thus it was decided to prepare a batch of corn to satisfy a specified moisture content, split the batch of corn into three parts (10 kg each), and then carry out three runs involving the other three factors. A response surface model is used to describe this milling process, specifically, a second-order polynomial response surface model, which is a regression model that consists of the linear terms, quadratic terms, and cross-product terms of the four factors. An optimal design selected from four factors at three levels with 30 runs was created using the OPTEX procedure in SAS/QC. The runs were grouped into sets of three, where each

set of three had the same level of moisture. The order of the three runs within a group was randomized and the order of the sets of three runs was randomized. The data along with the factor settings are displayed as Data Set 16.2, “Milling Corn,” in Appendix 2, “Data Sets.”

The three levels for each factor were equally spaced, so the coded values of -1, 0, and 1 are used in the analysis. There are four sets of runs at A = 1, four sets of runs at A = -1, and two sets of runs at A = 0. Because the three runs within a set were run together, the set of three runs or batch of corn becomes the experimental unit for the levels of A. The run or part of a batch of corn is the experimental unit for the levels of the other three factors.

The second-order response surface model used to describe these data is

$$Y_{ijk} = \beta_0 + \beta_1 a_i + \beta_2 b_{ijk} + \beta_3 c_{ijk} + \beta_4 d_{ijk} + \beta_5 a_i^2 + \beta_6 b_{ijk}^2 + \beta_7 c_{ijk}^2 + \beta_8 d_{ijk}^2 + \\ \beta_9 a_i b_{ijk} + \beta_{10} a_i c_{ijk} + \beta_{11} a_i d_{ijk} + \beta_{12} b_{ijk} c_{ijk} + \beta_{13} b_{ijk} d_{ijk} + \beta_{14} c_{ijk} d_{ijk} + s_{ij} + e_{ijk} \\ i=1,2,3, \quad j=1,\dots,n_i, \quad k=1,2,3$$

where

s_{ij} denote the batch error terms, assumed iid $N(0, \sigma_s^2)$

e_{ijk} denote the run error terms within a batch, assumed iid $N(0, \sigma_e^2)$

i is the index for the levels of A

j is the index for the batches within a level of A (j ranges from 1 to 4 for A = -1 and A=1 and from 1 to 2 for A=0)

k denotes the run within the j^{th} batch within the i^{th} level of A

We want to fit a regression model with the levels of A playing two roles.

First, we use the levels of A as quantitative in the regression model. Second, we use the levels of A as qualitative (a class variable) to extract the batch-to-batch error term. Because A plays two roles, we need to generate a second variable equivalent to A, in this case denoted by AA, so that we can let one play the quantitative role and one play the qualitative role. Thus, the term BATCH(AA) is used to extract the batch-to-batch variation from the data.

16.2.2 Analysis Using PROC MIXED

Program

The program used to read the data and generate the variable AA is as follows:

```
data design;
  input batch a b c d y;
  aa=a;
run;
```

Using these variables, here is the PROC MIXED program to fit the model above:

```
proc mixed data=design;
  class batch aa;
  model y = a|b|c|d@2 a*a b*b c*c d*d /solution;
  random batch(aa);
run;
```

The code A|B|C|D@2 generates the linear and cross-product terms in the model, and the effects A*A, B*B, C*C, and D*D add quadratic terms in the model. The SOLUTION option requests least-squares estimates of the coefficients in the response surface model, their estimated standard errors, *t*-values that test whether the parameter is zero, and the resulting significance level. The RANDOM BATCH(AA) enables PROC MIXED to compute the batch-to-batch variance component. By having the random effects in the RANDOM statement and the fixed effects in the MODEL, we can use A as a quantitative variable and a qualitative variable (AA) in the same model and extract all of the information about the parameters of the regression model and of the variance components.

Results

Selected results are shown in Output 16.1.

Output 16.1 PROC MIXED for the Response Surface Model with a Split-Plot Design without Approximate Degrees of Freedom

Class Level Information		
Class	Levels	Values
batch	10	1 2 3 4 5 6 7 8 9 10
aa	3	-1 0 1

Covariance Parameter Estimates	
Cov Parm	Estimate
batch(aa)	3.7853
Residual	5.2311

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	493.07	2.0847	7	236.52	<.0001
a	1.6627	0.8471	8	1.96	0.0853
b	1.5137	0.5356	8	2.83	0.0223
a*b	2.1685	0.5728	8	3.79	0.0053
c	2.7097	0.5665	8	4.78	0.0014
a*c	0.4458	0.6566	8	0.68	0.5163
b*c	0.3552	0.5963	8	0.60	0.5679
d	-0.06885	0.5000	8	-0.14	0.8939
a*d	-0.3813	0.5848	8	-0.65	0.5326
b*d	-1.6975	0.6473	8	-2.62	0.0305
c*d	2.6396	0.5924	8	4.46	0.0021
a*a	0.05378	1.9046	8	0.03	0.9782
b*b	2.5061	1.3276	8	1.89	0.0958

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
c*c	1.3726	1.1965	8	1.15	0.2845
d*d	0.5181	1.2642	8	0.41	0.6927

Interpretation

The estimated standard error of the estimated regression coefficient for A is larger than those for the other linear terms, and the estimated standard error of the regression coefficient for A \times A is larger than those for the other quadratic terms. The larger estimated standard errors occur because the levels of A are assigned to the batch experimental unit while the other factors are assigned to the run-within-a-batch experimental unit. The estimate of the batch-to-batch variance component is 3.785, and the estimate of the run-to-run variance component is 5.231.

The information about the parameters of the regression model comes from both between-batch comparisons and between-run-within-a-batch comparisons. PROC MIXED uses information from both parts of the model to compute combined estimates of the regression parameters.

Note that the denominator degrees of freedom for all effects in Output 16.1 correspond to the run-to-run effects. One would expect the A effects to use batch-to-batch degrees of freedom. However, the default containment method for denominator degrees of freedom depends on model information from the CLASS statement. It has a syntactic component that checks the effects in the MODEL statement against the effects in the RANDOM statement for containment. Since the effects in the MODEL statement consist of continuous variables, they are not contained (or are contained by) the BATCH(AA) classification effect. The containment method cannot provide the correct degrees of freedom in this case. You need to use the DDFM=SATTERTH option to compute appropriate approximate degrees of freedom for each of the estimated standard errors.

Program—Including DDFM=SATTERTH

The PROC MIXED program with the DDFM=SATTERTH option is as follows:

```
proc mixed data=design;
  class batch aa;
  model y=a|b|c|d@2 a*a b*b c*c d*d /solution ddfm=satterth;
  random batch(aa);
run;
```

Results

The results of using PROC MIXED with the DDFM=SATTERTH option are shown in Output 16.2.

Output 16.2 PROC MIXED for the Response Surface Model with a Split-Plot Design with Approximate Degrees of Freedom

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	493.07	2.0847	8.84	236.52	<.0001
a	1.6627	0.8471	4.57	1.96	0.1123
b	1.5137	0.5356	11.1	2.83	0.0163
a*b	2.1685	0.5728	10.3	3.79	0.0034
c	2.7097	0.5665	13.7	4.78	0.0003
a*c	0.4458	0.6566	14.5	0.68	0.5078
b*c	0.3552	0.5963	11.8	0.60	0.5626
d	-0.06885	0.5000	9.29	-0.14	0.8934
a*d	-0.3813	0.5848	10.3	-0.65	0.5287
b*d	-1.6975	0.6473	13.5	-2.62	0.0206
c*d	2.6396	0.5924	10.6	4.46	0.0011
a*a	0.05378	1.9046	4.68	0.03	0.9786
b*b	2.5061	1.3276	13.9	1.89	0.0801
c*c	1.3726	1.1965	10.6	1.15	0.2765
d*d	0.5181	1.2642	12.5	0.41	0.6888

Interpretation

The approximate degrees of freedom provide another indication as to which estimates are based on the batch-to-batch variance. The denominator degrees of freedom associated with the estimates of the standard errors of the coefficients for A and A × A are smaller than the degrees of freedom associated with the estimated standard errors of the other estimates, indicating the information is coming from between the batches instead of from within the batches.

16.3 Response Surface Experiment with Repeated Measures

16.3.1 Introduction

This example is from Bjerke et al. (2004). The data are from an experiment conducted to study the effects of different processing and storing conditions on quality of low-fat mayonnaise. There were three production variables, denoted A, B, and C. These were experimental factors involved in the production of the mayonnaise. The “recipe” involved varying amounts of A, B, and C. Factorial combinations of A, B, and C followed a central-composite response surface design as described in response surface design texts such as Khuri and Cornell (1996) or Myers

and Montgomery (2002). The different mayonnaise samples made from the production design were split and stored at two different temperatures (denoted D). Measurements were then taken at three different times, reflecting increasing length of storage. The data are given as Data Set 16.3, “Mayonnaise,” in Appendix 2, “Data Sets.”

The mixed model implied by the design of this experiment is

$$Y_{ijklm} = \mu_{ijklm} + s_{ijkn} + w_{ijkln} + e_{ijklmn}$$

where

μ_{ijklm} is the mean of the ijk^{th} A \times B \times C production variable combination at the l^{th} storage temperature (D) and at the m^{th} time of storage

s_{ijkn} is the effect of the n^{th} sample produced from the ijk^{th} A \times B \times C production variable combination

w_{ijkln} is the random error associated with the split of the n^{th} sample produced from the ijk^{th} A \times B \times C production variable combination stored at the l^{th} level of D

e_{ijklmn} is the random error associated with the m^{th} repeated measurement (time of storage) of the $ijkln^{\text{th}}$ half-sample

The random effects s_{ijkn} and w_{ijkln} are assumed normally and independently distributed with mean 0 and variance components σ_s^2 and σ_w^2 , respectively. The sets $\{e_{ijkl1n}, e_{ijkl2n}, e_{ijkl3n}\}$ of repeated measurement are assumed multivariate normal: repeated measures on the same half-sample are potentially correlated; half-samples are independent of one another.

The A \times B \times C \times D \times time mean, μ_{ijklm} , was partitioned into factor effects using the regression model

$$\begin{aligned} \mu_{ijklm} = & \beta_{0l} + \beta_{1AI}A_i + \beta_{1BI}B_j + \beta_{1CI}C_k + \beta_{2A}A_i^2 + \beta_{2B}B_j^2 + \beta_{2C}C_k^2 + \\ & \beta_{AB}A_iB_j + \beta_{AC}A_iC_k + \beta_{BC}B_jC_k + \\ & \beta_{1TI}T_m + \beta_{2T}T_m^2 + \beta_{AT}A_iT_m + \beta_{BT}B_jT_m + \beta_{CT}C_kT_m \end{aligned}$$

where β_{0l} is the intercept specific to the l^{th} storage level (of factor D); β_{1AI} , β_{1BI} , β_{1CI} , and β_{1TI} are linear regression coefficients over levels of A, B, C, and time, respectively, each specific to the l^{th} storage level; β_{2A} , β_{2B} , β_{2C} , and β_{2T} are the quadratic regression coefficients over levels of A, B, C, and time, respectively; and β_{AB} , β_{AC} , ..., β_{CT} are the linear-by-linear interaction regression coefficients for the respective factor combination levels.

16.3.2 Analysis Using Compound Symmetry

Program

You can use the following SAS statements to estimate the regression model that accounts for the various source of variation implied by the design. If you assume that the errors for the repeated measures are independent (and hence equivalent to a compound symmetry model), use these statements:

```

proc mixed data=mayo;
  class ca cb cc p d u;
  model y = d a(d) b(d) c(d) a*a b*b c*c a*b a*c b*c
    t(d) t*t t*a t*b t*c /noint solution htype=1 ddfm=kr;
  random p(ca cb cc) d*p(ca cb cc);
run;

```

Or equivalently, use these statements:

```

proc mixed data=mayo;
  class ca cb cc p d u;
  model y = d a(d) b(d) c(d) a*a b*b c*c a*b a*c b*c
    t(d) t*t t*a t*b t*c /noint solution htype=1 ddfm=kr;
  random p(ca cb cc);
  repeated / type=cs subject=d*p(ca cb cc);
run;

```

In the statements above, the variable P is a code for SAMPLE, U is a code for TIME OF STORAGE, T is a scaled value of TIME OF STORAGE used to fit the regression model, and CA, CB, and CC are values equal to A, B, and C, respectively, used as class variables to allow estimation of the random effects implied by the design, such as P(CA CB CC) to estimate the s_{ijkn} effects and of σ_s^2 . Following the dose-type example in Chapter 4, the NOINT and SOLUTION options allow you to print the estimated regression coefficients in immediately usable form.

Results

Output 16.3 shows relevant results.

Output 16.3 Compound Symmetry Analysis of Mayonnaise Quality Data

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
p(ca*cb*cc)		24.3200
CS	p*d(ca*cb*cc)	4.5151
Residual		11.4944

Fit Statistics	
-2 Res Log Likelihood	567.1
AIC (smaller is better)	573.1
AICC (smaller is better)	573.4
BIC (smaller is better)	575.8

Solution for Fixed Effects						
Effect	d	Estimate	Standard Error	DF	t Value	Pr > t
d	0	53.5687	2.3344	12.1	22.95	<.0001
d	1	31.7168	2.3344	12.1	13.59	<.0001
a(d)	0	16.8226	1.8101	10.5	9.29	<.0001
a(d)	1	11.2226	1.8101	10.5	6.20	<.0001

Solution for Fixed Effects						
Effect	d	Estimate	Standard Error	DF	t Value	Pr > t
b(d)	0	19.5049	1.8101	10.5	10.78	<.0001
b(d)	1	12.3715	1.8101	10.5	6.83	<.0001
c(d)	0	4.4019	1.8101	10.5	2.43	0.0344
c(d)	1	3.5352	1.8101	10.5	1.95	0.0781
a*a		0.4980	3.2427	8	0.15	0.8817
b*b		-2.5020	3.2427	8	-0.77	0.4626
c*c		5.1647	3.2427	8	1.59	0.1499
a*b		6.2083	1.8872	8	3.29	0.0110
a*c		-2.8333	1.8872	8	-1.50	0.1717
b*c		1.2083	1.8872	8	0.64	0.5399
t(d)	0	9.4200	0.5504	66	17.12	<.0001
t(d)	1	0.02442	0.5504	66	0.04	0.9647
t*t		-0.1487	1.1114	66	-0.13	0.8940
a*t		0.1160	0.5078	66	0.23	0.8200
b*t		1.7331	0.5078	66	3.41	0.0011
c*t		0.3513	0.5078	66	0.69	0.4915

Interpretation

The covariance parameter estimates denoted “P(CA CB CC),” “CS,” and “Residual” in the “Covariance Parameter Estimates” table give the estimates of σ_S^2 (24.32), σ_W^2 (4.5151), and σ^2 (error among e_{ijklmn} , estimate=1.4944), respectively. Recall that the independent error model gives identical results, except that the “CS” estimate is relabeled as “D*P(CA CB CC).”

The AICC for this model is 573.4. You can use this term to compare this model to correlated error models, an example of which is shown below. You can use the estimated regression coefficients to write the regression model and various LSMEANS and ESTIMATE statements to calculate predicted values of the response variable Y for various A \times B \times C \times D \times time combinations. Use the methods shown, for example, in Chapters 4 and 7. Output 16.4 shows the Type I tests of hypotheses for the various regression coefficients. You can use these results for possible model reduction. For example, the p-value associated with the B \times B effect tests $H_0: \beta_{2B} = 0$, i.e., the quadratic effect of factor B. The fact that $p = 0.8308$ suggests that this term does not significantly affect the response Y.

Output 16.4 Type I Tests Associated with Compound Symmetry Analysis of Mayonnaise Quality Data

Type I Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
d	2	12.9	770.48	<.0001
a(d)	2	12.9	41.65	<.0001
b(d)	2	12.9	55.12	<.0001
c(d)	2	12.9	2.75	0.1011
a*a	1	8	0.60	0.4606
b*b	1	8	0.05	0.8308
c*c	1	8	2.54	0.1499
a*b	1	8	10.82	0.0110
a*c	1	8	2.25	0.1717
b*c	1	8	0.41	0.5399
t(d)	2	66	154.27	<.0001
t*t	1	66	0.02	0.8940
a*t	1	66	0.05	0.8200
b*t	1	66	11.65	0.0011
c*t	1	66	0.48	0.4915

16.3.3 Analysis Using Correlated Error Model

Because the repeated measures on each half-sample imply the possibility of correlated errors, it is advisable to compare the fit of the compound symmetry model with a correlated error model. An AR(1) model would ordinarily be a natural alternative. In these data, however, the time periods were unequally spaced. To account for the unequal spacing, the SP(POW) model was used to model covariances. The SP(POW) model is essentially the same as the AR(1) model, except that the correlation between observations d units apart, ρ^d , reflects the actual distance rather than the AR(1)'s inflexible ρ for adjacent observations, ρ^2 for observations with one observation in between, and so forth. The AR(1) model expresses the decline in correlation as a function of separation of points in a numbered sequence. The SP(POW) model expresses the decline in correlations as a function of Euclidean distance.

Program

Use the following SAS statements to fit the model with correlated errors.

```
proc mixed data=mayo;
  class ca cb cc p d u;
  model y = d a(d) b(d) c(d) a*a b*b c*c a*b a*c b*c
         t(d) t*t t*a t*b t*c /noint solution htype=1 ddfm=kr;
  random p(ca cb cc);
  repeated / type=sp(pow) (t) subject=d*p(ca cb cc);
run;
```

The only difference between this and the compound symmetry model above is the TYPE=SP(POW) option. Note that distances between observations are figured on the basis of the coded T variable for storage time.

Results

Output 16.5 shows selected results.

Output 16.5 Correlated Error [SP(POW)] Analysis of Mayonnaise Quality Data

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
p(ca*cb*cc)		21.7209
SP(POW)	p*d(ca*cb*cc)	0.6463
Residual		22.5766

Fit Statistics	
-2 Res Log Likelihood	544.8
AIC (smaller is better)	550.8
AICC (smaller is better)	551.1
BIC (smaller is better)	553.5

Interpretation

Note that the AICC term is 551.1 versus 573.4 in the compound symmetry model. There is strong evidence that the repeated measurement errors are correlated and the SP(POW) model provides a better fit. The estimated correlation coefficient is $\hat{\rho} = 0.6463$. You can see that the estimates of the sample and error variance components are also affected.

Output 16.6 gives the revised estimates of the regression coefficients and Type I tests.

Output 16.6 Regression Estimates and Type I Tests Based on SP(POW) Analysis of Mayonnaise Quality Data

Solution for Fixed Effects						
Effect	d	Estimate	Standard Error	DF	t Value	Pr > t
d	0	53.6185	2.3086	10.9	23.23	<.0001
d	1	31.9996	2.3086	10.9	13.86	<.0001
a(d)	0	16.6958	1.9262	12.4	8.67	<.0001
a(d)	1	11.3771	1.9262	12.4	5.91	<.0001
b(d)	0	19.2674	1.9262	12.4	10.00	<.0001
b(d)	1	12.1931	1.9262	12.4	6.33	<.0001
c(d)	0	4.3007	1.9262	12.4	2.23	0.0447
c(d)	1	3.4667	1.9262	12.4	1.80	0.0963
a*a		0.1537	3.2946	8	0.05	0.9639
b*b		-2.5533	3.2946	8	-0.77	0.4606
c*c		5.2608	3.2946	8	1.60	0.1490
a*b		5.9371	1.9174	8	3.10	0.0147
a*c		-3.0256	1.9174	8	-1.58	0.1532
b*c		1.1635	1.9174	8	0.61	0.5608
t(d)	0	9.5458	0.6266	79.9	15.23	<.0001
t(d)	1	-0.1014	0.6266	79.9	-0.16	0.8719
t*t		-0.1487	0.6553	48.1	-0.23	0.8215
a*t		0.1011	0.5945	79.9	0.17	0.8653
b*t		1.9577	0.5945	79.9	3.29	0.0015
c*t		0.4429	0.5945	79.9	0.75	0.4584

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
d	2	13.3	650.07	<.0001
a(d)	2	13.1	36.01	<.0001
b(d)	2	13.1	45.67	<.0001
c(d)	2	13.1	2.43	0.1267
a*a	1	8	0.41	0.5409
b*b	1	8	0.05	0.8292
c*c	1	8	2.55	0.1490
a*b	1	8	9.59	0.0147
a*c	1	8	2.49	0.1532
b*c	1	8	0.37	0.5608
t(d)	2	79.9	114.23	<.0001

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
t*t	1	48.1	0.05	0.8215
a*t	1	79.9	0.03	0.8653
b*t	1	79.9	10.84	0.0015
c*t	1	79.9	0.56	0.4584

You can see that the point estimates of the regression estimates are only slightly affected, whereas the impact on the standard errors is more noticeable. Also, the p -values from the Type I tests vary somewhat. In this application, the conclusions are not affected, assuming standard α -levels (e.g., 0.05, 0.10, etc.). You can see that they can be affected in other instances.

The main point of this example is that it is possible to use split-plot and repeated measures methods discussed in Chapters 4 and 5 in conjunction with advanced, incomplete factorial designs such as those typically used in response surface estimation. You use the same ritual of determining the experimental unit with respect to each factor to determine the random effects needed in the model and note that observations made on the same unit over time must be treated as potentially correlated.

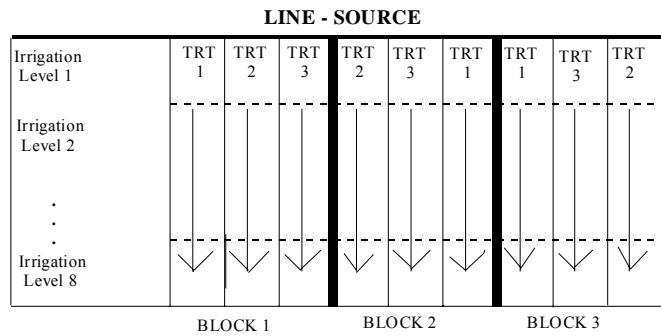
16.4 A Split-Plot Experiment with Correlated Whole Plots

16.4.1 Introduction

Experiments with correlated **errors** frequently occur. Common examples include repeated measures, discussed in Chapter 5, and spatially correlated experimental units, discussed in Chapter 11. In this example, a split-plot experiment was conducted in such a way that the **split-plot errors** were independent but the **whole-plot experimental units** were correlated. This section describes the design that led to this unusual situation, presents the PROC MIXED program required to analyze the data, and discusses the important differences between the analysis accounting for correlation and the standard, independent whole-plot analysis.

The purpose of this experiment was to compare the response to irrigation of three varieties of maize. A line-source irrigation system was used, similar to a system described in Johnson et al. (1983). In a line-source system, water emits from a linear source—essentially a long, straight pipe with regularly spaced nozzles. The amount of water reaching the ground decreases with increasing distance from the water source. In a standard line-source experiment, treatments are planted in long strips perpendicular to the line source. Within-a-strip measurements are taken at regularly spaced distances from the line source. Thus, distance corresponds to the amount of irrigation. The design is pictured in Figure 16.1.

Figure 16.1 Display of the Line-Source Arrangement of Treatments and Irrigation Levels



Because distance—and hence amount of irrigation—cannot be randomized, some correlation among the measurements at successive distances is possible. Thus, amount of irrigation behaves in the analysis like time in a repeated measures experiment. The analysis of this type of experiment using PROC MIXED is described in Example 46.6 of the *SAS/STAT User's Guide*.

The example considered here uses a different design. The strips perpendicular to the line source were divided into plots corresponding to increasing distance, i.e., amount of irrigation. Each plot was divided into three sub-plots, each of which was randomly assigned to a maize variety, as displayed in Figure 16.2.

Figure 16.2 Line-Source Layout Where Each Level of Irrigation Is Split into Three Sub-Plots within Each Block

LINE - SOURCE									
Irrigation Level 1	TRT 1	TRT 2	TRT 3	TRT 2	TRT 3	TRT 1	TRT 1	TRT 3	TRT 2
Irrigation Level 2	TRT 2	TRT 1	TRT 3	TRT 3	TRT 2	TRT 1	TRT 3	TRT 2	TRT 1
.
Irrigation Level 8	TRT 1	TRT 3	TRT 2	TRT 1	TRT 3	TRT 2	TRT 3	TRT 1	TRT 2
	Block 1			Block 2			Block 3		

There is a crucial difference between this design and the standard line-source experiment. In the *standard* experiment, the treatments are applied in whole plots and the irrigation amounts are divisions of the whole plot. In *this* example, the irrigation amounts are whole plots, and the treatments are applied to sub-plots within each whole plot. The whole plots cannot be randomized because they are defined in terms of their distance from the line source. Thus, the analysis of this experiment must account for the possibility of correlation among whole plots. The data for this experiment are given as Data Set 16.4, “Irrigation,” in Appendix 2, “Data Sets.”

A model for this experiment is

$$Y_{ijk} = \mu_{ij} + r_k + w_{ik} + e_{ijk}$$

where

Y_{ijk} is the observation on the i^{th} level of irrigation, j^{th} maize variety, and k^{th} block

μ_{ij} is the mean of the ij^{th} irrigation-by-variety combination

r_k is the k^{th} block effect

w_{ik} is the ik^{th} whole-plot effect

e_{ijk} is the ijk^{th} split-plot error

The treatment mean, μ_{ij} , can alternatively be expressed as

$$\mu_{ij} = \mu + \tau_i + \nu_j + (\tau\nu)_{ij}$$

where

μ is the intercept

τ_i is the i^{th} irrigation level effect

ν_j is the j^{th} variety effect

$(\tau\nu)_{ij}$ is the ij^{th} irrigation-by-variety interaction effect

The effects r_k , w_{ik} , and e_{ijk} are assumed to be random. The block effects and the split-plot errors are assumed to be normal and independent—specifically, $r_k \sim \text{iid } N(0, \sigma_B^2)$ and $e_{ijk} \sim \text{iid } N(0, \sigma_e^2)$. The whole-plot errors are assumed to be normal but correlated within a block. Specifically,

$$\text{Var}[w_{ik}] = \sigma_w^2$$

$$\text{Cov}[w_{ik}, w_{i'k}] = \sigma_{ii'}, \text{ where } i \neq i'$$

$$\text{Cov}[w_{ik}, w_{i'k'}] = 0, \text{ where } k \neq k'$$

The variable BLOC identifies the block, TRT identifies the maize variety, LEVEL identifies the irrigation level, and Y is the yield.

16.4.2 Analysis Using PROC GLIMMIX

Program

The model described above can be analyzed with either PROC MIXED or PROC GLIMMIX. The GLIMMIX procedure is used here because it makes the selective analysis of certain simple effects easier. The MIXED program is nearly identical, except that you must replace the LSMESTIMATE statement with ESTIMATE statements that are somewhat more cumbersome to define. See Section 4.4 for more details.

```

proc glimmix data=irrigation;
  class bloc trt level;
  model y = level trt trt*level / ddfm=kr;
  random intercept / subject=bloc;
  random level      / subject=bloc type=ar(1);
  lsmeans trt / diff;
  lsmeans level trt*level;
  estimate 'lev 1 vs lev 2' level 1 -1 0;
  estimate 'lev 1 vs lev 8' level 1  0 0  0  0  0  0  -1;
  lsmeans trt*level
    'trt 1 v 2 given lev 1'  1  0 0 0 0 0 0  0 -1 0,
    'lev 1 v 2 given trt 1'  1 -1 0
    'lev 1 v 8 given trt 1'  1  0 0 0 0 0 0 -1 0 ;
  contrast 'linear'   level -7 -5 -3 -1  1  3  5  7;
  contrast 'quadratic' level 7  1 -3 -5 -5 -3  1  7;
  contrast 'other'    level -7  5  7  3 -3 -7 -5  7,
    level 7 -13 -3  9  9 -3 -13 7,
    level -7 23 -17 -15 15 17 -23 7,
    level 1 -5  9 -5 -5  9 -5  1,
    level -1 7 -21 35 -35 21 -7  1;
run;

```

The CLASS and MODEL statements in this program are identical to those in standard split-plot analyses. The RANDOM statements for block and whole-plot error are handled differently. Because the block effects are independent but the whole-plot errors are correlated, separate RANDOM statements are needed for each effect. The RANDOM INTERCEPT / SUBJECT=BLOC statement defines the block effect. The second random statement defines the whole-plot error. The TYPE=AR(1) option defines the correlation—in this case a first-order autocorrelation structure—among the whole plots. The SUBJECT=BLOC option defines the correlation as being among whole plots *within* a block. We use the AR(1) model here to model the correlations. See the PROC GLIMMIX documentation for supported covariance structures and correlation models.

The ESTIMATE statements provide examples of main effect comparisons between levels that may be of interest. The LSMEANS statement provides examples of simple effects of possible interest. Other main effect or simple effect estimates can be defined. The method for constructing these statements was described in Chapter 4.

The LSMEANS TRT/DIFF statement causes PROC GLIMMIX to estimate all possible differences between varieties and their associated standard errors and *t*-test statistics. The CONTRAST statements given in this example reflect the researcher's interest in testing the fit of a hypothesized quadratic relationship between irrigation level and yield.

Selected results from this analysis are given in Output 16.7.

Results

Output 16.7 Analysis Using AR(1) Correlation among Whole-Plot Errors

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	bloc	5.8052	11.2249
Variance	bloc	12.3009	9.4385
AR(1)	bloc	0.7289	0.2174
Residual		1.2065	0.3449

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
level	7	16.43	1.46	0.2487
trt	2	24.53	29.58	<.0001
trt*level	14	25.72	1.56	0.1599

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
lev 1 vs lev 2	-1.8480	1.6892	19.56	-1.09	0.2872
lev 1 vs lev 8	-1.3120	2.8224	7.22	-0.46	0.6557

Least Squares Means Estimates						
Effect	Label	Estimate	Standard Error	DF	t Value	Pr > t
trt*level	trt 1 v 2 given lev 1	-2.285	0.981	25.29	-2.33	0.0282
trt*level	lev 1 v 2 given trt 1	-1.345	1.869	30.93	-0.72	0.4771
trt*level	lev 1 v 8 given trt 1	-0.986	2.934	8.84	-0.34	0.7447

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
linear	1	6.085	0.04	0.8477
quadratic	1	16.97	2.30	0.1477
other	5	17.27	1.56	0.2233

trt Least Squares Means					
trt	Estimate	Standard Error	DF	t Value	Pr > t
1	42.1612	1.5173	3.137	27.79	<.0001
2	43.2476	1.5173	3.137	28.50	<.0001
3	40.5901	1.5173	3.137	26.75	<.0001

level Least Squares Means					
level	Estimate	Standard Error	DF	t Value	Pr > t
1	39.6612	2.1598	6.496	18.36	<.0001
2	41.5092	2.1598	6.495	19.22	<.0001
3	44.1543	2.1598	6.495	20.44	<.0001
4	42.2788	2.1598	6.495	19.58	<.0001
5	41.9692	2.1598	6.495	19.43	<.0001
6	44.4925	2.1598	6.495	20.60	<.0001
7	40.9587	2.1598	6.495	18.96	<.0001
8	40.9732	2.1598	6.496	18.97	<.0001

Differences of trt Least Squares Means						
trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
1	2	-1.0864	0.3474	24.53	-3.13	0.0045
1	3	1.5710	0.3474	24.53	4.52	0.0001
2	3	2.6575	0.3474	24.53	7.65	<.0001

Interpretation

In the “Covariance Parameter Estimates” table the row labeled ‘intercept, bloc’ corresponds to the block variance, σ_B^2 . The estimated block variance is 5.805. PROC GLIMMIX by default also reports the asymptotic standard error of the variance component estimate, in this case 11.2249. These should be used with care. For example, interval estimates of variance component estimates based on asymptotic standard errors for small samples are notoriously inaccurate. The row labeled ‘variance bloc’ corresponds to σ_W^2 , and the row labeled ‘AR(1)’ corresponds to the autocorrelation parameter ρ . The estimates are 12.301 and 0.729, respectively. The ‘Residual’ row is for the split-plot error variance, σ_e^2 . Its estimate is 1.207.

In the “Type III Tests of Fixed Effects” table you find the F -tests for the main effects of irrigation (LEVEL) and variety (TRT), and the irrigation-by-variety (LEVEL \times TRT) interaction. These results show that there is no significant interaction between irrigation and variety ($F = 1.56, p = 0.1599$), nor is there a significant irrigation main effect ($F = 1.46, p = 0.2487$). There is a significant variety main effect ($F = 29.58, p < 0.0001$).

The “Estimates” and “Least Squares Mean Estimates” tables give results for specific main effects (ESTIMATE) and simple effects (LSMESTIMATE). The main thing to notice is that for the comparisons among LEVELS, which involve comparisons among correlated experimental units at varying distances apart, the standard error increases with distance. Thus, adjacent levels, such as LEVEL 1 and 2, are compared more precisely than levels farther apart, such as LEVEL 1 and 8. This is true of both LEVEL main effects and LEVEL simple effects holding TRT constant.

From the “Contrasts” table you see that there is very weak evidence of a quadratic relationship between LEVEL of irrigation and yield ($F = 2.30, p = 0.1477$) and no evidence of a more complex polynomial regression (“other,” $F=1.56, p = 0.2233$). Comparing these results to those obtained without the Kenward-Roger adjustment (results not shown here), the unadjusted F - and p -values are 3.66 ($p = 0.0695$) and $F = 2.24$ ($p = 0.0882$), respectively, for quadratic and “other.” As with correlated error models, the Kenward-Roger adjustment is needed for correlated whole-plot models to control the Type I error.

In the interest of space, only the LEVEL least-squares means are shown here. Looking at the pattern of the least-squares means across LEVEL, it is difficult to see any biologically intelligible relationship between irrigation level and yield. However, the pattern may suggest problems in correctly adjusting the equipment. That is, the “hot spots” at LEVEL 3 and 6 (whose means are 44.15 and 44.49, whereas other means range between 39.66 and 42.28) indicate misaligned nozzles in the line-source irrigator. This is a problem of interest to the irrigation engineers.

The “Differences of trt Least Squares Means” table shows that the estimated mean yield of variety (TRT) 1 is 1.086 units lower than variety (TRT) 2, with a standard error of 0.347. The difference between varieties 1 and 2 is statistically significant ($t = -3.13, p = 0.0045$). Also, TRT 1 and 2 both have significantly higher yields than TRT 3.

16.4.3 Comparison with Standard Split-Plot ANOVA

The standard split-plot analysis assumes that the whole-plot errors are independent. How does the analysis assuming correlated whole-plot errors compare to the standard split-plot analysis? You can obtain the usual analysis using the method described in Chapter 4.

Program

For these data, the PROC GLIMMIX statements are as follows:

```
proc glimmix data=irrigation;
  class bloc trt level;
  model y= level trt trt*level/ ddfm=kr;
  random bloc bloc*level;
  lsmeans trt/diff;
  lsmeans level trt*level;
  estimate 'lev 1 vs lev 2' level 1 -1 0;
  estimate 'lev 1 vs lev 8' level 1 0 0 0 0 0 -1;
  estimate 'trt 1 vs trt 2' trt 1 -1 0;
  lsmeans trt*level
    'trt 1 v 2 given lev 1' 1 0 0 0 0 0 0 -1 0,
    'lev 1 v 2 given trt 1' 1 -1 0,
    'lev 1 v 8 given trt 1' 1 0 0 0 0 0 -1 0;
```

```

contrast 'linear'    level -7   -5   -3   -1    1    3    5    7;
contrast 'quadratic' level  7    1    -3   -5   -5   -3   1    7;
contrast 'other'     level -7   5    7    3    -3   -7   -5   7,
                     level  7   -13  -3    9    9   -3   -13  7,
                     level -7   23   -17  -15  15   17   -23  7,
                     level  1   -5    9    -5   -5   9   -5   1,
                     level -1   7   -21   35  -35  21   -7   1;
run;

```

The LSMEANS, ESTIMATE, and CONTRAST statements are identical to those used in the previous section. The *only* difference is in the RANDOM statements.

The results of the analysis are given in Output 16.8.

Results

Output 16.8 Standard Split-Plot Analysis

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
bloc	8.5429	7.9124
bloc*level	8.3872	2.7829
Residual	1.2144	0.3499

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
level	7	20.73	1.18	0.3552
trt	2	24.17	29.39	<.0001
trt*level	14	24.75	1.48	0.1900

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
lev 1 vs lev 2	-1.8419	2.1146	20.73	-0.87	0.3937
lev 1 vs lev 8	-1.2766	2.1146	20.73	-0.60	0.5526
trt 1 vs trt 2	-1.0846	0.3486	24.17	-3.11	0.0047

Least Squares Means Estimates						
Effect	Label	Estimate	Standard Error	DF	t Value	Pr > t
trt*level	trt 1 v 2 given lev 1	-2.244	0.982	24.67	-2.29	0.0312
trt*level	lev 1 v 2 given trt 1	-1.455	2.261	26.55	-0.64	0.5254
trt*level	lev 1 v 8 given trt 1	-0.925	2.261	26.55	-0.41	0.6859

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
linear	1	20.73	0.12	0.7274
quadratic	1	20.73	4.31	0.0505
other	5	20.73	0.77	0.5839

Differences of trt Least Squares Means						
trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
1	2	-1.0846	0.3486	24.17	-3.11	0.0047
1	3	1.5730	0.3486	24.17	4.51	0.0001
2	3	2.6576	0.3486	24.17	7.62	<.0001

Interpretation

The following is a summary of the major differences between the two analyses in Output 16.7 and Output 16.8.

From the “Covariance Parameter Estimates” table we see that only the variance components, σ_B^2 , σ_W^2 , and σ_e^2 , are estimated. The estimate of the block variance is 8.543, compared with 5.805 in the previous analysis. For the whole-plot error, the estimate of σ_W^2 is 8.387, compared with 12.30 in Output 16.7. For the split-plot error, the estimate of σ_e^2 is 1.214, compared to 1.207. The AICC (not shown here) for the standard split-plot model is 249.05, compared to 241.71, indicating that the model with the autocorrelated errors provides the better fit.

Note that you could formally test the null hypothesis of no correlation among the whole plots, $H_0: \rho = 0$, using the “-2 Res Log Likelihood” statistics from the two models. For the correlated whole-plot model the statistic is 232.78; for the standard split plot it is 242.51. The resulting 1 d.f. χ^2 -statistic is 9.73. For any reasonable α -level, you would reject H_0 and conclude that significant correlation exists among the whole-plot experimental units.

The “Type III Tests of Fixed Effects” table indicates no significant interaction between irrigation and variety ($F = 1.48, p = 0.1900$) and no significant level main effect ($F = 1.18, p = 0.3552$). There is a highly significant variety effect ($F = 29.39, p < 0.0001$). These results are similar to the correlated whole-plot model assuming the Kenward-Roger adjustment is used. Because the main difference in the analyses is in the way whole-plot error is modeled, the greatest impact should be on the conclusions regarding the whole-plot treatment effect.

The main difference in the “Estimates” and “Least Squares Means Estimates” results is in the standard errors of the irrigation LEVEL main effects and simple effects holding variety constant. In the correlated whole-plot model, the standard error of a difference depends on the

distance between levels being compared. In the standard split-plot analysis, the standard errors are constant. Obviously, this can greatly affect the accuracy of interval estimation and hypothesis testing involving irrigation LEVEL.

The “Contrasts” results indicate a significant quadratic effect ($F = 4.31, p = 0.0505$) and no evidence of lack of fit (i.e., the contrast “other,” $F = 0.77$). This result is in marked contrast to the previous results, which showed no statistically significant evidence of any level effect.

If you use the results of the standard analysis, you may naively report a quadratic effect of irrigation. Biologically, this makes sense, so it may not occur to anyone to challenge the conclusion. However, the result is not consistent with the estimated least-squares means, nor is it consistent with the conclusions the irrigation engineers ultimately draw. That is, in the analysis accounting for whole-plot correlation, the engineers were alerted to improperly adjusted equipment.

The least-squares means in the two analyses are virtually identical, yet the CONTRAST results place them in an entirely different light. This is an example of the potential for rather serious misinterpretation that can result if the model assumptions do not closely reflect the actual design used to conduct the experiment.

The “Differences of trt Least Squares Means” table shows differences estimated between varieties. These are split-plot effects, which are not expected to be greatly affected by changes in the whole-plot component of the model. In fact, these results are virtually identical to the results obtained in the correlated whole-plot analysis.

16.5 A Complex Split Plot: Whole Plot Conducted as an Incomplete Latin Square

16.5.1 Introduction

This example presents the analysis of a split-plot experiment in which the whole plot has been set up as an incomplete Latin square. Such experiments are often desirable in view of the natural sources of variation among experimental units. However, using Latin squares and related designs as whole plots in split-plot experiments has been a traditional source of frustration for data analysts. The remainder of this section contains a description of the design and model. Section 16.5.2 contains a description of the main features of the analysis using PROC MIXED.

The experiment was conducted to evaluate the performance of two varieties of a crop at five levels of nitrogen fertilization. The five levels of nitrogen were mechanically applied to relatively large plots. Each plot was split into two sub-plots. Each genotype was randomly assigned to a sub-plot. Because of substantial north-south and east-west gradients, row-column, or Latin-square-like, blocking was used. However, as only 15 whole plots were available in a 5×3 grid, an incomplete Latin square was used. The layout of the experiment is given in Figure 16.3. The data are given as Data Set 16.5, “Nitrogen,” in Appendix 2, “Data Sets.”

Figure 16.3 Display of Randomization of Levels of Nitrogen to the Whole Plots and Levels of Genotype to the Sub-plots

	Col 1	Col 2	Col 3	Col 4	Col 5
Row 1	G1 Nit 1 ----- G2	G2 Nit 2 ----- G1	G2 Nit 5 ----- G1	G2 Nit 4 ----- G1	G1 Nit 3 ----- G2
Row 2	G1 Nit 2 ----- G2	G1 Nit 1 ----- G2	G1 Nit 3 ----- G2	G2 Nit 5 ----- G1	G1 Nit 4 ----- G2
Row 3	G2 Nit 3 ----- G1	G2 Nit 4 ----- G1	G1 Nit 1 ----- G2	G1 Nit 2 ----- G2	G2 Nit 5 ----- G1

A model for these data is

$$Y_{ijkl} = \mu_{ij} + c_k + r_l + w_{ikl} + e_{ijkl}$$

where

Y_{ijkl} is the observation on the i^{th} nitrogen (N) level, j^{th} genotype (G), k^{th} column (COL), and l^{th} row

μ_{ij} is the ij^{th} nitrogen-by-genotype mean

c_k is the k^{th} column effect, assumed iid $N(0, \sigma_c^2)$

r_l is the l^{th} row effect, assumed iid $N(0, \sigma_r^2)$

w_{ikl} is the ikl^{th} whole-plot error, assumed iid $N(0, \sigma_w^2)$

e_{ijkl} is the $ijkl^{\text{th}}$ split-plot error, assumed iid $N(0, \sigma_e^2)$

You can describe the nitrogen-by-genotype mean in terms of treatment effects as

$$\mu_{ij} = \mu + \eta_i + \gamma_j + (\eta\gamma)_{ij}$$

where

μ is the intercept

η_i is the i^{th} nitrogen level main effect

γ_j is the j^{th} genotype main effect

$(\eta\gamma)_{ij}$ is the ij^{th} nitrogen-by-genotype interaction

The researcher was interested in the following:

1. seeing if the nitrogen effect was similar for the two genotypes
2. characterizing the effect of nitrogen on response (Y)—separately for each genotype or averaged over genotypes, depending on the similarity of response
3. determining an optimum nitrogen rate for each genotype

16.5.2 Analysis Using PROC MIXED

Program

You can obtain the analysis for the model described above using the following PROC MIXED statements:

```
proc mixed data=nitrogen;
  class row col n g;
  model y = n g n*g / ddfm=satterth;
  random row col row*col*n;
  lsmeans n g n*g;
run;
```

The levels of nitrogen are denoted by N and the levels of genotype are denoted by G. The fixed effects part of the model consists of the two-way model for main effects for nitrogen and genotype and the interaction of nitrogen and genotype. The design structure consists of a rectangular plot of land split into rows and columns. The rows are a blocking factor and thus a random effect, and the columns are a blocking factor and thus a random effect. The whole-plot error is included by using the term ROW*COL*N in the RANDOM statement. The LSMEANS statement is included to show how the genotypes respond to the levels of nitrogen.

Results

The results of the estimates of the variance components and the tests for the fixed effects are shown in Output 16.9.

Output 16.9 Estimates of the Variance Components and Tests for the Fixed Effects

Covariance Parameter Estimates	
Cov Parm	Estimate
Row	13.2373
Col	0.8720
Row*Col*N	1.9586
Residual	0.3200

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
N	4	4.94	3.36	0.1093
G	1	10	234.03	<.0001
N*G	4	10	17.38	0.0002

Interpretation

From the “Covariance Parameter Estimates” table we see that the estimate of σ_R^2 is 13.2373; the estimate of σ_C^2 is 0.872; the estimate of σ_W^2 , the whole-plot experimental unit variance, is 1.9586; and the estimate of σ_e^2 , the split-plot error variance, is 0.32. The variance component of the row effects is much larger than any of the other variance components.

The “Type 3 Tests of Fixed Effects” table shows that the nitrogen level-by-genotype ($N \times G$) interaction is highly significant ($F = 17.38, p = 0.0002$) indicating the two genotypes did respond differently to the levels of nitrogen fertilizer. There is also a highly significant main effect of genotype ($F = 234.03, p < 0.0001$). We approach the interpretation of this effect with caution, pending developing a complete understanding of the $N \times G$ interaction.

Nitrogen is a quantitative factor with equally spaced levels. Therefore, it is reasonable to assess polynomial (linear, quadratic, etc.) regression. Orthogonal polynomial CONTRAST statements could be used to investigate the relationship of the response to the levels of the nitrogen fertilizer for each genotype. However, in the spirit of Section 4.6 and Chapter 7, it is easier to use direct regression in conjunction with analysis of covariance techniques.

Program—Evaluation of Regression over N Levels by Genotype

In order to use direct regression, a copy of the N variable is made so we have access to the nitrogen information as a continuous variable and a classification variable. The statement

```
NClass=N;
```

in the DATA step allows you to define regression over N levels and to use the class variable NClass to define the whole-plot error with the following program:

```
proc mixed data=nitrogen;
  class row col NClass G;
  model y = N|N|N|N|G / htype=1 ddfm=satterth;
  /*model y = N|N|G@2 NClass NCLASS*G /htype=1 ddfm=satterth;*/
  random row col row*col*NClass;
run;
```

Note that N is not included in the CLASS statement. When it appears in the MODEL statement, it is treated as a direct (regression) variable. The variable NClass does appear in the CLASS statement and is used in the RANDOM statement to define whole-plot error. The MODEL statement uses the vertical bar (|) notation to define N regression effects up to fourth order, the G main effect, and all interactions of G with the N regression effects. The second MODEL statement, commented out, shows how you can limit the regression—in the case to a second-order polynomial—and use the class variable to test lack of fit. Use the HTYPE=1 option to test elements of the model sequentially. The default HTYPE=3 option yields partial tests, which adjust lower-order terms for higher-order terms, and are not appropriate here.

The results of the first MODEL statement are shown in Output 16.10. The results of the alternative MODEL statement appear in Output 16.11.

Results

Output 16.10 Results Testing Regression over N Level by Genotype

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
N	1	5.71	8.10	0.0310
N*N	1	4.38	4.34	0.0997
N*N*N	1	4.36	0.28	0.6208
N*N*N*N	1	5.69	0.61	0.4666

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
G	1	10	234.03	<.0001
N*G	1	10	67.87	<.0001
N*N*G	1	10	0.05	0.8362
N*N*N*G	1	10	0.92	0.3604
N*N*N*N*G	1	10	0.69	0.4270

Interpretation

The N level by genotype interaction is confined to the $N \times G$ effect ($F = 67.87, p < 0.0001$). Because N refers here to linear regression over N levels, this result means that linear regression over levels has a different slope for each genotype. The significant G term ($F = 234.03$) indicates that the intercepts for the two genotypes are different. The second-order and higher N level effects ($N \times N$ – quadratic, $N \times N \times N$ – cubic, and $N \times N \times N \times N$ – quartic) all appear to be nonsignificant or at best marginally significant.

These results appear in somewhat different form in Output 16.11, using the alternative MODEL statement given above.

Output 16.11 Alternative Model to Test Regression over N Level by Genotype

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
N	1	5.71	8.10	0.0310
N*N	1	4.38	4.34	0.0997
G	1	10	234.03	<.0001
N*G	1	10	67.87	<.0001
nclass	2	4.92	0.45	0.6640
nclass*G	3	10	0.55	0.6596

Interpretation

Here only the second-order polynomial regression effects are shown explicitly. The cubic and quartic N main effects are pooled into a 2-degrees-of-freedom main effects term, NClass, which you interpret as main effect lack of fit. The interaction between genotype and quadratic, cubic, and quartic N effects, respectively, are pooled into the term NClass \times G. This term can be interpreted as a lack-of-fit term for the interaction between genotype and N level, or alternatively as a test for interaction between nonlinear N effects and genotype. From the results, $F = 0.55, p = 0.6596$, there is no evidence of such effects. The only significant effects are $N \times G$ and G, whose F -statistics and p -values and interpretation are identical to those given for Output 16.10.

Whether you use Output 16.10 or 16.11, the data indicate that there are separate linear regressions over N level for each genotype. The next step is to compute the **estimated** regression equation for each genotype. You can also plot the genotype-by-nitrogen means to provide a visual description of the regression.

PROC GLIMMIX Program Using ODS Graphics to Plot N x G Means

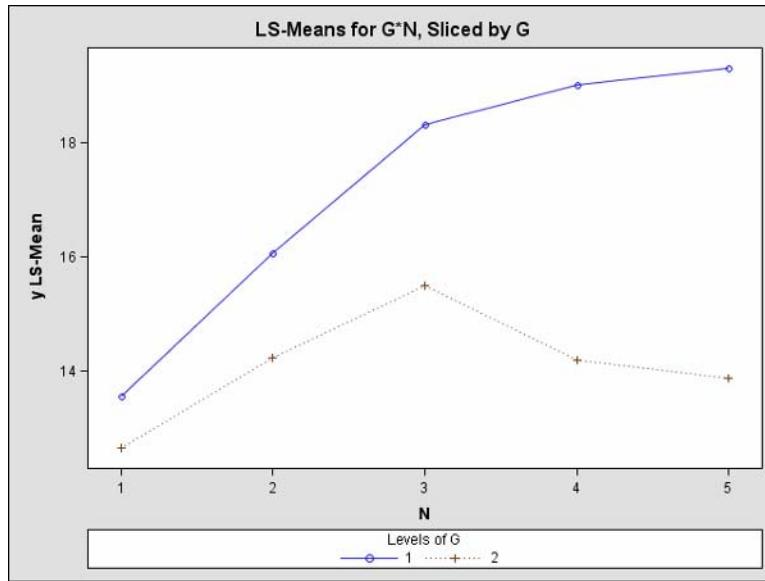
You can use the following program using PROC GLIMMIX and ODS graphics to obtain a plot of the genotype \times N level means. You can use PROC MIXED to obtain a similar plot (as shown in Section 4.6), but PROC GLIMMIX requires fewer steps.

```
ods html;
ods graphics on;
ods select MeanPlot;
proc glimmix data=nitrogen;
  class row col G N;
  model y = G|N;
  random row col row*col*nclass;
  lsmeans G*N/plot=MeanPlot (sliceby=G join);
run;
ods graphics off;
ods html close;
```

Figure 16.4 shows the plot.

Results

Figure 16.4 Plot of Means over N Levels by Genotype



Interpretation

You can see visual evidence that the response to N plateaus at higher levels of N for each genotype, even though there are no corresponding significant effects in the regression analysis in Output 16.10 or 16.11. One may need to consider some second-order or higher effect in order to obtain a realistic fit.

Program—Estimate Linear Regression for Each Genotype

The following SAS program allows you to estimate the linear regression models for each genotype. From Outputs 16.10 and 16.11 and Figure 16.4, a plausible regression model for the genotype \times N level means is

$$\mu_{ij} = \beta_{0i} + \beta_{1i}N_j + \beta_2N_j^2$$

where

- β_{0i} is the intercept for the i^{th} genotype
- β_{1i} is the slope (linear coefficient) for the i^{th} genotype
- N_j is the j^{th} N level
- β_2 is the quadratic coefficient

Notice that there are separate linear coefficients for each genotype but only the same quadratic coefficient for both genotypes. This is because the $G \times N \times N$ term in Output 16.10 was nonsignificant ($F = 0.54, p = 0.4808$). The quadratic main effect is included despite the fact that $N \times N$ was not significant ($F = 2.26, p = 0.2015$) because Figure 16.4 suggests that the curvature in the model should be considered.

The following SAS program computes the required estimates:

```
proc mixed data=nitrogen;
  class row col g nclass;
  model y = g n(g)n*n/ noint solution ddfm=satterth;
  random row col row*col*nclass;
run;
```

As with the programs that produce the results in Outputs 16.10 and 16.11, N is a direct regression variable and does not appear in the CLASS statement. The variable NCLASS does appear in the CLASS statement to allow you to estimate whole-plot error. Nesting N within G allows you to estimate the regression coefficients β_{1i} for each genotype. Note that $N \times N$ is not nested because the quadratic coefficient β_2 is assumed to be the same for both treatments. The NOINT option suppresses the default intercept term, so that G causes the intercept coefficients, β_{0i} , to be computed. Without the NOINT option, PROC MIXED would compute $\mu + \gamma$ instead. Relevant results appear in Output 16.12.

Results

Output 16.12 Estimates of Coefficients for Regression over N Level by Genotype

Solution for Fixed Effects						
Effect	G	Estimate	Standard Error	DF	t Value	Pr > t
G	1	9.6772	2.6901	4.68	3.60	0.0175
G	2	10.1272	2.6901	4.68	3.76	0.0148
N(G)	1	4.2540	1.1758	6.45	3.62	0.0098
N(G)	2	3.0507	1.1758	6.45	2.59	0.0384
N*N		-0.4714	0.1911	6.36	-2.47	0.0464

Interpretation

The ESTIMATE column gives the estimated regression coefficients. The intercept estimates, $\hat{\beta}_{0i}$, are 9.68 and 10.13, respectively, for genotypes 1 and 2. The estimated slopes, $\hat{\beta}_{1i}$, are 4.25 and 3.05, respectively. The estimate of the quadratic coefficient is $\hat{\beta}_2 = -0.47$. The standard errors of the estimates are given in the next column. The t -values and $\text{Pr} > |t|$ give statistics to

test the null hypotheses regarding each regression coefficient's equality to zero. Note that all p -values are statistically significant at the 0.05 level. Once lack of fit is pooled with relevant error terms, the quadratic term is statistically significant.

Plot of Predicted Regression Line

You can plot the predicted regression over N levels for each genotype. Use the regression coefficients from Output 16.12 to compute predicted values, $\hat{\mu}_{ij}$, for each N level \times genotype combination. Then you can use the PROC GPLOT statement in the following program to generate Figure 16.5.

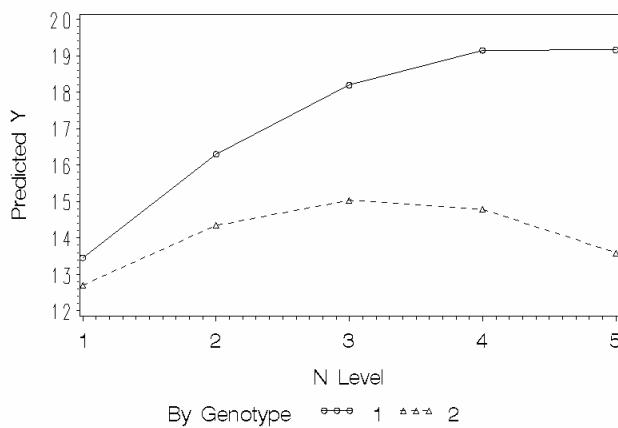
Program

```

axis1 value=(font=swiss2 h=2)
      label=(angle=90 f=swiss h=2 'Predicted Y');
axis2 value=(font=swiss h=2 )
      label=(f=swiss h=2 'N Level');
legend1 value=(font=swiss h=2 )
      label=(f=swiss h=2 'By Genotype');
symbol1 color=black interpol=join line=1 value=circle;
symbol2 color=black interpol=join line=20 value=triangle;
proc gplot data=now_plot;
  plot yhat*n=genotype / vaxis = axis1
                           haxis = axis2
                           legend = legend1;
run;

```

Figure 16.5 Plot of Predicted Mean Response over N Levels by Genotype based on Quadratic Model



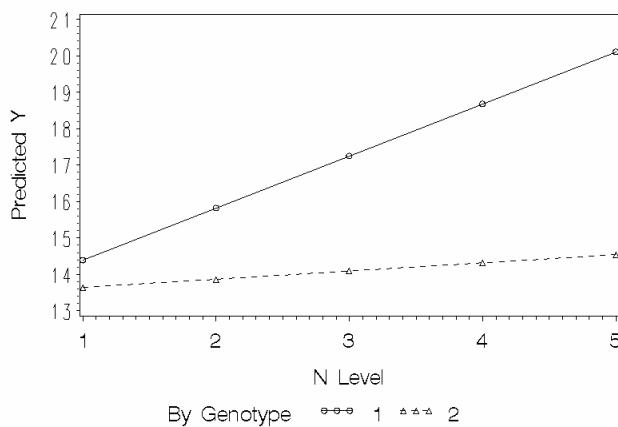
Interpretation

The predicted means in Figure 16.5 form a very similar pattern over N levels compared to the pattern of least-squares means shown in Figure 16.4. Given that the quadratic regression component was not significant in the analysis in Outputs 16.10 and 16.11, it is reasonable to ask how Figure 16.5 would compare to a regression without the quadratic component. You can delete the N*N term from the SAS program that generated the results in Output 16.12 and estimate the parameters of the model

$$\mu_{ij} = \beta_{0i} + \beta_{1i}N_j$$

The resulting regression produces the plot shown in Figure 16.6.

Figure 16.6 Plot of Predicted Mean Response over N Levels by Genotype based on Unequal Slopes Linear Model



You can see that the linear-only model does not provide a satisfactory explanation of the response to N level by genotype.

16.5.3 Problems Encountered with Using PROC GLM

Either PROC MIXED or PROC GLIMMIX can be used to analyze experiments with incomplete row-column and split-plot features such as the one shown here. For reasons fully discussed in Chapter 4, PROC GLM should not be used for these kinds of experiments. Users who attempt to do so will encounter spurious estimability problems, inaccurate or unavailable standard errors for most effects of interest, and, because of PROC GLM's inability to recover inter-block information, inefficient tests based on even those statistics that PROC GLM does compute correctly. As if these difficulties were not enough, you cannot obtain estimates of the regression coefficients without considerable heroics using the ESTIMATE statement, and even when you do, the standard errors will be wrong. Do not use PROC GLM to analyze experiments with split-plot features.

16.6 A Complex Strip-Split-Split-Plot Example

16.6.1 Introduction

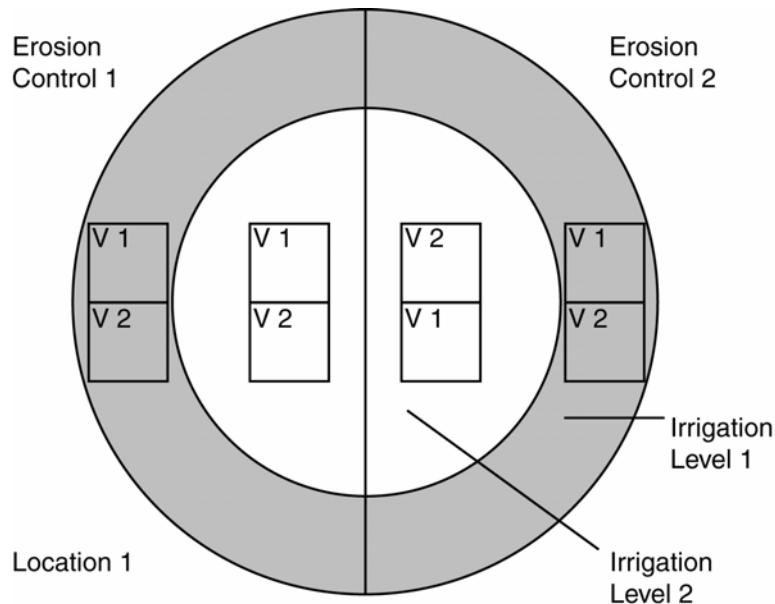
This example considers a complex extension of the split-plot principles introduced in Chapter 2. The data are based on an experiment comparing methods of rangeland reclamation in the Sandhills region of Nebraska.

Three treatment factors were studied: irrigation (IR) at two levels (1=low, 2=high); erosion control (EC), two methods; and seed mixes (V), two types. The objective of the study was to evaluate the effect of these three factors on the restoration of grassland areas that had been damaged by previous abuse.

Irrigation was applied using a center-pivot irrigator. This is a long, straight unit mounted on wheels that rotates around a central point. The center point acts both as an anchor for the unit and as the water source. The nozzles on the unit can be adjusted to various water flow rates, allowing level of irrigation to be varied in concentric circles around the center.

The circle defined by the center pivot was split into two halves. One side was managed using one erosion control method, and the other side was managed using the other method of erosion control. Within each irrigation level-by-erosion control sector, plots were assigned to the different seed mixes. Three center-pivot irrigation units were used in the experiment. Figure 16.7 shows the layout of the experiment for each location. The data are given in Data Set 16.6, “Sandhills,” in Appendix 2, “Data Sets.”

Figure 16.7 Layout of One Center-Pivot Irrigation Unit with the Levels of Irrigation, Erosion Control, and Seed Mixes



The design is an elaboration of the split-block layout shown in Figure 4.1.f. With respect to the irrigation and erosion control treatments, the design is identical to Figure 4.1.f; each center-pivot circle serves as a block. With respect to the seed mixes, the design is a split-split plot in the sense that each circle \times irrigation \times erosion control unit is split into units to which the seed mixes are applied. For lack of a better name, the overall design was called a “strip-split-split-plot” design.

Following the process described in Section 4.2.2, you can determine an appropriate mixed model by identifying the blocking criteria and the experimental unit with respect to each factorial treatment effect. These are shown in Table 16.1.

Table 16.1 Experimental Units with Respect to Factorial Effects for Range Reclamation Data

Effect	Experimental Unit with respect to
Block	Center-pivot circle
Erosion control (EC)	Circle \times EC
Irrigation level (IR)	Circle \times IR
IR \times EC	Circle \times EC \times IR
Seed mix (V)	Circle \times EC \times IR \times V
All interactions involving V and IR or EC	Circle \times EC \times IR \times V

Following Table 16.1, a mixed model for this experiment is

$$Y_{ijkl} = \mu_{ijk} + r_l + a_{il} + b_{jl} + c_{ijl} + e_{ijkl}$$

where

Y_{ijkl} is the observation on the l^{th} CIRCLE ($l = 1, \dots, L$), i^{th} EC level ($i=1, \dots, I$), j^{th} level of IR ($j = 1, \dots, J$), and k^{th} level of V ($k=1, \dots, K$)

μ_{ijk} is the mean of the ijk^{th} EC, IR, V treatment combination

r_l is the l^{th} CIRCLE effect, assumed $iid N(0, \sigma_r^2)$

a_{il} is the il^{th} CIRCLE \times EC effect, assumed $iid N(0, \sigma_a^2)$

b_{jl} is the jl^{th} CIRCLE \times IR effect, assumed $iid N(0, \sigma_b^2)$

c_{ijl} is the ijl^{th} CIRCLE \times ED \times IR effect, assumed $iid N(0, \sigma_c^2)$

e_{ijkl} is sub-plot error, assumed $iid N(0, \sigma_e^2)$

In this experiment, $I = J = K = 2$ and $L = 3$. You can extend the methods presented in this example to any number of treatment levels and blocks—we give the general range of subscripts to make extension easier. Also, you can express μ_{ijk} in terms of specific main effects and interactions—for example,

$$\mu_{ijk} = \mu + E_i + I_j + Ei_{ij} + V_k + VE_{ik} + VI_{jk} + VEI_{ijk}$$

where

μ is the intercept

E_i is the i^{th} EC main effect

I_j is the j^{th} IR main effect

Ei_{ij} is the ij^{th} EC*IR interaction

V_k is the k^{th} V main effect

VE_{ik} is the ik^{th} V \times EC interaction

VI_{jk} is the jk^{th} V \times IR interaction

VEI_{ijk} is the ijk^{th} V \times EC \times IR interaction

16.6.2 Analysis Using PROC GLIMMIX

Program

You can use either PROC MIXED or PROC GLIMMIX to obtain analysis for the model above. Because of its greater convenience in working with simple effects, PROC GLIMMIX is used here. The basic program to test model effects is as follows:

```
proc glimmix data=sandhills;
  class circle ec ir v;
  model y = ec ir ec*ir v*ec v*ir v*ec*ir/ddfm=satterth;
  random circle circle*ec circle*ir circle*ec*ir;
  ods select CovParms Tests3;
run;
```

The results appear in Output 16.13.

Results

Output 16.13 Covariance Parameter Estimates and Tests of Fixed Effects

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
circle	77.8981	83.1250
circle*ec	1.5500	6.0695
circle*ir	5.1254	9.2385
circle*ec*ir	5.0456	6.9541
Residual	3.6938	1.8469

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
ec	1	2	42.51	0.0227
ir	1	2	0.47	0.5650
ec*ir	1	2	0.16	0.7305
v	1	8	45.34	0.0001
ec*v	1	8	2.68	0.1406
ir*v	1	8	13.50	0.0063
ec*ir*v	1	8	0.20	0.6674

Interpretation

The “Covariance Parameter Estimates” show that none of the variance components for the various error terms are negative. If any were negative, you would need to use the methods discussed in Section 4.7 before proceeding. The “Type III Tests of Fixed Effects” table indicates that the IR \times V interaction and the EC and V main effect are statistically significant ($p = 0.0277$). The next step focuses on more detailed inference on the EC main effect and the IR \times V interaction. Inference on the V main effect must await examination of the IR \times V interaction.

The following LSMEANS options help explain the significant effects.

Program—LSMEANS

Insert the following LSMEANS statements after the RANDOM statement in the PROC GLIMMIX program shown above.

```
lsmeans ec      / diff          cl;
lsmeans ir*v   / slicediff=(ir v) cl;
lsmeans ec*ir*v / slicediff=ec*ir cl;
```

The LSMEANS EC / DIFF statement shows the difference between the two erosion control means. The CL option requests confidence limits for the difference. The LSMEANS statement for the IR \times V statement using the SLICEDIFF option for IR and V allows you to evaluate the simple effects of V given IR and vice versa. No V main effect statement is included because of the significant V \times IR interaction. Although the EC \times IR \times V interaction is not significant, EC \times IR \times V means sliced by EC \times IR are shown to illustrate how you obtain simple effects of a factor given two other factors, in this case V given EC \times IR. The SLICEDIFF option is available with PROC GLIMMIX but not PROC MIXED. It allows you to restrict estimates of differences between factorial means to desired simple effects only. PROC MIXED has a SLICE option that allows you to test simple effects (but not estimate them) and a DIFF option that indiscriminately computes all possible differences among factorial combination means.

The results of the LSMEANS statements for the significant effects (EC main effect and IR \times V simple effects) are given in Output 16.14. The EC \times IR \times V results appear in Output 16.15.

Results

Output 16.14 LSMEANS Results for EC Main Effect and IR \times V Simple Effect

ec Least Squares Means								
ec	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
1	27.9667	5.3372	2.12	5.24	0.0305	0.05	6.2074	49.7259
2	39.8667	5.3372	2.12	7.47	0.0148	0.05	18.1074	61.6259

Differences of ec Least Squares Means									
ec	_ec	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
1	2	-11.9000	1.8251	2	-6.52	0.0227	0.05	-19.7526	-4.0474

ir*v Least Squares Means									
ir	v	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
1	1	29.0167	5.4212	2.253	5.35	0.0255	0.05	8.0297	50.0037
1	2	37.1833	5.4212	2.253	6.86	0.0149	0.05	16.1963	58.1703
2	1	33.5333	5.4212	2.253	6.19	0.0187	0.05	12.5463	54.5203
2	2	35.9333	5.4212	2.253	6.63	0.0161	0.05	14.9463	56.9203

Simple Effect Comparisons of ir*v Least Squares Means By ir										
Simple Effect Level	v	_v	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
ir 1	1	2	-8.1667	1.1096	8	-7.36	<.0001	0.05	-10.7254	-5.6079
ir 2	1	2	-2.4000	1.1096	8	-2.16	0.0625	0.05	-4.9588	0.1588

Simple Effect Comparisons of ir*v Least Squares Means By v										
Simple Effect Level	ir	_ir	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
v 1	1	2	-4.5167	2.5160	2.447	-1.80	0.1909	0.05	-13.6512	4.6178
v 2	1	2	1.2500	2.5160	2.447	0.50	0.6604	0.05	-7.8845	10.3845

Interpretation

The least-squares means of EC levels 1 and 2 are 27.97 and 39.87, respectively. Their estimated mean difference is 11.90 units with a standard error of 1.83 and a 95% confidence interval of 4.05 to 19.75 units. Using the SLICEDIFF results, the significant IR \times V interaction results from a highly significant seed mix effect at irrigation level IR1 ($t = -8.17, p < 0.0001$) but only a marginally significant V effect at IR2 ($t = -2.40, p = 0.0625$). Seed mix V1 has a mean response 8.17 units lower than V2 at IR1 but only 2.40 units lower at IR2. The standard error for these simple effect differences is 1.11. Alternatively, you can characterize the V \times IR interaction in terms of simple effects of irrigation for a given seed mix. For seed mix V1, IR1 is 4.52 units lower than IR2, whereas for V2, IR1 has a mean response 1.25 units *higher* than IR2. Neither of these simple effects is statistically significant, however, as the standard error is 2.52. You can also use ODS graphics in conjunction with the MEANPLOT option to produce an interaction plot. Use GLIMMIX statements similar to those shown in Section 16.5.2 to produce Figure 16.4. In the interest of space, the plot is not shown here.

Results for EC x IR x V Means

Output 16.15 LSMEANS Results for EC x IR x V Simple Effect

ec*ir*v Least Squares Means										
ec	ir	v	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
1	1	1	23.8333	5.5771	2.52	4.27	0.0331	0.05	4.0097	43.6570
1	1	2	31.0667	5.5771	2.52	5.57	0.0179	0.05	11.2430	50.8903
1	2	1	28.1000	5.5771	2.52	5.04	0.0226	0.05	8.2763	47.9237
1	2	2	28.8667	5.5771	2.52	5.18	0.0212	0.05	9.0430	48.6903
2	1	1	34.2000	5.5771	2.52	6.13	0.0142	0.05	14.3763	54.0237
2	1	2	43.3000	5.5771	2.52	7.76	0.0080	0.05	23.4763	63.1237
2	2	1	38.9667	5.5771	2.52	6.99	0.0104	0.05	19.1430	58.7903
2	2	2	43.0000	5.5771	2.52	7.71	0.0082	0.05	23.1763	62.8237

Simple Effect Comparisons of ec*ir*v Least Squares Means By ec*ir										
Simple Effect Level	v	_v	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
ec*ir 1 1	1	2	-7.2333	1.5692	8	-4.61	0.0017	0.05	-10.8520	-3.6147
ec*ir 1 2	1	2	-0.7667	1.5692	8	-0.49	0.6383	0.05	-4.3853	2.8520
ec*ir 2 1	1	2	-9.1000	1.5692	8	-5.80	0.0004	0.05	-12.7187	-5.4813
ec*ir 2 2	1	2	-4.0333	1.5692	8	-2.57	0.0331	0.05	-7.6520	-0.4147

Interpretation

The first table shows the least-squares means for all 8 factorial treatment combinations. The second table shows the difference between seed mixes V1 and V2 for each EC \times IR combination. You can see that the seed mix differences are greater for the two EC \times IR combinations involving IR level 1 than those involving IR level 2. This is consistent with the IR \times V results in Output 16.14. The fact that the seed mix difference is not statistically significant for the EC₁ \times IR₂ combination but is statistically significant for all other EC \times IR combinations may be an artifact of multiple means testing, or it may be worth noting for further investigation.

16.6.3 Standard Errors in Complex Designs

From Outputs 16.14 and 16.15, you can see that the standard errors vary with the particular least-squares mean or least-squares mean difference. Moreover, there is no straightforward relationship between standard errors of factorial combination means and their corresponding differences. This is a serious issue for editorial policies of scientific journals. Many journals publish standard errors of treatment means. Often, only one standard error is published even when there is a split-plot design structure. This is a problem even with simple split plots, but the more complex the design, the more serious the problem. Standard errors of treatment differences for those main effects or simple effects under discussion must be included as an essential part of scientific reporting. Table 16.2 shows the variances of several treatment differences of potential interest in the design discussed in this case study.

Table 16.2 Variance of Estimated Difference for Various Main Effects and Simple Effects

Contrast	Comparison	Variance
EC main effect	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk}$	$(2/JKL)(JK\sigma_A^2 + K\sigma_C^2 + \sigma_e^2)$
IR main effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ij'k}$	$(2/IKL)(IK\sigma_B^2 + K\sigma_C^2 + \sigma_e^2)$
V main effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ijk'}$	$(2/IJL)\sigma_e^2$
IR EC _i simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ij'k}$	$(2/KL)[K(\sigma_B^2 + \sigma_C^2) + \sigma_e^2]$
EC IR _j simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk}$	$(2/KL)[K(\sigma_A^2 + \sigma_C^2) + \sigma_e^2]$
EC×IR means, different EC and IR	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk'}$	$(2/KL)[K(\sigma_A^2 + \sigma_B^2 + \sigma_C^2) + \sigma_e^2]$
V EC _i simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ijk'}$	$(2/JL)\sigma_e^2$
EC V _k simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk}$	$(2/JL)(J\sigma_A^2 + \sigma_C^2 + \sigma_e^2)$
V×EC means, different EC and V	$\bar{\mu}_{ijk} - \bar{\mu}_{ijk'}$	$(2/JL)(J\sigma_A^2 + \sigma_C^2 + \sigma_e^2)$
V EC _i ×IR _j simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ijk'}$	$(2/L)\sigma_e^2$
IR EC _i ×V _k simple effect	$\bar{\mu}_{ijk} - \bar{\mu}_{ij'k}$	$(2/L)(\sigma_A^2 + \sigma_C^2 + \sigma_e^2)$
Same EC, different levels of IR and V	$\bar{\mu}_{ijk} - \bar{\mu}_{ij'k}$	$(2/L)(\sigma_A^2 + \sigma_C^2 + \sigma_e^2)$
Same V, different EC and IR levels	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk}$	$(2/L)(\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_e^2)$
Different EC, IR, and V	$\bar{\mu}_{ijk} - \bar{\mu}_{i'jk'}$	$(2/L)(\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_e^2)$

Estimated standard errors are obtained by taking the square roots of the terms in Table 16.3 with variance component estimates in place of the variance parameters. You can see that most differences have unique variances. Statistical reports from complex designs must reflect the design complexity. Note also that PROC GLM *cannot* compute these standard errors correctly and *should not* be used to analyze such designs. You must use true mixed model software, such as PROC MIXED or PROC GLIMMIX.

16.7 Unreplicated Split-Plot Design

16.7.1 Introduction

This example is provided by Hain (2004). The data are from an experiment to evaluate the effect of detergent pH, washing temperature, and mode of drying on the wear of an experimental fabric. The treatment design was a $2 \times 2 \times 3$ factorial with two levels of pH (8 and 10), two washing temperatures (35°C and 55°C), and three drying cycles (air dry, delicate dry, and normal dry). The experiment was conducted as follows.

A large sheet of fabric was divided into halves, each half forming a block. Within each block, the fabric was cut into 12 pieces, one for each treatment combination. Each of the 12 pieces per

block was washed and dried a total of 50 times using the assigned pH \times wash temp \times dry cycle combination. Breaking-strength measurements were taken after 10, 20, 30, 40, and 50 wash-dry cycles. The data are given in Data Set 16.7, “Breaking Strength,” in Appendix 2, “Data Sets.”

Note that there is only one observation per pH \times water temperature \times dry cycle \times number of washes. This occurred because the observations on the two blocks were combined during the experiment and the ability to distinguish between the two blocks in the data was lost.

The mixed model implied by the design of this experiment is

$$Y_{ijklm} = \mu_{ijkl} + r_m + w_{ijkm} + e_{ijklm}$$

where

μ_{ijkl} is the mean of the ijk^{th} pH \times water temperature \times dry cycle ($i = 8, 10; j = 35, 55; k = \text{air, delicate, normal}$) at the l^{th} time of washing ($l = 10, 20, 30, 40, 50$)

r_m is the effect of the m^{th} block ($m = 1, 2$ in the design, but $m = 1$ only in the data)

w_{ijkm} is the $ijkm^{\text{th}}$ between-subjects (or whole-plot) error effect, assumed $iid N(0, \sigma_w^2)$

e_{ijklm} is the within-subjects (or split-plot) error effect, assumed $iid N(0, \sigma^2)$

This model can be viewed as a split plot with the pH \times temp \times dry combinations as whole plots and number of washes as the split-plot factor, or, as a repeated measures design with number of washes as the factor with repeated observations over time. In either case, Table 16.3 shows the form of the analysis of variance that would have been appropriate if the observations had been taken according to the design.

Table 16.3 Nominal Analysis of Variance for Fabric Breaking-Strength Data

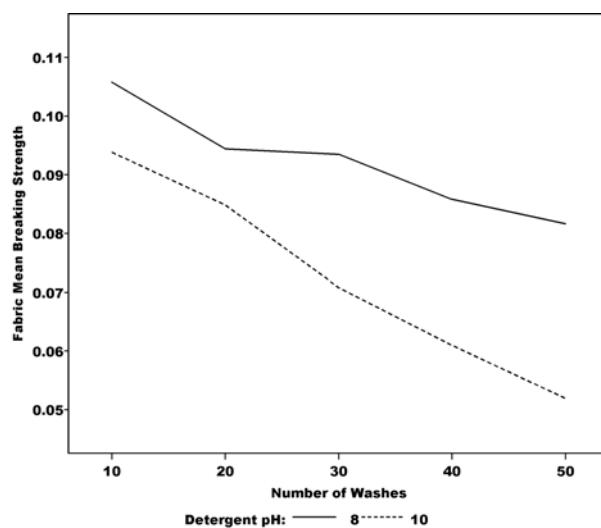
Source of Variation	d.f.
block	1
pH (P)	1
wash temp (T)	1
dry cycle (D)	2
P \times T	1
P \times D	2
T \times D	2
P \times T \times D	2
<i>between subject error</i>	11
no. of washes (W)	4
W \times P	4
W \times T	4
W \times D	8
W \times P \times T	4
W \times P \times D	8
W \times T \times D	8
W \times P \times T \times D	8
<i>within subject error</i>	48

You can see that if block distinctions had been maintained, the data could be analyzed using standard methods discussed in Chapters 4 and 5. However, the lack of a separate observation per block means that in the actual data observed there are 0 degrees of freedom for block, between-subject error, and within-subject error, *if one uses the standard full factorial model* shown above. Is the data analyst out of luck, or is there an alternative?

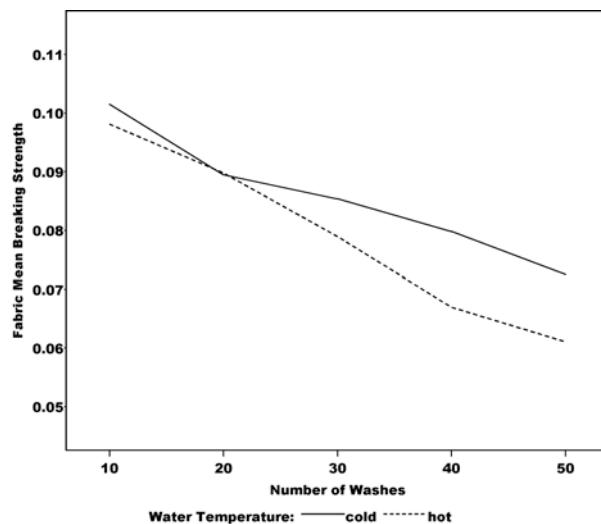
Figure 16.8 shows the plots of response means by pH, water temperature, and dry cycle, respectively. These plots hint at a possibly appropriate alternative to the full factorial model. Notice that breaking strength decreases over the number of washes. In every case the decrease is approximately linear. Only the slope, or rate of decrease, appears to be affected by treatment. Because the main objective of the study is to determine the effect of detergent pH, water temperature during washing, and drying cycle on fabric wear, you can view the slope, or rate of decrease in breaking strength, as the appropriate measure of wear.

Figure 16.8 Plot of Breaking Strength over Number of Washes by Treatment Levels

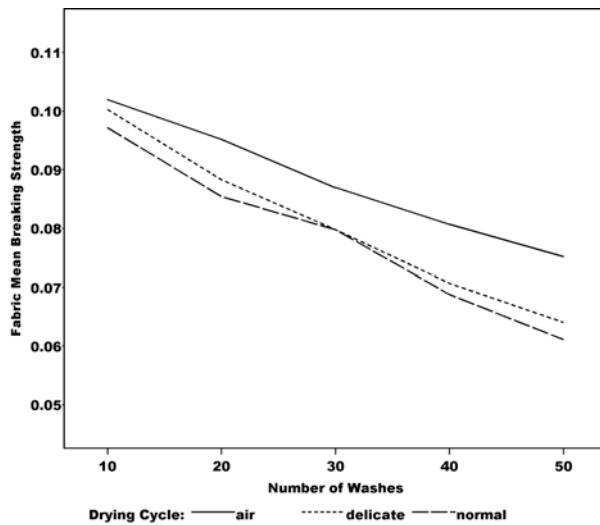
a. By pH



b. By Wash Temperature



c. By Drying Cycle



Half-normal plots were computed using SAS procedures described by Milliken and Johnson (1989). These confirmed that quadratic and higher-order polynomial effects over number of washed were negligible, as were all interactions among treatments. This leaves the reduced analysis of variance shown in Table 16.4.

Table 16.4 Reduced Analysis of Variance for Fabric Breaking-Strength Data

Source of Variation	d.f.
pH (P)	1
wash temp (T)	1
dry cycle (D)	2
<i>between subject error</i>	7
linear effect of no. of washes (W Lin)	1
W Lin × P	1
W Lin × T	1
W Lin × D	2
<i>within subject error</i>	43

The between-subject error term is estimated by pooling the 2- and 3-way interactions among pH, water temperature, and drying cycle. Supported by the half-normal plots, these effects can be assumed negligible. Hence, their expected mean squares are all effectively $\sigma^2 + 5\sigma_w^2$, meaning that they all effectively estimate between-subject error. Similarly, you can estimate the within-subject error by pooling all quadratic and higher-order polynomial effects of number of washes and their interactions with the other treatment factors. Because these effects are assumed negligible (again, supported by half-normal plots), they effectively estimate σ^2 .

The analysis of variance in Table 16.4 describes the mixed model

$$Y_{ijkl} = \beta_0 + \tau_i^P + \tau_j^T + \tau_k^D + (\beta + \delta_i^P + \delta_j^T + \delta_k^D)(nw)_l + w_{ijk} + e_{ijkl}$$

where

β_0 is the intercept

τ_i^P , τ_j^T , and τ_k^D are the i^{th} pH, j^{th} water temperature, and k^{th} drying cycle effects

β is the overall slope, or rate of decrease in breaking strength over number of washes

δ_i^P , δ_j^T , and δ_k^D are the changes in the slope resulting from the i^{th} pH, j^{th} water temperature, and k^{th} drying cycle

$(nw)_l$ denotes the number of washes

w_{ijk} and e_{ijkl} are the between-subject and within-subject error terms as previously defined

From the model, the regression of the number of washes on breaking strength for the ijk^{th} pH \times water temperature \times dry cycle combination is

$$\hat{Y}_{ijkl} = \hat{\beta}_0 + \hat{\tau}_i^P + \hat{\tau}_j^T + \hat{\tau}_k^D + (\hat{\beta} + \hat{\delta}_i^P + \hat{\delta}_j^T + \hat{\delta}_k^D)(nw)_l$$

The model parameters δ_i^P , δ_j^T , and δ_k^D measure the treatment effects on the rate of decrease in breaking strength. In other words, they measure treatment effects on fabric wear and are hence the terms of primary interest in the analysis.

16.7.2 Analysis Using PROC MIXED

Program

The following SAS statements yield the parameter estimates of the model above:

```
proc mixed data=breaking_strength;
  class pH water_temp dry_cycle;
  model breaking_strength=pH water_temp dry_cycle
    w w*pH w*water_temp w*dry_cycle/solution;
  random pH*water_temp*dry_cycle;
  contrast 'air vs dryer effect on wear'
    w*dry_cycle 2 -1 -1;
  contrast 'delicate v normal effect on wear'
    w*dry_cycle 0 1 -1;
run;
```

The pH, water_temp, and dry_cycle terms in the MODEL statement correspond to the parameters τ_i^P , τ_j^T , and τ_k^D , respectively. Similarly, the terms w*pH, w*water_temp, and w*dry_cycle correspond to δ_i^P , δ_j^T , and δ_k^D , respectively. The RANDOM statement acts as a catch-all for interactions among pH, water_temp, and dry_cycle, and thus provides an estimate of between-subjects error. The contrasts compare the δ_k^D , i.e., the rate of wear terms for the dry cycles. The first contrast compares wear rate of the air dry cycle to the average wear rate of the two machine dry cycles; the second contrast compares the wear rate for delicate machine dry cycle to that of the normal dry cycle.

Results

The results are shown in Output 16.16.

Output 16.16 PROC MIXED Analysis of Fabric Breaking-Strength Data

Covariance Parameter Estimates	
Cov Parm	Estimate
pH*water_te*dry_cycl	0.000020
Residual	0.000024

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
pH	1	7	1.32	0.2877
water_temp	1	7	0.26	0.6231
dry_cycle	2	7	0.26	0.7773
w	1	43	343.34	<.0001
w*pH	1	43	32.70	<.0001
w*water_temp	1	43	10.88	0.0020
w*dry_cycle	2	43	2.60	0.0862

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
air vs dryer effect on wear	1	43	5.18	0.0279
delicate v normal effect on wear	1	43	0.01	0.9039

Solution for Fixed Effects								
Effect	water_temp	dry_cycle	pH	Estimate	Standard Error	DF	t Value	Pr > t
Intercept				0.1038	0.004395	7	23.62	<.0001
pH			8	0.004522	0.003931	7	1.15	0.2877
pH			10	0
water_temp	cold			-0.00202	0.003931	7	-0.51	0.6231
water_temp	hot			0
dry_cycle		air		0.003319	0.004814	7	0.69	0.5127
dry_cycle		delicate		0.002563	0.004814	7	0.53	0.6109
dry_cycle		normal		0
w				-0.00129	0.000099	43	-12.97	<.0001
w*pH			8	0.000508	0.000089	43	5.72	<.0001
w*pH			10	0
w*water_temp	cold			0.000293	0.000089	43	3.30	0.0020

Solution for Fixed Effects								
Effect	water_temp	dry_cycle	pH	Estimate	Standard Error	DF	t Value	Pr > t
w*water_temp	hot			0
w*dry_cycle		air		0.000208	0.000109	43	1.91	0.0629
w*dry_cycle		delicate		-0.00001	0.000109	43	-0.12	0.9039
w*dry_cycle		normal		0

Interpretation

The first terms of real interest are the Type 3 tests of $W \times PH$, $W \times WATER_TEMP$, and $W \times DRY_CYCLE$. Their p -values are <0.0001 , 0.0020 , and 0.0862 , respectively, indicating statistically significant pH and water_temp effects on rate of wear. The $W \times DRY_CYCLE$ p -value appears to provide somewhat weaker evidence of a dry_cycle effect, but this has to be taken in perspective. It is a 2-degrees-of-freedom test. Partitioning it into its single-degree-of-freedom components using the CONTRAST statements shows that there is a significant difference in the wear rate using air drying compared to machine drying ($p = 0.0279$) but no statistically significant difference between the wear rates resulting from the delicate and normal dry settings ($p = 0.9039$). These results confirm the visual evidence in Figure 16.8.

Note that the “main effect” tests for pH, WATER_TEMP, and DRY_CYCLE are all nonsignificant: the p -values are all 0.2877 or greater. For these data, this is as it should be, since they test differences in the intercept of the regression, i.e., treatment differences after 0 washings have occurred. One would expect that treatment differences would begin to be observed only *after* several washes, not at the beginning of the experiment.

The “Solution for Fixed Effects” table gives the information needed to determine the regression equations for any given treatment combination. For example, for pH=8, washing at low water temperature, and using the air dry cycle, the regression equation is

$$\begin{aligned}\hat{Y}_{8,C,A} &= 0.1038 + 0.0045 - 0.0020 + 0.0033 + \\ &\quad (-0.0013 + 0.0005 + 0.0003 + 0.0002)(nw), \\ &= 0.1114 - 0.0003(nw),\end{aligned}$$

The parameter estimates are consistent with Figure 16.8. For example, the $W \times DRY_CYCLE$ term for air dry—i.e., $\hat{\delta}_A^D = 0.0002$ —indicates that the slope (rate of wear) for the air dry cycle is a little greater (i.e., less negative). Since more washes implies reduced breaking strength, the greater the negative slope coefficient, the greater the wear. The air dry cycle has the “least negative” slope among the dry cycles and hence the least rate of wear.

Program to Estimate Model Parameters

The SAS program shown above is best suited for testing the equality of the regression slopes among different treatment factors, i.e., the $W \times PH$, $W \times WATER_TEMP$, and $W \times DRY_CYCLE$ effects. The SOLUTION output from this program, shown in Output 16.1, is not convenient when you want to obtain the slopes for the various treatments. The following PROC MIXED program is better suited for estimating the regression parameters:

```

proc mixed data=breaking_strength;
  class pH water_temp dry_cycle;
  model breaking_strength = w(pH) w(water_temp)
    w(dry_cycle) / solution;
  random pH*water_temp*dry_cycle;
  estimate 'slope: ph 8, cold, air'
    w(ph) 1 0 w(water_temp) 1 0 w(dry_cycle) 1 0 0;
  estimate 'slope: ph 8, cold, delicate'
    w(ph) 1 0 w(water_temp) 1 0 w(dry_cycle) 0 1 0;
  estimate 'slope: ph 8, cold, normal'
    w(ph) 1 0 w(water_temp) 1 0 w(dry_cycle) 0 0 1;
  estimate 'slope: ph 8, hot, air'
    w(ph) 1 0 w(water_temp) 0 1 w(dry_cycle) 1 0 0;
  estimate 'slope: ph 8, hot, delicate'
    w(ph) 1 0 w(water_temp) 0 1 w(dry_cycle) 0 1 0;
  estimate 'slope: ph 8, hot, normal'
    w(ph) 1 0 w(water_temp) 0 1 w(dry_cycle) 0 0 1;
  estimate 'slope: ph 10, cold, air'
    w(ph) 0 1 w(water_temp) 1 0 w(dry_cycle) 1 0 0;
  estimate 'slope: ph 10, cold, delicate'
    w(ph) 0 1 w(water_temp) 1 0 w(dry_cycle) 0 1 0;
  estimate 'slope: ph 10, cold, normal'
    w(ph) 0 1 w(water_temp) 1 0 w(dry_cycle) 0 0 1;
  estimate 'slope: ph 10, hot, air'
    w(ph) 0 1 w(water_temp) 0 1 w(dry_cycle) 1 0 0;
  estimate 'slope: ph 10, hot, delicate'
    w(ph) 0 1 w(water_temp) 0 1 w(dry_cycle) 0 1 0;
  estimate 'slope: ph 10, hot, normal'
    w(ph) 0 1 w(water_temp) 0 1 w(dry_cycle) 0 0 1;
  estimate 'avg slope: ph 8'
    w(ph) 6 0 w(water_temp) 3 3 w(dry_cycle) 2 2 2
    / divisor=6;
  estimate 'avg slope: ph 10'
    w(ph) 0 6 w(water_temp) 3 3 w(dry_cycle) 2 2 2
    / divisor=6;
  estimate 'avg slope: cold water'
    w(ph) 3 3 w(water_temp) 6 0 w(dry_cycle) 2 2 2
    / divisor=6;
  estimate 'avg slope: hot water'
    w(ph) 3 3 w(water_temp) 0 6 w(dry_cycle) 2 2 2
    / divisor=6;
  estimate 'avg slope: air dry'
    w(ph) 1 1 w(water_temp) 1 1 w(dry_cycle) 2 0 0
    / divisor=2;
  estimate 'avg slope: delicate dry'
    w(ph) 1 1 w(water_temp) 1 1 w(dry_cycle) 0 2 0
    / divisor=2;
  estimate 'avg slope: normal dry'
    w(ph) 1 1 w(water_temp) 1 1 w(dry_cycle) 0 0 2
    / divisor=2;
run;

```

The MODEL statement assumes a common intercept for all treatment combinations, following the results of the tests of PH, WATER_TEMP, and DRY_CYCLE from the “Type 3 Tests of Fixed Effects” in Output 16.16. The model can be written as

$$Y_{ijkl} = \beta_0 + (\beta_i^P + \beta_j^T + \beta_k^D)(nw)_l + w_{ijk} + e_{ijkl}$$

where

β_0 is the intercept

β_i^P , β_i^T , and β_k^D are the slope components for the ijk^{th} pH, water temperature, and dry cycle combinations, respectively

w_{ijk} and e_{ijkl} are the between- and within-subject error terms as before

The W(PH), W(WATER_TEMP), and W(DRY_CYCLE) terms in the MODEL statement correspond to β_i^P , β_i^T , and β_k^D , respectively. The first 12 ESTIMATE statements compute the slopes for the entire slope coefficient, $\beta_i^P + \beta_i^T + \beta_k^D$, for the ijk^{th} pH \times water temperature \times dry cycle combination. The next seven ESTIMATE statements compute the slopes for each treatment level averaged over the levels of the other two factors. For example, the ESTIMATE statement labeled 'avg slope: ph 8' estimates the slope for pH=8 averaged over the levels of water temperature and dry cycle:

$$\beta_8^P + \frac{1}{2}(\beta_{Cold}^T + \beta_{Hot}^T) + \frac{1}{3}(\beta_{Air}^D + \beta_{Delicate}^D + \beta_{Normal}^D)$$

Results

Output 16.17 shows the relevant results.

Output 16.17 Regression (Rate of Wear) Estimates for Fabric Breaking-Strength Data

Solution for Fixed Effects								
Effect	water_temp	dry_cycle	pH	Estimate	Standard Error	DF	t Value	Pr > t
Intercept				0.1070	0.001895	11	56.46	<.0001
w(pH)			8	-0.00076	0.000077	43	-9.86	<.0001
w(pH)			10	-0.00134	0.000077	43	-17.41	<.0001
w(water_temp)	cold			0.000260	0.000063	43	4.12	0.0002
w(water_temp)	hot			0
w(dry_cycle)		air		0.000261	0.000077	43	3.38	0.0016
w(dry_cycle)		delicate		0.000028	0.000077	43	0.36	0.7175
w(dry_cycle)		normal		0

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
slope: ph 8, cold, air	-0.00024	0.000077	43	-3.08	0.0036
slope: ph 8, cold, delicate	-0.00047	0.000077	43	-6.11	<.0001
slope: ph 8, cold, normal	-0.00050	0.000077	43	-6.48	<.0001
slope: ph 8, hot, air	-0.00050	0.000077	43	-6.47	<.0001
slope: ph 8, hot, delicate	-0.00073	0.000077	43	-9.50	<.0001
slope: ph 8, hot, normal	-0.00076	0.000077	43	-9.86	<.0001

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
slope: ph 10, cold, air	-0.00082	0.000077	43	-10.63	<.0001
slope: ph 10, cold, delicate	-0.00105	0.000077	43	-13.66	<.0001
slope: ph 10, cold, normal	-0.00108	0.000077	43	-14.02	<.0001
slope: ph 10, hot, air	-0.00108	0.000077	43	-14.01	<.0001
slope: ph 10, hot, delicate	-0.00131	0.000077	43	-17.04	<.0001
slope: ph 10, hot, normal	-0.00134	0.000077	43	-17.41	<.0001
avg slope: ph 8	-0.00053	0.000054	43	-9.83	<.0001
avg slope: ph 10	-0.00111	0.000054	43	-20.56	<.0001
avg slope: cold water	-0.00069	0.000054	43	-12.79	<.0001
avg slope: hot water	-0.00095	0.000054	43	-17.60	<.0001
avg slope: air dry	-0.00066	0.000063	43	-10.50	<.0001
avg slope: delicate dry	-0.00089	0.000063	43	-14.22	<.0001
avg slope: normal dry	-0.00092	0.000063	43	-14.67	<.0001

Interpretation

From this output, you can readily see the regression coefficients for any treatment combination you select. For example, for the pH=8, cold water, air-dried treatment, the regression equation is $\hat{y} = 0.1070 - 0.00024 \times (nw)$. For the air-dried treatment, averaged over all detergent pH and water temperature combinations, the regression is $\hat{y} = 0.1070 - 0.00066 \times (nw)$. You can see that the slope is greater, and hence the rate of wear is less, for pH 8 (-0.00053 versus -0.00111), cold water (-0.00069 versus -0.00095), and air dried (-0.00066 compared with -0.00089 and -0.00092 for the delicate and normal machine dried treatments).

Final Comment

This example demonstrates that you can use principles borrowed from regression, response surface analysis, and half-normal plot analysis to “salvage” the analysis of an unreplicated split-plot or repeated measures experiment. However, we must conclude this example with a disclaimer that this is not a design we recommend. In this case, the researcher was lucky in the sense that the response to number of washes was strictly linear and there were no higher-order interactions. This allowed these terms to be “borrowed” in order to estimate between- and within-subject error and hence construct tests based on the appropriate error terms. If the higher-order polynomial regression terms and the higher-order interactions had not been negligible, the analysis shown here would not be possible.

This example is presented to show data analysts how to implement such an analysis as a last resort when the assumptions warrant and there is no alternative. However, if you are designing an experiment, remember that failure of the assumptions of negligible higher-order polynomial regression or negligible higher-order interaction would have been fatal to this analysis. It is best to plan for legitimate replication.

16.8 2^3 Treatment Structure in a Split-Plot Design with the Three-Way Interaction as the Whole-Plot Comparison

16.8.1 Introduction

When the number of treatments is large, an incomplete block design structure is the most efficient way to carry out the experiment. When an incomplete block design is used, some information about the treatment comparisons is confounded with blocks. If the design is carried out in sets of blocks with complete replication of the treatment combinations within each set of blocks, you can construct the blocks so that the same effect or effects are confounded with blocks within each replication. Such a design is a type of split plot: the confounded effects are the whole-plot comparisons and the other effects are the sub-plot comparisons. The experiment described in this section has a 2^3 treatment design. The experiment design consists of six sets or replications each containing two blocks with four experimental units per block. The treatment combinations are assigned to the blocks by confounding the three-way interaction with blocks within each replication. The assignment of the treatment combinations to the two blocks within each replication is displayed in Tables 16.5 and 16.6. The symbols for the treatment combinations in Table 16.5 use the notation of Cochran and Cox (1957), and the symbols for the treatment combinations in Table 16.6 use the notation of Box, Hunter, and Hunter (1978).

Table 16.5 Treatment Combinations Assigned to the Two Blocks within Each Replication to Confound the ABC Interaction Using the Notation of Cochran and Cox (1957)

BLOCK 1	(0) ¹	ab	ac	bc
BLOCK 2	a	b	c	abc

¹The presence of a letter indicates that factor at the high level and the absence of a letter indicates that factor at the low level.

Table 16.6 Treatment Combinations Assigned to the Two Blocks within Each Replication to Confound the ABC Interaction Using the Notation of Box, Hunter, and Hunter (1978)

BLOCK 1	- - - ¹	- + +	+ - +	+ + -
BLOCK 2	- - +	- + -	+ - -	+ + +

¹The order of the “+” and “-” signs corresponds to the levels of A, B, and C.

The data in Data Set 16.8, “ 2^3 Factorial,” in Appendix 2, “Data Sets,” are from an experiment studying the effect of three factors, pressure (A) at two levels, temperature (B) at two levels, and flow rate (C) at two levels, on the thickness (Y) of a silicon nitrate (Si_3N_4) layer deposited on silicon wafers used in the manufacture of computer chips. Only four batches of wafers could be processed during a shift, so blocks of size four were needed to carry out the experiment. Two blocks of size four were run on consecutive days, or one complete replication of the set of treatment combinations could be obtained in two days. The measurement is the average thickness of the layer of Si_3N_4 deposited on all of the wafers in a batch (25 wafers are in a batch).

The two blocks of four within each replication were determined using the arrangements in Table 16.5 or 16.6. Thus, the ABC interaction was confounded with blocks within each replication. Because there are six replications, information about the ABC interaction can be obtained from the data by averaging over the differences between the two blocks within each replication. The confounding of the ABC interaction with blocks within each replication means that the block or group of four batches is the experimental unit with respect to the components of the ABC interaction. The batch within a block is the experimental unit for all other factorial effects. Hence, two error terms are required in the model used to describe the data, one for the blocks within a replication and one for the batches within a block. Because the pairs of consecutive days correspond to a replication, the model must contain a random effect for these pairs of days. The randomization scheme was as follows:

1. Assign one of the two blocks to day 1.
2. Randomize the order of the runs within day 1.
3. During day 2, run the treatment combinations in the other block in random order.

This randomization process was then repeated for the other five pairs of consecutive days or replications. In the data set, REP corresponds to replication or pair of days and BLK corresponds to block or day within a replication. The data are listed in nonrandomized order.

This data set can be analyzed in two ways. First, you can use analysis of variance to analyze the means of the treatment combinations. Alternatively, the analysis can be carried out by using a regression model.

The analysis of variance model used to describe the means of the data is

$$Y_{ijkmn} = \mu + \alpha_k + \beta_m + \gamma_n + (\alpha\beta)_{km} + (\alpha\gamma)_{kn} + (\beta\gamma)_{mn} + (\alpha\beta\gamma)_{kmn} + r_i + a_{ij} + e_{ijkmn}$$

where

μ denotes the intercept

a_k , ($k = -1, 1$), denotes the k^{th} level of factor A

β_m , ($m = -1, 1$), denotes the m^{th} level of factor B

γ_n , ($n = -1, 1$), denotes the n^{th} level of factor C

$(\alpha\beta)_{km}, \dots, (\alpha\beta\gamma)_{kmn}$ denote the interactions among the factors

r_i , ($i = 1, 2, \dots, 6$), denotes the i^{th} replication, assumed iid $N(0, \sigma_r^2)$

a_{ij} , ($j = 1, 2$), denotes the j^{th} block of the i^{th} replication, assumed iid $N(0, \sigma_a^2)$

e_{ijkmn} denotes random error, assumed iid $N(0, \sigma_e^2)$

The regression model or response surface model is

$$\begin{aligned} Y_{ijk} = & \beta_0 + \beta_A A_k + \beta_B B_k + \beta_C C_k + \beta_{AB} A_k B_k + \beta_{AC} A_k C_k \\ & + \beta_{BC} B_k C_k + \beta_{ABC} A_k B_k C_k + r_i + a_{ij} + e_{ijkmn} \end{aligned}$$

where

β_0 is the intercept

A_k , B_k , and C_k denote values of factors A, B, and C, respectively

β_A , β_B , and β_C are linear regression coefficients for A, B, and C

$\beta_{AB}, \dots, \beta_{ABC}$ are cross-product interaction regression coefficients

r_i , a_{ij} , and e_{ijkmn} are defined as above

16.8.2 Analysis Using PROC MIXED and PROC GLIMMIX

This section shows the mixed model analysis of both models—the analysis of variance and the regression—using SAS software. You can use either PROC GLIMMIX or PROC MIXED for both models. PROC GLIMMIX makes working with the analysis of variance model somewhat easier because there are two-way interactions and the SLICEDIFF option makes it easier to evaluate the associated simple effects. There is little difference between GLIMMIX and MIXED for the regression model.

Program—Analysis of Variance Model

The following PROC GLIMMIX program implements the analysis of variance model:

```
proc glimmix data=fac_sp;
  class a b c rep blk;
  model y=a|b|c / ddfm=kr;
  random rep blk(rep);
  lsmeans a*b b*c / slicediff=b;
  lsmeans a*b*c;
run;
```

With A, B, and C in the CLASS statement, the MODEL term A|B|C generates all main effects, two-factor interactions, and the three-way interaction. The RANDOM statement contains REP for the replication effect and BLK(REP) for the block or group of four batches. Note that BLK(REP) is the error term for the three-way interaction. The DDFM=KR option is used to obtain approximate degrees of freedom for all fixed effect tests. If you use DDFM=SATTERTH, you obtain identical results for these data. The SLICEDIFF=B option obtains estimates of the simple effects of A given B and C given B to help interpret the $A \times B$ and $B \times C$ interactions, which are statistically significant for these data.

Results

The results of PROC GLIMMIX are shown in Output 16.18.

Output 16.18 Results of PROC GLIMMIX for Analyzing the Cell Means

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
rep	4.4875	4.7527
blk(rep)	3.9931	3.2681
Residual	4.6111	1.1906

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
a	1	30	26.10	<.0001
b	1	30	24.74	<.0001
a*b	1	30	18.51	0.0002
c	1	30	115.66	<.0001
a*c	1	30	1.46	0.2358
b*c	1	30	26.10	<.0001
a*b*c	1	5	0.10	0.7632

a*b Least Squares Means						
a	b	Estimate	Standard Error	DF	t Value	Pr > t
-1	-1	122.00	1.2103	7.672	100.80	<.0001
-1	1	127.75	1.2103	7.672	105.55	<.0001
1	-1	121.50	1.2103	7.672	100.38	<.0001
1	1	121.92	1.2103	7.672	100.73	<.0001

Simple Effect Comparisons of a*b Least Squares Means By b							
Simple Effect Level	a	_a	Estimate	Standard Error	DF	t Value	Pr > t
b -1	-1	1	0.5000	0.8767	30	0.57	0.5727
b 1	-1	1	5.8333	0.8767	30	6.65	<.0001

b*c Least Squares Means						
b	c	Estimate	Standard Error	DF	t Value	Pr > t
-1	-1	120.00	1.2103	7.672	99.15	<.0001
-1	1	123.50	1.2103	7.672	102.04	<.0001
1	-1	119.92	1.2103	7.672	99.08	<.0001
1	1	129.75	1.2103	7.672	107.20	<.0001

Simple Effect Comparisons of b*c Least Squares Means By b							
Simple Effect Level	c	_c	Estimate	Standard Error	DF	t Value	Pr > t
b -1	-1	1	-3.5000	0.8767	30	-3.99	0.0004
b 1	-1	1	-9.8333	0.8767	30	-11.22	<.0001

a*b*c Least Squares Means							
a	b	c	Estimate	Standard Error	DF	t Value	Pr > t
-1	-1	-1	119.67	1.4771	14.66	81.01	<.0001
-1	-1	1	124.33	1.4771	14.66	84.17	<.0001
-1	1	-1	122.67	1.4771	14.66	83.04	<.0001
-1	1	1	132.83	1.4771	14.66	89.93	<.0001
1	-1	-1	120.33	1.4771	14.66	81.46	<.0001
1	-1	1	122.67	1.4771	14.66	83.04	<.0001
1	1	-1	117.17	1.4771	14.66	79.32	<.0001
1	1	1	126.67	1.4771	14.66	85.75	<.0001

Interpretation

Because the ABC interaction is confounded with blocks in each replication, there are six estimates of the ABC effect, one from each replication. The variance of those six estimates—which is the estimate of the block-to-block variance or whole-plot variance—has five degrees of freedom. Hence, there are five denominator degrees of freedom for testing the ABC interaction. All other factorial effects have 30 denominator degrees of freedom, indicating they are sub-plot comparisons.

There is a significant $A \times B$ interaction ($p = 0.0002$) and a significant $B \times C$ interaction ($p < 0.0001$). You can use the results of LSMEANS statements computing the $A \times B$ and $B \times C$ means to interpret these interactions. The SLICEDIFF option provides estimates of the simple effects of A and C, respectively, given B. You can see that the difference between the A least-squares means at level $B = -1$ is negligible (122.0 versus 121.5 for $A = -1$ and $A = 1$, respectively), whereas the difference is significant at $B = 1$ ($p < 0.0001$). At $B = 1$, the mean of $A = -1$ is 127.75; the observed mean for $A = 1$ is 121.92. The standard error of the $A|B$ simple effect difference is 0.877. For the $C|B$ simple effects, the $C = 1$ mean is significantly greater than the $C = -1$ mean for both levels of B, but the magnitude of difference is greater given $B = 1$. At $B = -1$ the C difference is 3.5 units; given $B = 1$ it is 9.83 units.

The LSMEANS statement is used to compute the $A \times B \times C$ means or three-way cell means. These are shown for comparison with PROC GLM in Section 16.8.3.

Program—Regression Model

The PROC MIXED program used to fit the regression model with the random effects is as follows:

```
proc mixed data=fac_sp;
  class rep blk;
  model y=a|b|c / solution ddfm=satterth;
  random rep blk(rep);
run;
```

The MODEL statement term $A|B|C$ generates a regression model with the linear terms, two-way cross-product terms, and the three-way cross-product term. The program is essentially identical to the PROC GLIMMIX program shown above, except A, B, and C do not appear in the CLASS statement, SOLUTION is used in the MODEL statement, and no LSMEANS statements can be used. As with the previous program, with these data DDFM=KR and DDFM=SATTERTH yield identical results.

Results

The results of the analysis are shown in Output 16.19.

Output 16.19 Results from PROC MIXED Analyzing the Regression Model

Covariance Parameter Estimates	
Cov Parm	Estimate
rep	4.4875
blk(rep)	3.9931
Residual	4.6111

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	123.29	1.0848	5	113.66	<.0001
a	-1.5833	0.3099	30	-5.11	<.0001
b	1.5417	0.3099	30	4.97	<.0001
a*b	-1.3333	0.3099	30	-4.30	0.0002
c	3.3333	0.3099	30	10.75	<.0001
a*c	-0.3750	0.3099	30	-1.21	0.2358
b*c	1.5833	0.3099	30	5.11	<.0001
a*b*c	0.2083	0.6548	5	0.32	0.7632

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
a	1	30	26.10	<.0001
b	1	30	24.74	<.0001
a*b	1	30	18.51	0.0002
c	1	30	115.66	<.0001
a*c	1	30	1.46	0.2358
b*c	1	30	26.10	<.0001
a*b*c	1	5	0.10	0.7632

Interpretation

The estimates of the variance components are identical to those obtained in Output 16.18 because the regression model has one parameter for each cell of data in the treatment structure. If a simpler regression model were used, the estimates of the variance components would change. The SOLUTION output provides estimates of the regression coefficients and their standard errors. The standard error for $\hat{\beta}_{ABC}$ is 0.6548 and is based on 5 degrees of freedom, the degrees of freedom corresponding to the whole-plot error. As in the analysis of the means in

Output 16.18, there are significant coefficients for $A \times B$ and $B \times C$ as well as significant linear coefficients.

16.8.3 Comparison with PROC GLM

As with other designs with split-plot features discussed in Chapter 4 and previous sections of this chapter, the authors strongly discourage analysis with PROC GLM. In previous sections of this chapter, the issues are similar to those illustrated in Chapter 4. The design in this section presents some unique issues not previously discussed.

This section shows two GLM analyses. The first uses the same CLASS, MODEL, and RANDOM statements—adapted to GLM syntax—as the GLIMMIX program for the analysis of variance model. The second shows the considerable adaptation you need in order to get PROC GLM to retrieve most—but not all—of the correct analysis.

The blocking structure is confounded with the ABC interaction, and it takes a new variable to separate the ABC effect and leave the whole-plot error intact. When you use BLK(REP) as the whole-plot error, all of the sub-plot error is combined with the whole-plot error, and the error term in the analysis is left with zero degrees of freedom. To get around this, you must compute a new indicator variable, denoted ABC, that indicates which blocks have the “1” treatment combinations and the “+1” treatment combinations of the ABC interaction. Including ABC and ABC \times REP in the model enables the whole-plot comparisons to be extracted and leaves the sub-plot error intact.

Program—Naive PROC GLM

Including random effects in a regression model for PROC GLM causes some difficulties when one of the effects is confounded with some of the blocking structure or random effects. If you adapt the PROC MIXED program for the regression model shown in Section 16.8.2 to GLM syntax, you use the following:

```
proc glm data=fac_sp;
  class a b c rep blk;
  model y=rep blk(rep) a|b|c/solution;
  random rep blk(rep)/test;
  lsmeans a*b*c;
run;
```

Results

The results are shown in Output 16.20.

Output 16.20 Results from PROC GLM to Fit the Analysis of Variance Model

Source	Type III Expected Mean Square
a	Var(Error) + Q(a,a*b,a*c,a*b*c)
b	Var(Error) + Q(b,a*b,b*c,a*b*c)
a*b	Var(Error) + Q(a*b,a*b*c)
c	Var(Error) + Q(c,a*c,b*c,a*b*c)
a*c	Var(Error) + Q(a*c,a*b*c)
b*c	Var(Error) + Q(b*c,a*b*c)

Source	Type III Expected Mean Square
a*b*c	0
rep	Var(Error) + 4 Var(blk(rep)) + 8 Var(rep)
blk(rep)	Var(Error) + 4 Var(blk(rep))

*The GLM Procedure
Least Squares Means*

a	b	c	y LSMEAN
-1	-1	-1	Non-est
-1	-1	1	Non-est
-1	1	-1	Non-est
-1	1	1	Non-est
1	-1	-1	Non-est
1	-1	1	Non-est
1	1	-1	Non-est
1	1	1	Non-est

Interpretation

The SAS log will also provide the following note:

NOTE: No tests performed due to a MS with zero expectation.

You can see from the table of “Type III Expected Mean Squares” that the confounding of the A \times B \times C interaction and the BLK(REP) whole-plot error term results in the A \times B \times C interaction having 0 degrees of freedom. This occurs because PROC GLM does not distinguish between fixed and random effects (i.e., between the X and Z matrices) and hence cannot correctly compute confounded effects. This also causes PROC GLM to incorrectly declare the A \times B \times C least-squares means to be nonestimable, even though they clearly are estimable, provided that estimability theory is correctly applied.

Program—Modified PROC GLM to Obtain Tests of Model Effects

You can modify the program above as follows to obtain tests all model effects. Replace the MODEL statement with the following:

```
model y = a|b|c rep blk(rep) / e1;
```

This statement restricts analysis to the Type I Sums of Squares. It prevents the BLK(REP) and A \times B \times C terms from mutually canceling one another as they do with the Type III Sums of Squares. Output 16.21 shows the results.

Results

Output 16.21 Results from PROC GLM to Analysis of Variance Model with MODEL Statement Containing E1 Restriction

Source	Type I Expected Mean Square
a	Var(Error) + Q(a,a*b,a*c,a*b*c)
b	Var(Error) + Q(b,a*b,b*c,a*b*c)
a*b	Var(Error) + Q(a*b,a*b*c)
c	Var(Error) + Q(c,a*c,b*c,a*b*c)
a*c	Var(Error) + Q(a*c,a*b*c)
b*c	Var(Error) + Q(b*c,a*b*c)
a*b*c	Var(Error) + 4 Var(blk(rep)) + Q(a*b*c)
rep	Var(Error) + 4 Var(blk(rep)) + 8 Var(rep)
blk(rep)	Var(Error) + 4 Var(blk(rep))

Tests of Hypotheses for Mixed Model Analysis of Variance

	Source	DF	Type I SS	Mean Square	F Value	Pr > F
*	a	1	120.333333	120.333333	26.10	<.0001
*	b	1	114.083333	114.083333	24.74	<.0001
*	a*b	1	85.333333	85.333333	18.51	0.0002
*	c	1	533.333333	533.333333	115.66	<.0001
*	a*c	1	6.750000	6.750000	1.46	0.2358
*	b*c	1	120.333333	120.333333	26.10	<.0001
	blk(rep)	5	102.916667	20.583333	4.46	0.0037
	Error: MS(Error)	30	138.333333	4.611111		

* This test assumes one or more other fixed effects are zero.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
a*b*c	1	2.083333	2.083333	0.10	0.7632
rep	5	282.416667	56.483333	2.74	0.1461
Error: MS(blk(rep))	5	102.916667	20.583333		

Interpretation

You can see that the tests for all model effects are now the same as those obtained for both mixed model analyses. If you run the regression model with PROC GLM, you cannot obtain a valid estimate of β_{ABC} , the regression coefficient for the three-way interaction.

Program—Modified PROC GLM to Obtain LSMEANS

With considerable heroics using the ESTIMATE statement, you can obtain LSMEANS for the A \times B \times C treatment combinations. First, you define a variable ABC=A*B*C after the INPUT statement in the DATA step. This is the indicator variable mentioned in the introduction to

Section 16.8.3. The resulting PROC GLM statements needed to fit the model with the least-squares means are as follows:

```
proc glm data=fac_sp;
  class abc a b c rep ;
  model y=rep abc rep*abc a|b|c@2;
  random rep rep*abc / test;
  lsmeans a*b b*c / pdiff;
  estimate '(0)' intercept 1 abc 1 0 a 1 0 b 1 0 c 1 0
           a*b 1 0 0 0 a*c 1 0 0 0 b*c 1 0 0 0;
  estimate 'a' intercept 1 abc 0 1 a 0 1 b 1 0 c 1 0
           a*b 0 0 1 0 a*c 0 0 1 0 b*c 1 0 0 0;
  estimate 'b' intercept 1 abc 0 1 a 1 0 b 0 1 c 1 0
           a*b 0 1 0 0 a*c 1 0 0 0 b*c 0 0 1 0;
  estimate 'c' intercept 1 abc 0 1 a 1 0 b 1 0 c 0 1
           a*b 1 0 0 0 a*c 0 1 0 0 b*c 0 1 0 0;
  estimate 'ab' intercept 1 abc 1 0 a 0 1 b 0 1 c 1 0
           a*b 0 0 0 1 a*c 0 0 1 0 b*c 0 0 1 0;
  estimate 'ac' intercept 1 abc 1 0 a 0 1 b 1 0 c 0 1
           a*b 0 0 1 0 a*c 0 0 0 1 b*c 0 1 0 0;
  estimate 'bc' intercept 1 abc 1 0 a 1 0 b 0 1 c 0 1
           a*b 0 1 0 0 a*c 0 1 0 0 b*c 0 0 0 1;
  estimate 'abc' intercept 1 abc 0 1 a 0 1 b 0 1 c 0 1
           a*b 0 0 0 1 a*c 0 0 0 1 b*c 0 0 0 1;
run;
```

The terms ABC and ABC × REP provide the whole-plot analysis, and the term A|B|C@2 generates the main effects and the two-factor interactions (recall that ABC is extracting the ABC interaction and is not needed in this part of the definition of the model). The RANDOM statement identifies the random effects in the model and provides the expected mean squares for the Type III mean squares. The ESTIMATE statements define the A × B × C least-squares means in terms of the effects that produce estimable functions for each cell mean. The least-squares mean labels follow the Cochran and Cox notation described in the introduction to this section.

The results are shown in Output 16.22.

Results

Output 16.22 Results from PROC GLM Analyzing the Cell Means

Parameter	Estimate	Standard Error	t Value	Pr > t
(0)	119.666667	0.87665188	136.50	<.0001
a	120.333333	0.87665188	137.26	<.0001
b	122.666667	0.87665188	139.93	<.0001
c	124.333333	0.87665188	141.83	<.0001
ab	117.166667	0.87665188	133.65	<.0001
ac	122.666667	0.87665188	139.93	<.0001
bc	132.833333	0.87665188	151.52	<.0001
abc	126.666667	0.87665188	144.49	<.0001

Interpretation

You can see that the values in the “Estimate” column are identical to the least-squares means shown for the mixed model analysis in Output 16.18. The standard errors, however, are not. The PROC GLM output shows standard errors of 0.877, whereas the mixed model results you get from PROC GLIMMIX (or MIXED) have standard errors of 1.477. The difference is that GLM does not include the whole-plot variance in the standard errors. The PDIFF results for the A \times B and B \times C means and simple effects (not shown here in the interest of space) produce equivalent results to the GLIMMIX analysis for the simple effect comparisons. If you use PROC GLM to compare A \times B \times C means, you will not be able to obtain the correct standard errors of differences.

As with previous examples involving split-plot features, you can see that PROC MIXED and PROC GLIMMIX can compute appropriate analyses using CLASS, MODEL, and RANDOM statements that follow straightforwardly from the assumed models. On the other hand, PROC GLM produces at best a compromised, partially incorrect analysis and only then with considerable, often unintuitive, program “tricks.” Although the specific issues with the design in this section are unique, the bottom line is the same: *PROC GLM should not be used to analyze split-plot data.*

16.9 2^3 Treatment Structure in an Incomplete Block Design Structure with Balanced Confounding

16.9.1 Introduction

Incomplete block designs are useful when there are a large number of treatments or treatment combinations and it is not possible to observe each of the treatments in a homogeneous time period. When incomplete block designs are used, some of the information about the treatment effects is confounded with the blocks. One strategy, described in Section 16.8, is to confound one higher-order interaction with the block effects and generate a split-plot design structure. Another strategy, shown in this section, is to confound all interactions of the same order with blocks an equal number of times within some replication. Unlike the experiment in Section 16.8, different interactions are confounded with blocks in each of the replications.

The following design involves a 2^3 treatment structure in five replications of two blocks each. The ABC interaction is confounded in replications one and two, the AB interaction is confounded in replication three, the AC interaction is confounded in replication four, and the BC interaction is confounded in replication five. Intra-block (or fixed block) analysis extracts information from the design from the within-block (or inter-block) comparisons. The mixed model analysis with random block effects also extracts information from the between-block comparisons (inter-block analysis) and combines them with inter-block comparisons into a common estimate. For this design, all main effects information comes from the within-block comparisons, but for the interactions, some of the information comes from the within-block comparisons and some from the between-block comparisons. This, the combined analysis, obtained via the mixed model with random block effects, is more efficient.

PROC GLIMMIX and PROC MIXED can provide either inter-block analysis only or combined inter- and intra-block analysis. PROC GLM provides inter-block analysis only. The example in this section demonstrates combined inter- and intra-block mixed model analysis.

The data in Data Set 16.9, “Edge Slope,” in Appendix 2, “Data Sets,” are data similar to a photoresist coating experiment in the semiconductor industry (the design structure is the same but the variables and response values are altered to preserve confidentiality). The objective was to study the effect of developing time at two levels (A), developing temperature at two levels (B), and exposure at two levels (C) on the edge slope (EDGSLP) of silicon wafers. Groups of six wafers were subjected to a set of conditions and the edge slope was determined as the average from the set of six wafers. Four runs could be carried out during a shift, and thus blocks of size four were used in the design structure (these are denoted as BLK in the data set). A complete replication, consisting of two shifts with four runs each, could be run in one day (these are denoted as REP in the data set). The experiment was repeated for five days, thus providing five replications. The ABC interaction was confounded in replication one and two, and the AB, AC, and CB interactions were each confounded in replications three, four, and five, respectively, generating a balanced confounded design.

As in Section 16.8, the data can be analyzed using an analysis of variance of the means or using a regression model.

The analysis of variance model is

$$Y_{ijkmn} = \mu + \alpha_k + \beta_m + \gamma_n + (\alpha\beta)_{km} + (\alpha\gamma)_{kn} + (\beta\gamma)_{mn} + (\alpha\beta\gamma)_{kmn} + r_i + a_{ij} + e_{ijkmn}$$

where

μ denotes the intercept

α_k , ($k = -1, 1$), denotes the k^{th} level of factor A

β_m , ($m = -1, 1$), denotes the m^{th} level of factor B

γ_n , ($n = -1, 1$), denotes the n^{th} level of factor C

$(\alpha\beta)_{km}, \dots, (\alpha\beta\gamma)_{kmn}$ denote the interactions among the factors

r_i , ($i = 1, 2, \dots, 6$), denotes the i^{th} replication, or day,
assumed iid $N(0, \sigma_r^2)$

a_{ij} , ($j = 1, 2$), denotes the j^{th} block, or shift within a day, of the i^{th} replication,
assumed iid $N(0, \sigma_a^2)$

e_{ijkmn} denotes random error, corresponding to the group of six wafers, assumed iid $N(0, \sigma_e^2)$

The regression model or response surface model is

$$\begin{aligned} Y_{ijk} = & \beta_0 + \beta_A A_k + \beta_B B_k + \beta_C C_k + \beta_{AB} A_k B_k + \beta_{AC} A_k C_k \\ & + \beta_{BC} B_k C_k + \beta_{ABC} A_k B_k C_k + r_i + a_{ij} + e_{ijkmn} \end{aligned}$$

where

β_0 is the intercept

A_k , B_k , and C_k denote values of factors A, B, and C, respectively

β_A , β_B , and β_C are linear regression coefficients for A, B, and C

$\beta_{AB}, \dots, \beta_{ABC}$ are cross-product interaction regression coefficients

r_i , a_{ij} , and e_{ijkmn} are defined as above

16.9.2 Analysis Using PROC GLIMMIX and PROC MIXED

As with the analyses in Section 16.8, you can use either PROC MIXED or PROC GLIMMIX for the analysis of variance or regression model. PROC GLIMMIX is shown for the analysis of variance (means) model because the SLICEDIFF option with the LSMEANS statement makes it easier to evaluate simple effects for significant interactions.

Program—Analysis of Variance (Means) Model

The PROC GLIMMIX program to fit the model for analyzing the means of the treatment combinations is as follows:

```
proc glimmix data=resist;
  class a b c rep blk;
  model eds1p = a|b|c / ddfm=kr;
  random rep blk(rep);
  lsmeans a*b b*c / slicediff=(a b c);
run;
```

The REP and BLK(REP) terms in the RANDOM statement provide the estimates of the variance components for day-to-day variation and shift-to-shift (within a day) variation, respectively. The DDFM=KR option is used to obtain approximate degrees of freedom for all fixed effect tests. The DDFM=SATTERTH method yields identical results for these data. Use the SLICEDIFF option to obtain estimates of the simple effects of A given B and vice versa and B given C and vice versa. These aid in interpreting the $A \times B$ and $B \times C$ interactions, which are statistically significant for these data. The results of the PROC GLIMMIX are shown in Output 16.23.

Results

Output 16.23 PROC GLIMMIX Results for the Factorial Treatment Structure in a Balanced Confounded Design Structure for the Analysis of the Means

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
rep	0.08993	0.1382
blk(rep)	0.1672	0.1129
Residual	0.03831	0.01128

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
a	1	23.05	2.61	0.1198
b	1	23.05	11.51	0.0025
a*b	1	23.66	0.09	0.7639
c	1	23.05	84.82	<.0001
a*c	1	23.66	9.27	0.0056
b*c	1	23.66	6.87	0.0151
a*b*c	1	24.61	0.04	0.8503

a*b Least Squares Means						
a	b	Estimate	Standard Error	DF	t Value	Pr > t
-1	-1	4.7505	0.1969	4.719	24.12	<.0001
-1	1	4.9395	0.1969	4.719	25.08	<.0001
1	-1	4.8295	0.1969	4.719	24.53	<.0001
1	1	5.0605	0.1969	4.719	25.70	<.0001

Simple Effect Comparisons of a*b Least Squares Means By a							
Simple Effect Level	b	_b	Estimate	Standard Error	DF	t Value	Pr > t
a -1	-1	1	-0.1890	0.09283	23.39	-2.04	0.0532
a 1	-1	1	-0.2310	0.09283	23.39	-2.49	0.0204

b*c Least Squares Means						
b	c	Estimate	Standard Error	DF	t Value	Pr > t
-1	-1	4.5957	0.1969	4.719	23.34	<.0001
-1	1	4.9843	0.1969	4.719	25.31	<.0001
1	-1	4.6243	0.1969	4.719	23.48	<.0001
1	1	5.3757	0.1969	4.719	27.30	<.0001

Simple Effect Comparisons of b*c Least Squares Means By c							
Simple Effect Level	b	_b	Estimate	Standard Error	DF	t Value	Pr > t
c -1	-1	1	-0.02865	0.09283	23.39	-0.31	0.7603
c 1	-1	1	-0.3913	0.09283	23.39	-4.22	0.0003

Interpretation

From the “Covariance Parameter Estimates” table you see that the shift, or BLK(REP), variance component estimate is 0.1672, about twice as large as the day (REP) variance component (0.0899).

The A × C and B × C interactions are statistically significant ($p = 0.0056$ and $p = 0.0151$, respectively, in the “Type III Tests of Fixed Effects” table). Note that the approximate denominator degrees of freedom vary according to the confounding structure of the design. The main effects use intra-block information only, the two-way interactions gain inter-block information from being confounded with blocks in one replication, and the three-way interaction gains inter-block information from being confounded in two replications.

In the interest of space, only the simple effects providing the greatest insight into the interaction are shown for the least-squares mean analysis. For the A × B means, the simple effect

differences between B levels for given levels of A provide the greatest insight into the $A \times B$ interaction. At $A = -1$, the least-squares mean for $B = -1$ is 4.75 and the least-squares mean for $B = 1$ is 4.94, a difference of 0.19 with a standard error of a difference of 0.0923. The p -value for this difference is 0.0523. At $A = 1$, the least-squares means are 4.83 and 5.06, respectively, for $B = -1$ and $B = 1$, respectively. The difference is 0.23 with a standard error of 0.0923 and a p -value of 0.0204. Thus the B effect is somewhat greater given $A = 1$.

For the $B \times C$ means the simple effect differences of B given C provide the greatest insight. At $C = -1$, the least-squares means are 4.60 and 4.62, respectively, for $B = -1$ and $B = 1$. At $C = 1$, the least-squares means are 4.98 and 5.37. Thus, there is a large B effect (difference 0.39, standard error 0.093, $p = 0.0003$) at $C = 1$, but not at $C = -1$ ($p = 0.7603$). The $B \times C$ least-squares means also reveal a substantial C effect over and above the interaction, which is consistent with the F -value for C of 84.82 in the “Type III Tests of Fixed Effects” table.

Program—Regression Model

The PROC MIXED program to fit the regression model is as follows:

```
proc mixed data=resist;
  class rep blk;
  model edsdp = a|b|c / solution ddfm=satterth;
  random rep blk(rep);
run;
```

Alternatively, you could run the same CLASS, MODEL, and RANDOM statements with PROC GLIMMIX and obtain the same results. Also, the DDFM=KR option generates identical results for this model.

Results

The results are shown in Output 16.24.

Output 16.24 PROC MIXED Results for the Factorial Treatment Structure in a Balanced Confounded Design Structure for the Regression Model

Covariance Parameter Estimates	
Cov Parm	Estimate
rep	0.08993
blk(rep)	0.1672
Residual	0.03831

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	4.8950	0.1889	4	25.92	<.0001
a	0.05000	0.03095	23.1	1.62	0.1198
b	0.1050	0.03095	23.1	3.39	0.0025
a*b	0.01051	0.03459	23.7	0.30	0.7639
c	0.2850	0.03095	23.1	9.21	<.0001
a*c	-0.1053	0.03459	23.7	-3.05	0.0056

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
b*c	0.09067	0.03459	23.7	2.62	0.0151
a*b*c	-0.00761	0.03990	24.6	-0.19	0.8503

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
a	1	23.1	2.61	0.1198
b	1	23.1	11.51	0.0025
a*b	1	23.7	0.09	0.7639
c	1	23.1	84.82	<.0001
a*c	1	23.7	9.27	0.0056
b*c	1	23.7	6.87	0.0151
a*b*c	1	24.6	0.04	0.8503

Interpretation

Because the regression model has the same number of parameters as the number of cells in the treatment structure, the estimates of the variance components are identical to those obtained from the analysis of variance model. If the model is simplified by excluding nonsignificant terms, the estimates of the variance components will change. The analysis of the Type 3 Tests of Fixed Effects yields results identical to the analysis of variance model. The “Solution for Fixed Effects” output contains the estimated coefficients, $\hat{\beta}_A$, $\hat{\beta}_B$, ..., $\hat{\beta}_{AB}$. The standard errors and degrees of freedom reflect the confounding structure. You do not need to know the type of confounding in the design to carry out the analysis. PROC MIXED’s estimation procedure extracts the available information about the regression coefficients and provides the combined estimates.

16.10 Product Acceptability Study with Crossover and Repeated Measures

16.10.1 Introduction

The data in Data Set 16.10, “Use Products,” in Appendix 2, “Data Sets,” represent results from an experiment to evaluate the acceptability of four types of daily use products (such as toothpaste, facial soap, tissue, automobile, etc.; the data are simulated). Each person used three of the possible four products in a crossover type of process and rated a given product after one week of use, two weeks of use, and three weeks of use. The four products were grouped into four blocks of size three forming a balanced incomplete block, as indicated in Table 16.7.

Table 16.7 The Balanced Incomplete Block Used to Group the Four Products into Blocks of Size Three

Block	Products
1	1, 2, 3
2	1, 2, 4
3	1, 3, 4
4	2, 3, 4

The three products within each block were used in all six possible sequences, thus providing a crossover design with the three products in each block, as shown in Table 16.8.

Table 16.8 The Six Sequences of Three Products Used within Each Block

BLOCK	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6
1	1 2 3	1 3 2	2 1 3	2 3 1	3 2 1	3 1 2
2	1 2 4	1 4 2	2 1 4	2 4 1	4 2 1	4 1 2
3	1 3 4	1 4 3	3 1 4	3 4 1	4 3 1	4 1 3
4	2 3 4	2 4 3	3 2 4	3 4 2	4 3 2	4 2 3

There are 24 total sequences, six from each of the four blocks. Twenty-four women were selected from a group of college-age women residing in a university town. The women were randomly assigned to the 24 sequences, providing one woman per sequence. Each woman used the three products in the designated sequence and rated the acceptability of the products at the end of one week, two weeks, and three weeks of exposure. The acceptability scale is from 0 to 7, with 0 being unacceptable and 7 extremely acceptable. Using a product means that the woman used the product daily for the seven days prior to providing an acceptability rating.

This particular data structure has several very interesting features. First, the woman uses the three specified products in a given sequence. Each product is used for three weeks with a one-week interval of using a control product. Thus, the woman's time in the experiment consists of three four-week periods, and she uses a different product within each period. The experimental unit for product is one four-week interval. Each woman is repeatedly measured three times as she uses three products in the three four-week intervals, thus providing a repeated measures aspect to the experiment. Second, each woman rates each product three times (one week apart), providing another level of repeated measures to the experiment.

The researcher was interested in evaluating the acceptability of products and as well as determining if the acceptability changes over time with use. The treatment structure is a three-way with levels of products (four levels) crossed with levels of time (three levels measured weekly within a product). A third factor is the period (with three levels) in which the products were used, where periods 1, 2, and 3 correspond to the first four weeks, second four weeks, and third four weeks, respectively.

The design structure consists of (1) the blocks of three treatments, (2) the women within each block where the three treatments were observed in the six sequences, (3) the four weeks of time

a product was used, and (4) the one-week time intervals between acceptability ratings. A model to describe the data is

$$Y_{ijkmn} = \mu_{mn} + \rho_k + \lambda_{t_{k-1}} + b_i + w_{j(i)} + p_{k(ij)} + e_{ijkmn}$$

where

μ_{mn} is the expected mean acceptability rating for the m^{th} product after the n^{th} week of exposure

ρ_k is the effect of the k^{th} period or set of four weeks

$\lambda_{t_{k-1}}$ is the carry-over effect of treatment t_{k-1} occurring in the previous four-week period prior to the current four-week period ($\lambda_{t_{k-1}}$ does not occur in the model when $k = 1$)

b_i is the random block effect, assumed $iid N(0, \sigma_b^2)$

$w_{j(i)}$ is the random effect for the variation in the women participating in the study, assumed $iid N(0, \sigma_w^2)$

$p_{k(ij)}$ is the random effect corresponding to the four-week time intervals or periods, with assumed distribution

$$\begin{pmatrix} p_{1(ij)} \\ p_{2(ij)} \\ p_{3(ij)} \end{pmatrix} \sim iid N\left(0, \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}\right)$$

e_{ijkmn} is the random effect of the one-week time intervals within a product or period with assumed distribution

$$\begin{pmatrix} e_{ijkml} \\ e_{ijkm2} \\ e_{ijkm3} \end{pmatrix} \sim iid N\left(0, \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}\right)$$

Let matrices \mathbf{R}^* and \mathbf{A} be defined by

$$\mathbf{R}^* = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

The covariance matrix of the vector of data is

$$\text{Var}[\mathbf{Y}] = \sigma_b^2 \mathbf{I}_4 \otimes \mathbf{J}_6 \otimes \mathbf{J}_3 \otimes \mathbf{J}_3 + \sigma_w^2 \mathbf{I}_4 \otimes \mathbf{I}_6 \otimes \mathbf{J}_3 \otimes \mathbf{J}_3 + \mathbf{I}_4 \otimes \mathbf{I}_6 \otimes \mathbf{A} \otimes \mathbf{J}_3 + \mathbf{I}_4 \otimes \mathbf{I}_6 \otimes \mathbf{I}_3 \otimes \mathbf{R}$$

where $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of the matrices \mathbf{A} and \mathbf{B} .

The analysis is carried out with simplifying assumptions on \mathbf{A} and \mathbf{R}^* . The independent errors assumptions are $\mathbf{A} = \sigma_p^2 \mathbf{I}$ and $\mathbf{R} = \sigma_e^2 \mathbf{I}$. The first-order autoregressive error structure assumptions are

$$\mathbf{A} = \sigma_p^2 \begin{bmatrix} 1 & \rho_a & \rho_a^2 \\ \rho_a & 1 & \rho_a \\ \rho_a^2 & \rho_a & 1 \end{bmatrix} \text{ and } \mathbf{R} = \sigma_e^2 \begin{bmatrix} 1 & \rho_e & \rho_e^2 \\ \rho_e & 1 & \rho_e \\ \rho_e^2 & \rho_e & 1 \end{bmatrix}$$

The analysis of the data set is carried out in three parts. The first analysis uses the independent errors model for both the $p_{k(ij)}$ and the e_{ijkmn} . The second analysis uses the independent errors model for the $p_{k(ij)}$ and the first-order autoregressive errors model for the e_{ijkmn} . The third analysis uses the first-order autoregressive errors model for both the $p_{k(ij)}$ and the e_{ijkmn} .

16.10.2 Variable Definitions

The examples that follow use the following variable definitions. The variables BLK and SEQ denote the block number and the sequence number, respectively, from Table 16.8. PERSON is the identification code assigned to each woman in the study. PERIOD corresponds to the four-week interval. PROD denotes the product in use during the current period, and PRIORPRD denotes the product used in the previous four-week period. PRIORPRD has a value of zero for data from first four-week period and thus can take on the value 0, 1, 2, 3, or 4. TIME refers to the one-week interval within the period when a product is in use. L_i is one if the i^{th} product is the prior product and zero if the i^{th} product is not the prior product. The variables L1, L2, L3, and L4 are used to enable the $\lambda_{k_{-1}}$ to be included in the model.

16.10.3 Independent Error Model

Program—Including PRIORPRD

Use the following program to fit the independent errors model to the data and to determine if the carry-over effects are different.

```
proc mixed data=prior;
  class blk person seq prod period time priorprd;
  model y=prod period time time*prod priorprd/ddfm=satterth;
  random blk person(blk) period*person(blk);
  lsmeans prod time*prod;
  estimate 'prod 1 v 2' prod 1 -1 0 0;
run;
```

The MODEL option DDFM=SATTERTH is used to provide approximate denominator degrees of freedom for all tests and comparisons. The independent errors model is specified by the RANDOM statement that includes no additional options and by not including a REPEATED statement.

Relevant results appear in Output 16.25.

Results

Output 16.25 Results of Fitting the Independent Errors Model to the Acceptability Data Using the PRIORPRD Class Variable

Covariance Parameter Estimates	
Cov Parm	Estimate
blk	0.4982
person(blk)	0.3171
person*period(blk)	0.1549
Residual	0.09831

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	40.9	4.42	0.0088
period	1	39.9	0.05	0.8254
time	2	136	29.38	<.0001
prod*time	6	136	15.01	<.0001
priorprd	3	43.4	4.68	0.0064

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
prod 1 v 2	-0.2689	0.1692	40.9	-1.59	0.1196

Interpretation

There is a significant PRIORPRD effect ($F = 4.68, p = 0.0064$), indicating that the carry-over effects are not equal. Fortunately, this model allows for the estimation of the PROD effects in the presence of unequal carry-over effects. Unfortunately, using PRIORPRD provides a model that is over-parameterized and the PROD and PROD \times TIME least-squares means are not estimable. However, you can use ESTIMATE statements to compute differences between the least-squares means—e.g., the example showing the comparison of product 1 versus 2.

Program—including Li Terms

It is easier to carry out the analysis using a model with the L_i values instead of the PRIORPRD values. You can make the substitution in the PROC MIXED program shown above, or you can use PROC GLIMMIX. While both procedures give essentially interchangeable results, PROC GLIMMIX is used in this example mainly because the SLICEDIFF option allows you to evaluate the TIME \times COND interaction, which in this case is significant, more easily.

```
proc glimmix data=prior;
  class blk person seq prod period time PRIORPRD;
  model y=prod period time time*prod L1 L2 L3 L4/ ddfm=satterth;
  random blk person(blk) prod*person (blk);
  lsmeans time*prod/slicediiff=(time prod);
run;
```

The variance component and Type III test results appear in Output 16.26. Selected LSMEAN results appear in Output 16.27.

Results

Output 16.26 Type III F-Statistics for the Fixed Effects from the Independent Errors Model to the Acceptability Data Using the L_i 's as Variables

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
blk	0.4982	0.4621
person(blk)	0.3171	0.1217
person*prod(blk)	0.1549	0.04220
Residual	0.09831	0.01192

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	40.89	4.42	0.0088
period	1	39.92	0.05	0.8254
time	2	136	29.38	<.0001
prod*time	6	136	15.01	<.0001
l1	0	.	.	.
l2	0	.	.	.
l3	0	.	.	.
l4	0	.	.	.

Interpretation

The estimates of the variance components are identical to those in Output 16.25 obtained using the PRIORPRD. The results for PROD, PERIOD, TIME, and PROD \times TIME are identical to those in Output 16.25. The least-squares means for PROD and TIME \times PROD, however, are estimable with this model.

The Type III F-statistics for L1, L2, L3, and L4 are zero because there is a dependency or collinearity between the levels of PERIOD and the L_i 's.

Output 16.27 Least-Squares Means and Simple Effect Comparisons among COND by TIME Fitting the Independent Errors Model to the Acceptability Data

prod*time Least Squares Means						
prod	time	Estimate	Standard Error	DF	t Value	Pr > t
1	1	3.9111	0.3934	3.591	9.94	0.0010
1	2	3.6334	0.3934	3.591	9.24	0.0013
1	3	2.8556	0.3934	3.591	7.26	0.0029
2	1	4.0319	0.3934	3.591	10.25	0.0009
2	2	3.6986	0.3934	3.591	9.40	0.0012
2	3	3.4763	0.3934	3.591	8.84	0.0015
3	1	3.8112	0.3934	3.591	9.69	0.0011
3	2	3.7001	0.3934	3.591	9.41	0.0012
3	3	3.6446	0.3934	3.591	9.26	0.0012
4	1	3.9124	0.3934	3.591	9.95	0.0010
4	2	4.1902	0.3934	3.591	10.65	0.0008
4	3	4.1346	0.3934	3.591	10.51	0.0008

Simple Effect Comparisons of prod*time Least Squares Means By time							
Simple Effect Level	prod	_prod	Estimate	Standard Error	DF	t Value	Pr > t
time 1	1	2	-0.1208	0.1895	63.12	-0.64	0.5262
time 1	1	3	0.09992	0.1895	63.12	0.53	0.5998
time 1	1	4	-0.00126	0.1895	63.12	-0.01	0.9947
time 1	2	3	0.2207	0.1895	63.12	1.16	0.2485
time 1	2	4	0.1195	0.1895	63.12	0.63	0.5305
time 1	3	4	-0.1012	0.1895	63.12	-0.53	0.5952
time 2	1	2	-0.06520	0.1895	63.12	-0.34	0.7319
time 2	1	3	-0.06674	0.1895	63.12	-0.35	0.7258
time 2	1	4	-0.5568	0.1895	63.12	-2.94	0.0046
time 2	2	3	-0.00154	0.1895	63.12	-0.01	0.9935
time 2	2	4	-0.4916	0.1895	63.12	-2.59	0.0118
time 2	3	4	-0.4901	0.1895	63.12	-2.59	0.0120
time 3	1	2	-0.6208	0.1895	63.12	-3.28	0.0017
time 3	1	3	-0.7890	0.1895	63.12	-4.16	<.0001
time 3	1	4	-1.2790	0.1895	63.12	-6.75	<.0001
time 3	2	3	-0.1682	0.1895	63.12	-0.89	0.3781

Simple Effect Comparisons of prod*time Least Squares Means By time							
Simple Effect Level	prod	_prod	Estimate	Standard Error	DF	t Value	Pr > t
time 3	2	4	-0.6583	0.1895	63.12	-3.47	0.0009
time 3	3	4	-0.4901	0.1895	63.12	-2.59	0.0120

Simple Effect Comparisons of prod*time Least Squares Means By prod							
Simple Effect Level	time	_time	Estimate	Standard Error	DF	t Value	Pr > t
prod 1	1	2	0.2778	0.1045	136	2.66	0.0088
prod 1	1	3	1.0556	0.1045	136	10.10	<.0001
prod 1	2	3	0.7778	0.1045	136	7.44	<.0001

Interpretation

The standard errors of the differences between pairs of PROD means at a common level of TIME are all identical (0.1895), and the Satterthwaite approximation to the associated degrees of freedom is 63.1. There are no significant differences among the levels of PROD at TIME = 1. Level 4 of PROD is significantly more acceptable than the other three levels of PROD at TIME = 2. At TIME = 3, level 1 of PROD is significantly less acceptable than the other three levels, levels 2 and 3 are not different, and level 4 of PROD is significantly more acceptable than the other three levels.

Also included in Output 16.28 are comparisons among the levels of TIME for PROD 1. The standard errors of the differences between pairs of TIME means for PROD 1 (0.1045) depend on the residual variance only. Therefore, they are based on the residual degrees of freedom (136). You can use ODS graphics with PROC GLIMMIX to plot least-squares means over TIME for each PROD level.

Program—ODS Graphics with PROC GLIMMIX to Plot TIME Effect by PROD

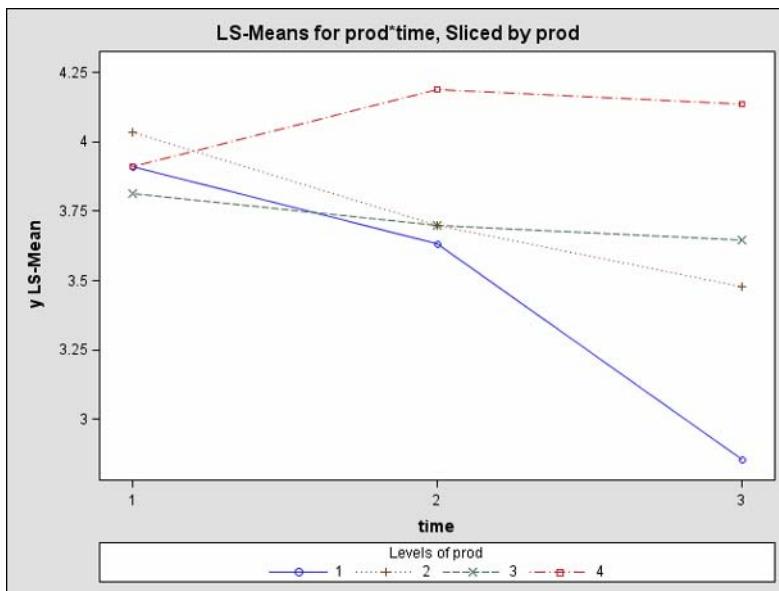
Use the following program to generate the interaction plot.

```

ods html;
ods graphics on;
ods select MeanPlot;
proc glimmix data=prior;
  class blk person seq prod period time PRIORPRD;
  model y=prod period time time*prod L1 L2 L3 L4/ ddfm=satterth;
  random blk person(blk) prod*person (blk);
  lsmeans time*prod/plot=meanplot (sliceby=prod join);
run;
ods graphics off;
ods html close;
run;

```

The results appear in Figure 16.9. You can see that for PROD 1, acceptability tends to decrease over TIME. For PROD 2 it remains essentially constant over TIME. For PROD 3, acceptability tends to increase.

Figure 16.9 Plot of Product Means across Time

16.10.4 Autoregressive Errors for the Time Interval Part of the Model

Program—Autoregressive Errors

The program to fit a model where \mathbf{R}^* satisfies the autoregressive error structure and \mathbf{A} still has the independent errors structure is as follows:

```
proc mixed data=prior;
  class blk person seq prod period time priorprd;
  model y=prod period time time*prod priorprd/ ddfm=satterth;
  random blk person(blk) period(person blk);
  repeated time/subject=period*person(blk) type=ar(1);
run;
```

The REPEATED statement is used to specify a first-order autoregressive error structure for the three measurements made after one, two, and three weeks of use within a product or period for each person. The option TYPE = AR(1) specifies the covariance structure for the three observations common to a person within a PERIOD, specified as SUBJECT=PERIOD*PERSON*BLK. Note that this program uses DDFM=SATTERTH to determine approximate degrees of freedom. In general this is not recommended for correlated error models because it does not correct for potential bias in the standard errors and test statistics. You should use DDFM=KR instead. The Satterthwaite option is retained here so that you can compare the results (shown in Output 16.28) to those for the KR option (shown in Output 16.30).

Readers comparing this edition to the first edition of Littell et al. (1996) will notice that the PARMS statement providing starting values to speed the computations was used in the first edition and does not appear here. This is a reflection of improvements both in PROC MIXED algorithms and in the quality and speed of computer hardware: the PARMS statement no longer has a perceptible impact on computing speed.

The results of using the statements above to analyze the acceptability data are shown in Output 16.28.

Results

Output 16.28 Results of Fitting the Autoregressive Time Errors and Independent Period Errors Model to the Acceptability Data Using the PRIORPRD Class Variable

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
blk		0.4997
person(blk)		0.3150
period(blk*person)		0.1380
AR(1)	person*period(blk)	0.2013
Residual		0.1154

Fit Statistics	
-2 Res Log Likelihood	309.8
AIC (smaller is better)	319.8
AICC (smaller is better)	320.2
BIC (smaller is better)	316.8

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	41	4.41	0.0089
period	1	39.9	0.05	0.8315
time	2	97.3	26.52	<.0001
prod*time	6	97.3	13.67	<.0001
priorprd	3	43.4	4.71	0.0062

Interpretation

The estimate of the AR(1) parameter is $\hat{\rho} = 0.20$. You can compare the various criteria given in the “Fit Statistics” output to see if the AR(1) model provides a better fit. For example, the AICC for the AR(1) model is 320.2 versus 319.2 for the independent errors model, indicating that the AR(1) model does not provide improved fit. Alternatively, you can use the “-2 Res Log Likelihood” from the AR(1) and independent errors analysis to compute a likelihood ratio test of $H_0: \rho = 0$. The resulting chi-square statistic with 1 degree of freedom is $311.0 - 309.8 = 1.2$. You do not reject H_0 , and hence you conclude that the AR(1) model does not improve the fit.

As expected, since the AR(1) model provided a fit similar to the independent errors model, the resulting tests of fixed effects are similar. The Type III F -values for the fixed effects part of the model are somewhat smaller than the corresponding values for the independent errors model, but the overall conclusions are essentially identical.

Program—AR(1) Time Errors and the L_i Terms

The program for fitting the model with the L_i 's to provide for the estimation of the LSMEANS is as follows:

```
proc glimmix data=prior;
  class blk person seq prod period time priorprd;
  model y = prod period time time*prod L1 L2 L3 L4/ ddfm=kr;
  random blk person(blk) period*person (blk);
  random _residual_ / subject=period*person(blk) type=ar(1);
  lsmeans time*prod / slicediff=(time prod);
run;
```

For the AR(1) model, PROC GLIMMIX uses different syntax than PROC MIXED. In place of the REPEATED statement, GLIMMIX uses the RANDOM _RESIDUAL_ statement. In other words, PROC GLIMMIX does not have a REPEATED statement. The corresponding R-side covariance structures are specified in the RANDOM statement as well. Otherwise, the syntax for TYPE= and SUBJECT= options are the same. This model uses the DDFM=KR option, which is recommended, whereas the MIXED program that generates Output 16.28 used DDFM=SATTERTH (which is not recommended with correlated error models).

Results

The estimates of the variance components and the tests of the fixed effects appear in Output 16.29. Output 16.30 shows LSMEANS results.

Output 16.29 Type III F-Statistics for the Fixed Effects from the Independent Errors Model to the Acceptability Data Using the L_i 's as Variables

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
blk		0.4997	0.4630
person(blk)		0.3150	0.1210
person*period(blk)		0.1380	0.04737
AR(1)	person*period(blk)	0.2013	0.1954
Residual		0.1154	0.02686

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	40.98	4.45	0.0086
period	1	39.92	0.05	0.8302
time	2	92.53	24.96	<.0001
prod*time	6	108.3	12.83	<.0001
l1	0	.	.	.
l2	0	.	.	.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
l3	0	.	.	.
l4	0	.	.	.

Interpretation

As with the independent errors models, the PRIORPRD and L_i parameterizations yield identical variance component estimates. The Fit Statistics (not shown here) are also the same. Again the Type III F -statistics for the L_i 's are zero as there is a collinearity between the L_i 's and the levels of PERIOD. There are slight discrepancies between the denominator degrees of freedom (Den DF), f - and p -values, resulting from the use of the KR option instead of the SATTERTH option used in Output 16.28. For example, for PROD \times TIME using DDFM=KR, $F = 12.83$ (DenDF = 108.3, $p < 0.0001$) versus $F = 13.67$, DenDF = 97.3 using DDFM=SATTERTH. In general, for correlated error models you should use the KR option. Note that these differences are not due to having used PROC MIXED to generate Output 16.28 and PROC GLIMMIX to generate Output 16.29. They are only due to having used different DDFM= options, which you can verify by changing DDFM=SATTERTH to DDFM=KR in the SAS code that generates Output 16.28.

Output 16.30 Least-Squares Means and Simple Effect Comparisons among PROD by TIME Fitting the AR(1) Model to the Acceptability Data

Simple Effect Comparisons of prod*time Least Squares Means By time							
Simple Effect Level	prod	_prod	Estimate	Standard Error	DF	t Value	Pr > t
time 1	1	2	-0.1218	0.1900	63.2	-0.64	0.5237
time 1	1	3	0.1040	0.1900	63.2	0.55	0.5859
time 1	1	4	-0.00057	0.1900	63.2	-0.00	0.9976
time 1	2	3	0.2258	0.1900	63.2	1.19	0.2389
time 1	2	4	0.1212	0.1900	63.2	0.64	0.5256
time 1	3	4	-0.1046	0.1900	63.2	-0.55	0.5838
time 2	1	2	-0.06625	0.1900	63.2	-0.35	0.7284
time 2	1	3	-0.06264	0.1900	63.2	-0.33	0.7427
time 2	1	4	-0.5561	0.1900	63.2	-2.93	0.0047
time 2	2	3	0.003616	0.1900	63.2	0.02	0.9849
time 2	2	4	-0.4899	0.1900	63.2	-2.58	0.0123
time 2	3	4	-0.4935	0.1900	63.2	-2.60	0.0117
time 3	1	2	-0.6218	0.1900	63.2	-3.27	0.0017
time 3	1	3	-0.7849	0.1900	63.2	-4.13	0.0001
time 3	1	4	-1.2783	0.1900	63.2	-6.73	<.0001
time 3	2	3	-0.1631	0.1900	63.2	-0.86	0.3939
time 3	2	4	-0.6565	0.1900	63.2	-3.46	0.0010
time 3	3	4	-0.4935	0.1900	63.2	-2.60	0.0117

Simple Effect Comparisons of prod*time Least Squares Means By prod							
Simple Effect Level	time	_time	Estimate	Standard Error	DF	t Value	Pr > t
prod 1	1	2	0.2778	0.1036	118.9	2.68	0.0084
prod 1	1	3	1.0556	0.1140	70.62	9.26	<.0001
prod 1	2	3	0.7778	0.1036	118.9	7.51	<.0001

Because the error models are different, there are slight differences between the least-squares means in Outputs 16.27 and 16.30 as well as between their estimated standard errors. The TIME \times PROD least-squares mean estimates (not shown here in the interest of space) for the two model are virtually identical (usually agreeing to the second decimal place). Their standard errors are 0.3938 (versus 0.3934 for the independent errors model). The main impact of the AR(1) error structure for the time intervals is seen in the estimated standard errors for the difference between two TIME means at the same level of PROD. The estimated standard error for the difference between two means one time period apart is 0.1036 and two time periods apart is 0.1140. Notice that the degrees of freedom also depend on the distance between observations. The estimated standard error of the differences of pairs of PROD means at the same level of TIME is 0.1900 versus 0.1895 for the independent error model. For these data, the overall conclusions are similar.

16.10.5 Autoregressive Errors for the Time Interval and Period Parts of the Model

As with the independent errors and AR(1) errors models, you can use either PROC MIXED or PROC GLIMMIX to fit the model with autoregressive period effects. As in Sections 16.10.3 and 16.10.4, the model with PRIORPRD is demonstrated with PROC MIXED and the model with the Li's is computed with PROC GLIMMIX to take advantage of its greater convenience in working with simple effect differences.

PROC MIXED Program—Model with PRIORPRD Effect and AR(1) Error Structure for Time Interval and Period Effects

Use the following PROC MIXED program to specify autoregressive errors for both the time intervals and the periods part of the model:

```
proc mixed data=prior;
  class blk person seq prod period time priorprd;
  model y = prod period time time*prod priorprd / ddfm=kr;
  random int person / subject=blk;
  random period / subject=person(blk) type=ar(1);
  repeated time / subject=period*person(blk) type=ar(1);
run;
```

The REPEATED statement is the same as in the previous program (Section 16.10.4), but the RANDOM statement has been split into two parts. The first RANDOM statement specifies the block and the persons within block effects, i.e., random effects assumed to be independent (this is a reasonable assumption unless the women in the study are all genetically or socially related). The second RANDOM statement specifies the error structure for the PERIOD errors or the four-week time interval effects, which are assumed to be correlated. TYPE=AR(1) specifies the covariance structure among period effects within a given person, and

SUBJECT=PERSON(BLK) defines the three period effects as dependent with a person and independent between persons. The DDFM=KR is used to compute approximate denominator degrees of freedom and to adjust standard errors and test statistics for bias.

Notice that the two RANDOM statements and the REPEATED statement are specified with SUBJECT= options and that the SUBJECT effects are nested. That is, the BLK effect is contained in the PERSON(BLK) effect, which is contained in the PERIOD*PERSON(BLK) effect. For a definition of containment, see Chapter 11, “The Four Types of Estimable Functions,” in the *SAS/STAT User’s Guide*. Briefly, an effect containing only class variables is contained in another effect if the second effect contains more class variables and contains the class variables in the first effect (for example, A is contained in A*B, if A and B are classification variables). The result of nested subject effects in PROC MIXED or PROC GLIMMIX is that the procedure can process the data by subjects. The subjects for processing are determined by the SUBJECT= effect contained in all others, in this case, the BLK effect. The “Dimensions” table in the MIXED or GLIMMIX output informs you how many subjects there are for data processing. If you replace the RANDOM statement

```
random int person / subject=blk;
```

in the code above with the equivalent

```
random blk person(blk);
```

there will not be an overall subject for the analysis and estimation will be noticeably slower.

The results appear in Output 16.31.

Results

Output 16.31 Results of Fitting the Autoregressive Time Errors and Independent Period Errors Model to the Acceptability Data Using the PRIORPRD Class Variable

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
blk		0.5042
person(blk)		0.3265
Variance	person(blk)	0.1260
AR(1)	person(blk)	-0.1361
AR(1)	person*period(blk)	0.2098
Residual		0.1164

Fit Statistics	
-2 Res Log Likelihood	309.8
AIC (smaller is better)	321.8
AICC (smaller is better)	322.2
BIC (smaller is better)	318.1

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	33.3	3.97	0.0161
period	1	31.3	0.04	0.8391
time	2	91.6	24.77	<.0001
prod*time	6	108	12.73	<.0001
priorprd	3	42.7	3.91	0.0148

Interpretation

The variance component part of the model now has two parameters for the PERIOD part of the error structure (that is, the components identified with PERSON(BLK) in the Subject column): AR(1) for the correlation coefficient ($\hat{\rho}_a = -0.1361$) and PERIOD diagonal for the variance ($\hat{\rho}_e = 0.1260$). The “Fit Statistics” table indicates that the AR(1) covariance structure for PERIOD does not improve the fit of the model relative to either of the previous models—for example, AICC=322.2 compared with 319.2 for the independent errors model and 320.2 for the AR(1) errors only model. As a consequence, the “Type 3 Tests of Fixed Effects” table is almost the same as for the previous two models.

Program—Replace PRIORCND with Li's to Estimate LSMEANS

The PROC GLIMMIX program replacing PRIORCND Li's to provide LSMEANS and simple effect estimates using SLICEDIFF is as follows:

```
proc glimmix data=prior;
  class blk person seq prod period time priorprd;
  model y = prod period time time*prod L1 L2 L3 L4/ddfm=kr;
  random int person / subject=blk;
  random period / subject=person(blk) type=ar(1);
  random time / subject=period*person*blk type=ar(1)
    residual;
  lsmeans time*prod / slicediff=(time prod);
run;
```

As in the previous PROC MIXED code we can take advantage of nested subject effects in specifying the model. The REPEATED statement in the PROC MIXED code specified the TIME effect after the REPEATED statement. This can be necessary, if, for example, the data contain missing values and the observation sequence numbers within subject would not provide the correct level for determining the AR(1) covariances, or, if the input data is not properly arranged. To replace the MIXED statement

```
repeated time / subject=period*person(blk) type=ar(1);
```

in PROC GLIMMIX use

```
random time / subject=period*person*blk type=ar(1)
  residual;
```

All covariance structures are specified in the RANDOM statement in PROC GLIMMIX, regardless of whether they apply to the G matrix or the R matrix. If it is necessary to specify an R-side effect for levelization, add the RESIDUAL option to the RANDOM statement to indicate an R-side random effect.

Results

The estimates of the variance components and the model-fitting information are the same as the previous model with the PRIORPRD class variable. The tests for the fixed effects part of the model are shown in Output 16.32 and the LSMEANS are shown in Output 16.33.

Output 16.32 Type III F-Statistics for Acceptability Data Fit with the Fixed Effects from AR(1) Errors and AR(1) PERIOD Effect Model Using the Li's

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
prod	3	33.26	3.97	0.0161
period	1	31.29	0.04	0.8391
time	2	91.65	24.77	<.0001
prod*time	6	107.6	12.73	<.0001
l1	0	.	.	.
l2	0	.	.	.
l3	0	.	.	.
l4	0	.	.	.

As with the independent and AR(1) error only models, these results reproduce the same tests for PROD, PERIOD, TIME, and PROD \times TIME as the program using PRIORPRD in place of the Li's. The Li's are zero because of the collinearity between the Li's and the levels of PERIOD.

Output 16.33 Least-Squares Means and Selected Simple Effect Comparisons from Fitting the Autoregressive Errors and Autoregressive Period Effects Model to the Acceptability Data

prod*time Least Squares Means						
prod	time	Estimate	Standard Error	DF	t Value	Pr > t
1	1	3.9112	0.3965	3.57	9.86	0.0010
1	2	3.6334	0.3965	3.57	9.16	0.0013
1	3	2.8556	0.3965	3.57	7.20	0.0030
2	1	4.0370	0.3965	3.57	10.18	0.0009
2	2	3.7037	0.3965	3.57	9.34	0.0012
2	3	3.4814	0.3965	3.57	8.78	0.0015
3	1	3.8050	0.3965	3.57	9.60	0.0011
3	2	3.6939	0.3965	3.57	9.32	0.0012
3	3	3.6383	0.3965	3.57	9.18	0.0013
4	1	3.9135	0.3965	3.57	9.87	0.0010
4	2	4.1913	0.3965	3.57	10.57	0.0008
4	3	4.1357	0.3965	3.57	10.43	0.0008

Simple Effect Comparisons of prod*time Least Squares Means By time							
Simple Effect Level	prod	_prod	Estimate	Standard Error	DF	t Value	Pr > t
time 1	1	2	-0.1258	0.1994	52.26	-0.63	0.5309
time 1	1	3	0.1062	0.1994	52.26	0.53	0.5964
time 1	1	4	-0.00230	0.1994	52.26	-0.01	0.9909
time 1	2	3	0.2320	0.1994	52.26	1.16	0.2498
time 1	2	4	0.1235	0.1994	52.26	0.62	0.5384
time 1	3	4	-0.1085	0.1994	52.26	-0.54	0.5885
time 2	1	2	-0.07023	0.1994	52.26	-0.35	0.7261
time 2	1	3	-0.06043	0.1994	52.26	-0.30	0.7630
time 2	1	4	-0.5579	0.1994	52.26	-2.80	0.0072
time 2	2	3	0.009806	0.1994	52.26	0.05	0.9610
time 2	2	4	-0.4876	0.1994	52.26	-2.45	0.0179
time 2	3	4	-0.4974	0.1994	52.26	-2.49	0.0158
time 3	1	2	-0.6258	0.1994	52.26	-3.14	0.0028
time 3	1	3	-0.7827	0.1994	52.26	-3.93	0.0003
time 3	1	4	-1.2801	0.1994	52.26	-6.42	<.0001
time 3	2	3	-0.1569	0.1994	52.26	-0.79	0.4350
time 3	2	4	-0.6543	0.1994	52.26	-3.28	0.0018
time 3	3	4	-0.4974	0.1994	52.26	-2.49	0.0158

Simple Effect Comparisons of prod*time Least Squares Means By prod							
Simple Effect Level	time	_time	Estimate	Standard Error	DF	t Value	Pr > t
prod 1	1	2	0.2778	0.1037	119.2	2.68	0.0084
prod 1	1	3	1.0556	0.1145	69.28	9.22	<.0001
prod 1	2	3	0.7778	0.1037	119.2	7.50	<.0001

Interpretation

The values of the least-squares means are a little different from those in the two previous analyses. The major difference is that the standard errors of the difference between pairs of PROD means at the same level of TIME are a little larger than for the other two models (the standard error of a difference is 0.1994 compared with 0.1895 for the independent errors model in Section 16.10.3 and 0.1900 for the AR(1) errors model in Section 16.10.4). The standard errors of the difference between two TIME means for the same level of PROD are slightly larger than those of the previous models. The standard errors for differences one time period apart (TIME 1 versus 2 or 2 versus 3) are 0.1037 (compared with 0.1036 for the AR(1) error model). For a difference two times periods apart (TIME 1 versus 3) the standard error is 0.1145 compared with 0.1140 for the AR(1) error model.

16.10.6 Conclusions

The acceptability data were analyzed using three specifications of the error structure, and the results were compared. The major differences in the analyses are in the magnitudes of the Type III F -values and in the estimated standard errors. The values of the least-squares means were affected least. The likelihood ratio tests and the other fit criteria (e.g., AICC) for comparing the models indicate that the independent errors model is adequate to describe the data. Either PROC GLIMMIX or PROC MIXED can be used. PROC GLIMMIX is more convenient for estimating simple effects given the presence of time \times product interaction.

16.11 Random Coefficients Modeling of an AIDS Trial

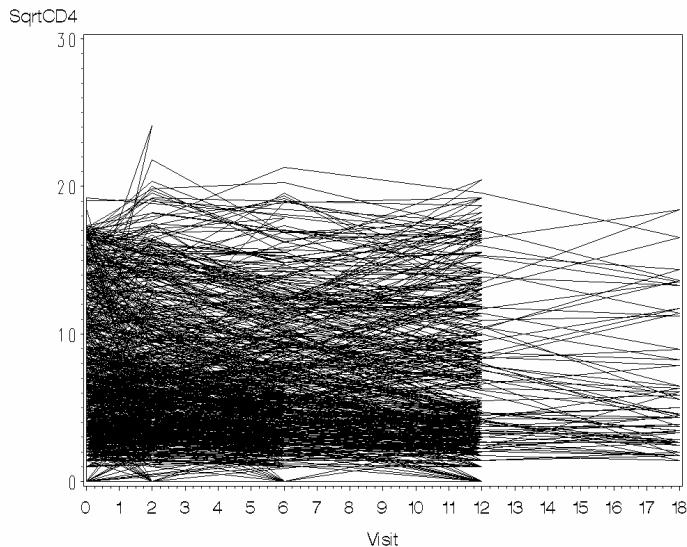
16.11.1 Introduction

The data for this example are from a clinical trial for AIDS patients comparing two drugs, ddI and ddC (Abrams, Goldman, Launer, et al. 1994, Fleming, Neaton, Goldman, et al. 1995; see also Lange, Carlin, and Gelfand 1992). The study was a multicenter, randomized, open-label comparative trial of ddI to ddC in HIV-infected patients who were intolerant to or had failed ZDV therapy. Participants were men and women, aged 15–67, who had either advanced to AIDS or had a CD4 lymphocyte count less than or equal to 300. The primary objective of the study was to determine which monotherapy was better in this patient population. Some key features of the data are as follows:

- repeated measures on CD4 counts and several explanatory variables
- 467 subjects at 5 visits: baseline, 2, 6, 12, and 18 months, with many missing values
- values well below the normal range of 22–31 on the square root scale

Figure 16.10 provides profile plots of the longitudinal measurements.

Figure 16.10 CD4 Data



Scientific theory suggests an initial increase in the CD4 counts over the first few months and then a linear decline (on the square root scale) thereafter. The population mean of the responses is therefore modeled as a piecewise linear model with a join point at 2 months. This mean model is a “hockey stick” because it has one short line and one long one. The objectives for this case study analysis are as follows:

1. Fit a random coefficients model using a “hockey stick” population curve for SQRTCD4 with baseline as a dependent variable.
2. Compute estimates and standard errors for various linear combinations.
3. Compute and plot predicted values.

The first objective involves using the baseline measurement as a dependent variable; however, you can also use baseline as a covariate. Some pros and cons of each approach are shown in Table 16.9.

Table 16.9 Pros and Cons of Using Baseline as a Dependent Variable or as a Covariate

Baseline as Dependent	Baseline as Covariate
Models baseline to first measurement	Starts with first measurement
Unconditional prediction	Conditional prediction
Must account for no treatment effect at baseline	Treatment effect straightforward
Account for covariance structure at baseline	Covariance structure easier to model
Type III tests may not be meaningful	Type III tests straightforward

These characteristics indicate that fitting baseline as a dependent variable can be more challenging than fitting it as a covariate from both a modeling and an interpretation perspective. Indeed, fitting baseline as a covariate is accomplished by just reformulating the original SAS data set so that each subject has one fewer observation and the baseline measurement is a new variable and then including this new variable in the MODEL statement as an additional fixed effect. Instead, this case study focuses on techniques for handling the baseline measurement as a dependent variable.

The use of the standard random coefficients model (see Chapter 7) for these data relies on the assumption that the data are Gaussian and missing at random. Although these assumptions could be violated here, they are made for illustration purposes. A likelihood approach to fitting the random coefficients model nicely handles unequal spacing and random missingness. It also provides a reasonable mechanism for prediction beyond the range of the data.

16.11.2 Analysis Using PROC MIXED

The SAS data set CD4 is shown as Data Set 16.11, “CD4 Count,” in Appendix 2, “Data Sets.” The SAS data set for analysis is generated with the DATA step as follows:

```
data cd4;
  input randgrp stratum unit @@;
  seq = _n_;
  do visit = 0,2,6,12,18;
    input cd4 @@;
```

```

sqrtcd4 = sqrt(cd4);
if visit ne 0 then notbase = 1;
else notbase = 0;
if visit ne 0 then visitm2 = visit - 2;
else visitm2 = 0;
output;
end;
datalines;
2   2   2   114      .    71    89    .
1   2   2   40       .    66    21    25
1   2   3   12       13   38    .     .
2   1   3   15       21   7     3     .
1   1   3   53       74   74    45    .
2   1   3   21       .     .     .     .
<more datalines>

```

The key variables are as follows:

SEQ	a 3-digit ID for each subject (1–467)
UNIT	the clinical center (2–18)
RANDGRP	the treatment group (1=ddI, 2=ddC)
STRATUM	a time-invariant variable used to stratify the subjects (1=ZDV failure, 2=ZDV intolerant)
CD4	the CD4 count
VISIT	the month at which the CD4 count was measured (0=baseline, 2=2 months, 6=6 months, 12=12 months, 18=18 months)
SQRTCD4	the square root of the CD4 count
NOTBASE	a dummy variable indicating if the measurement is baseline (0=baseline, 1=not baseline)
VISITM2	equals max(0,VISIT-2)

Program—Model Assuming Homogeneous Residual Variance Group

Use the following PROC MIXED program to estimate the initial random coefficients model:

```

proc mixed data=cd4;
  class randgrp stratum unit seq;
  model sqrtcd4 = stratum unit notbase
    notbase*randgrp notbase*stratum
    visitm2 visitm2*randgrp visitm2*stratum
    / ddfm=kr;
  random int notbase visitm2 / type=un sub=seq g gcorr;
run;

```

The MODEL statement specifies the mean model for the data. STRATUM and UNIT are both considered fixed main effects, but note that the treatment variable RANDGRP is included in interaction terms only to account for the fact that it does not affect the baseline measurements.

NOTBASE and VISITM2 are the variables used to specify the piecewise linear “hockey stick” mean model. NOTBASE is effectively a new intercept starting at 2 months, and VISITM2 models a linear slope from there. Both NOTBASE and VISITM2 interact with RANDGRP and STRATUM to provide different “hockey sticks” for these classifications. As with all models involving correlated errors, the DDFM=KR option is recommended.

The RANDOM statement models three random coefficients using the following three effects: INT (an overall intercept), NOTBASE (a random jump at 2 months), and VISITM2 (a random slope starting at 2 months). Thus, each subject has a random starting point for its hockey stick, a random location for the bend in the stick, and a random angle for the stick. TYPE=UN models a full 3×3 covariance matrix for the random coefficients, and SUB=SEQ identifies the subjects. The G and GCORR options request a printout of the G matrix and its correlations.

In general mixed model notation, \mathbf{X} is 1408×29 (over-parameterized form). \mathbf{Z} is 1408×1401 and is block diagonal with 467 blocks (one per subject) having dimensions $n_i \times 3$, where n_i is the number of observations for the i^{th} subject. \mathbf{G} is 1401×1401 and is block diagonal with 467 unstructured blocks, each of dimension 3×3 , $\mathbf{R} = \sigma^2 \mathbf{I}_{1408}$.

Output 16.34 shows relevant results for this initial model.

Results

Output 16.34 Results for “Hockey Stick” Random Coefficients Model

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	seq	17.7568
UN(2,1)	seq	1.5132
UN(2,2)	seq	0.2508
UN(3,1)	seq	-0.3360
UN(3,2)	seq	0.07111
UN(3,3)	seq	0.02800
Residual		2.9952

Estimated G Matrix					
Row	Effect	seq	Col1	Col2	Col3
1	Intercept	1	17.7568	1.5132	-0.3360
2	notbase	1	1.5132	0.2508	0.07111
3	visitm2	1	-0.3360	0.07111	0.02800

Estimated G Correlation Matrix					
Row	Effect	seq	Col1	Col2	Col3
1	Intercept	1	1.0000	0.7171	-0.4765
2	notbase	1	0.7171	1.0000	0.8485
3	visitm2	1	-0.4765	0.8485	1.0000

Fit Statistics		
-2 Res Log Likelihood		7032.8
AIC (smaller is better)		7046.8
AICC (smaller is better)		7046.9
BIC (smaller is better)		7075.8

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
stratum	1	451	32.00	<.0001
unit	15	450	1.50	0.1006
notbase	1	381	4.31	0.0386
notbase*randgrp	1	378	7.35	0.0070
notbase*stratum	1	382	4.21	0.0409
visitm2	1	193	62.41	<.0001
visitm2*randgrp	1	180	0.61	0.4352
visitm2*stratum	1	194	0.24	0.6245

Interpretation

The REML algorithm converges successfully after four iterations. The “Covariance Parameter Estimates” table indicates that there is some additional variability of the 2-month random effect beyond that modeled by the random intercept. The variance estimate, denoted UN(2,2), of this 2-month jump equals 0.2508. This estimate appears in the “Estimated G Matrix” output as the diagonal element for NOTBASE (row 2, column 2). The “Estimated G Correlation Matrix” table reveals that the 2-month jump is positively correlated with the random intercept (0.717) and that the random slope is negatively correlated with the random intercept (-0.477) and positively correlated with the 2-month jump (0.849).

The residual variance estimate is 2.995. The “Fit Statistics” table includes the “-2 Res Log Likelihood,” equal to 7032.8, and the AICC criterion, equal to 7046.9. These can be used to evaluate the fit of the covariance model relative to competing covariance structures.

The “Type 3 Tests of Fixed Effects” table reveals some evidence for significant treatment effect (the NOTBASE × RANDGRP effect has a *p*-value of 0.0070). Also, the UNIT effect, with a *p*-value of 0.1006, and the interactions involving VISITM2, with *p*-values greater than 0.43, do not appear to contribute much to the model. Before deleting them from the model, note that the model above assumes homogeneous residual variance. This assumption may not be justified for these data and the resulting model misspecification may mask significant UNIT effects or interactions involving VISITM2. To assess this possibility, consider an enhanced model.

Program—Heterogeneous Residual Variance Groups

The enhancement involves modeling heterogeneous residual variance groups, a topic discussed further in Chapter 9. The PROC MIXED program for this model is as follows:

```
proc mixed data=cd4;
  class randgrp stratum unit seq visit;
  model sqrtcd4 = stratum unit notbase
    notbase*randgrp notbase*stratum
    visitm2 visitm2*randgrp visitm2*stratum
    / ddfm=kr;
  random int notbase visitm2 / type=un sub=seq;
  repeated visit / sub=seq group=randgrp*stratum;
run;
```

Note the addition of the REPEATED statement, whose purpose is to model **R** with four different variance components down its diagonal via the GROUP= option. These variance components correspond to the different combinations of treatment and stratum.

Abbreviated output from this model is displayed in Output 16.35.

Results

Output 16.35 Results for “Hockey Stick” Heterogeneous Model

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	seq		18.8029
UN(2,1)	seq		0.3585
UN(2,2)	seq		1.0135
UN(3,1)	seq		-0.3564
UN(3,2)	seq		0.08717
UN(3,3)	seq		0.02956
visit	seq	randgrp*stratum 1 1	1.1308
visit	seq	randgrp*stratum 1 2	4.0471
visit	seq	randgrp*stratum 2 1	1.8942
visit	seq	randgrp*stratum 2 2	2.7210

Fit Statistics	
-2 Res Log Likelihood	6983.1
AIC (smaller is better)	7003.1
AICC (smaller is better)	7003.3
BIC (smaller is better)	7044.6

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
stratum	1	436	31.60	<.0001
unit	15	445	1.54	0.0874
notbase	1	327	5.11	0.0245
notbase*randgrp	1	343	7.52	0.0064
notbase*stratum	1	362	5.03	0.0254
visitm2	1	158	75.67	<.0001
visitm2*randgrp	1	178	0.47	0.4953
visitm2*stratum	1	175	0.22	0.6429

Interpretation

The enhanced residual variance model results in a substantially better fit to the data. You can document this improved fit either with a likelihood ratio test or by comparing information criteria, e.g., the AICC. For the homogeneous model, the AICC is 7046.9 versus 7003.3 for the heterogeneous model. Smaller is better. For the likelihood ratio test, the χ^2 statistic is the difference in “–2 Res Log Likelihood” for the two models, $7032.8 - 6986.2 = 46.6$, with 3 degrees of freedom. From the “Covariance Parameter Estimates” table you can see definite differences in the residual variance component estimates for the four groups. They range from 1.131 to 4.047, with both estimates for stratum 2 smaller than the average value of 2.99 obtained in the previous model. The fixed effects tests change somewhat as a result of this different variance-covariance model. Notably, the UNIT effect has a p -value of 0.0874, versus a p -value greater than 0.10 in the homogeneous residual variance model. Depending on the α -level, this could obviously affect your conclusions.

The final steps of the analysis are as follows: (1) fit a reduced model with nonsignificant effects deleted; (2) estimate and test any terms of interest; and (3) compute and plot predicted values.

Program—Fit the Final Reduced “Hockey Stick” Model

Assuming that the main effect of UNIT and the interactions with VISITM2 are nonsignificant, the final reduced “hockey stick” heterogeneous model, along with some relevant inferential statements, is as follows:

```
proc mixed data=cd4;
  class randgrp stratum visit seq;
  model sqrtcd4 = stratum notbase notbase*randgrp
    notbase*stratum visitm2 /solution ddfm=kr;
  random int notbase visitm2 / type=un sub=seq g gcorr;
  repeated visit / sub=seq group=randgrp*stratum;
  estimate 'ddI 2 - base s 2'
    notbase 1 notbase*randgrp 1 0 visitm2 0 / cl;
  estimate 'ddC 2 - base s 2'
    notbase 1 notbase*randgrp 0 1 visitm2 0;
  estimate 'ddI 6 - base s 2'
    notbase 1 notbase*randgrp 1 0 visitm2 4;
  estimate 'ddC 6 - base s 2'
    notbase 1 notbase*randgrp 0 1 visitm2 4;
  estimate 'ddI - ddC s 2' notbase*randgrp 1 -1;
  estimate 'ddI 24 s 1' int 1 stratum 1 0 notbase 1
    notbase*randgrp 1 0 notbase*stratum 1 0 visitm2 22;
  estimate 'ddC 24 s 1' int 1 stratum 1 0 notbase 1
    notbase*randgrp 0 1 notbase*stratum 1 0 visitm2 22;
  estimate 'ddI 24 s 2' int 1 notbase 1 notbase*randgrp 1 0
    visitm2 22;
  estimate 'ddC 24 s 2' int 1 notbase 1 notbase*randgrp 0 1
    visitm2 22;
run;
```

The SOLUTION option in the MODEL statement requests a table of the fixed effects parameter estimates. The DDFM=KR option requests the computation of approximate degrees of freedom and bias-corrected standard errors and test statistics.

The ESTIMATE statements define various linear combinations of the fixed effects parameter estimates, also known as estimable functions, chosen for potential interest in the context of the research. The first two construct differences between baseline and the 2-month visit for each treatment for the second stratum. The next two ESTIMATE statements construct differences between baseline and the 6-month visit for each treatment for the second stratum. The fifth

ESTIMATE statement computes the overall treatment difference for the second stratum. The final four statements compute mean estimates at 24 months. The CL option in the first ESTIMATE statement requests confidence limits for all of the estimates.

Partial output from this model is shown in Output 16.36.

Results

Output 16.36 Results for Reduced “Hockey Stick” Heterogeneous Model

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	seq		18.7283
UN(2,1)	seq		0.5173
UN(2,2)	seq		0.9961
UN(3,1)	seq		-0.3272
UN(3,2)	seq		0.09241
UN(3,3)	seq		0.02885
visit	seq	randgrp*stratum 1 1	1.1246
visit	seq	randgrp*stratum 1 2	4.0547
visit	seq	randgrp*stratum 2 1	1.9319
visit	seq	randgrp*stratum 2 2	2.7279

Estimated G Matrix					
Row	Effect	seq	Col1	Col2	Col3
1	Intercept	1	18.7283	0.5173	-0.3272
2	notbase	1	0.5173	0.9961	0.09241
3	visitm2	1	-0.3272	0.09241	0.02885

Estimated G Correlation Matrix					
Row	Effect	seq	Col1	Col2	Col3
1	Intercept	1	1.0000	0.1198	-0.4451
2	notbase	1	0.1198	1.0000	0.5452
3	visitm2	1	-0.4451	0.5452	1.0000

Solution for Fixed Effects							
Effect	randgrp	stratum	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			8.0589	0.2734	497	29.48	<.0001
stratum		1	-2.4777	0.4300	447	-5.76	<.0001
stratum		2	0
notbase			-0.3077	0.1913	286	-1.61	0.1088
notbase*randgrp	1		0.5904	0.2245	362	2.63	0.0089
notbase*randgrp	2		0

Solution for Fixed Effects							
Effect	randgrp	stratum	Estimate	Standard Error	DF	t Value	Pr > t
notbase*stratum		1	-0.5027	0.2302	375	-2.18	0.0296
notbase*stratum		2	0
visitm2			-0.1511	0.01694	183	-8.92	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
stratum	1	447	33.20	<.0001
notbase	1	320	5.11	0.0245
notbase*randgrp	1	362	6.92	0.0089
notbase*stratum	1	375	4.77	0.0296
visitm2	1	183	79.50	<.0001

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
ddI 2 base s 2	0.03144	0.1634	251	0.19	0.8475	0.05	-0.2903	0.3532
ddC 2 base s 2	-0.5590	0.1605	289	-3.48	0.0006	0.05	-0.8749	-0.2431
ddI 6 base s 2	-0.5729	0.1676	325	-3.42	0.0007	0.05	-0.9026	-0.2431
ddC 6 base s 2	-1.1633	0.1632	316	-7.13	<.0001	0.05	-1.4843	-0.8422
ddI ddC s 2	0.5904	0.2245	362	2.63	0.0089	0.05	0.1489	1.0320
ddI 24 s 1	2.0376	0.4829	417	4.22	<.0001	0.05	1.0884	2.9868
ddC 24 s 1	1.4472	0.4789	382	3.02	0.0027	0.05	0.5056	2.3888
ddI 24 s 2	3.5278	0.3960	276	8.91	<.0001	0.05	2.7483	4.3073
ddC 24 s 2	2.9374	0.3907	246	7.52	<.0001	0.05	2.1679	3.7069

Interpretation

The variance of the 2-month jump is positive again, but still small as compared to the variance of the random intercept. Also, the correlations involving the 2-month jump have dropped considerably from the initial analysis in Output 16.34. The four combinations of treatment and stratum continue to exhibit heterogeneity in their residual variances.

The “Solution for Fixed Effects” table lists estimates of the fixed effects parameters. The order and form of these estimates are important for formulating the ESTIMATE statements shown previously. Specifically, coefficients in the ESTIMATE statement must correspond to the order in which the parameter estimates appear in the “SolutionF” output.

From the “Type 3 Tests of Fixed Effects” we see that all fixed effects have *p*-values less than 0.05, and the *p*-value for treatments (NOTBASE × RANDGRP) equals 0.0089.

The first two rows of the “Estimates” table reveals that in stratum 2, the ddI treatment exhibits a slight increase from baseline at 2 months (0.03), which is not significantly different from zero, but the ddC treatment exhibits a significant decrease from baseline (-0.56). The corresponding 6-month estimates are both significantly below baseline (-0.57 and -1.16 for ddI and ddC, respectively). The “ddI ddC s 2” estimate has the same *p*-value as the NOTBASE × RANGRP test; both are testing the same null hypothesis of no difference between treatments. The estimate reveals that ddC patients have significantly lower counts than ddI patients. The estimated difference is 0.59 with a standard error of 0.22. The 24-month mean predictions are on the same order of magnitude, with the ddI patients who are ZDV intolerant having the largest predicted mean at 24 months.

Finally, you can plot the predicted values with the following program.

Program—Plot Predicted Values

Before computing the predicted values, some additional code is required.

```
data cd4x;
  input seq randgrp stratum sqrtcd4 visit notbase;
  if (visit ne 0) then visitm2 = visit - 2;
  else visitm2 = 0;
  datalines;
1000 1 1 . 0 0
1000 1 1 . 2 1
1000 1 1 . 6 1
1000 1 1 . 12 1
1000 1 1 . 18 1
1001 1 2 . 0 0
1001 1 2 . 2 1
1001 1 2 . 6 1
1001 1 2 . 12 1
1001 1 2 . 18 1
1002 2 1 . 0 0
1002 2 1 . 2 1
1002 2 1 . 6 1
1002 2 1 . 12 1
1002 2 1 . 18 1
1003 2 2 . 0 0
1003 2 2 . 2 1
1003 2 2 . 6 1
1003 2 2 . 12 1
1003 2 2 . 18 1
;
proc append base=cd4 data=cd4x;
run;
```

The resulting new CD4 data set allows population predictions to be constructed across the range of the data. Once you do this, use the following program to create the data set of predicted values:

```
proc mixed data=cd4;
  class randgrp stratum visit seq;
  model sqrtcd4 = stratum
    notbase notbase*randgrp notbase*stratum
    visitm2 / ddfm=res outp=p;
  random int notbase visitm2 / type=un sub=seq g gcorr;
  repeated visit / sub=seq group=randgrp*stratum;
  id visit seq;
run;
```

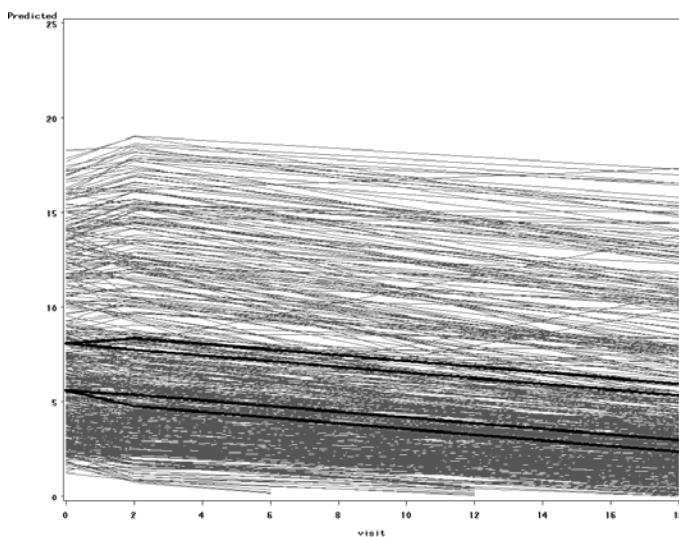
The OUTP=P option in the MODEL statement causes the creation of a new SAS data set, called "P," containing predicted values. The DDFM=RES option replaces DDFM=KR. You should not use the Kenward-Roger option while computing predicted values for this model because the covariance structure is complex. Depending on your hardware, you may run out of memory or simply experience excessive computing time as PROC MIXED attempts to compute bias-corrected standard errors and approximate degrees of freedom for all predicted values. The KR adjustments are needed for the statistics in Output 16.36, but not for the predicted values.

Once the predicted values are computed, the following statements allow you to plot the predicted values using PROC GPLOT:

```
symbol1 i=join c=green w=1 r=467;
symbol2 i=join c=black w=4 r=4;
proc gplot data=p;
  plot pred*visit=seq / nolegend
    vaxis=0 to 25 by 5 vminor=0
    haxis=0 to 18 by 2 hminor=0;
run;
```

Figure 16.11 shows these predictions, with the population "hockey sticks" in bold. Note that predictions are available for all subjects at all time periods, even if a subject has only a few observations. For example, the sixth subject has only a baseline measurement, but a full, predicted, individual "hockey stick" is available based on that single measurement.

The predicted profiles are all essentially a weighted average between the original data profiles and the appropriate population curve, with the weights determined by the method of empirical Bayes. The predicted profiles for most subjects follow the basic pattern of the population curves. Several curves drop below zero and are truncated in Figure 16.11. On the other hand, a few exceptional subjects deviate from the trend, exhibiting significant growth and good recovery potential.

Figure 16.11 Predicted Values for CD4 Data

16.12 Microarray Example

16.12.1 Introduction

Microarrays are widely used in molecular genetics to evaluate gene response to various treatments. Typical microarrays are slides containing material representing anywhere from several hundred to tens of thousands of genes. Mixed model methods are among the statistical tools used to analyze microarray data. Several excellent texts are available to help the intermediate to advanced reader implement microarray analysis (for example, Allison et al. 2006). This case study presents a highly simplified example with the basic structure and mix model issues involved in microarray data. Its purpose is to introduce the subject to readers who are either statisticians with little or no previous experience with microarrays, or molecular biologists who are familiar with microarrays but have little experience with mixed model analysis. The setup uses simulated data, but shares the basic structure and experimental design with common microarray data sets.

Microarrays typically have a row-column grid of spots containing genetic material from g genes. The exact genetic material depends on the type of microarray, but is commonly single-stranded segments of DNA from genes of interest. (The lengths of these segments can vary from oligonucleotides of length 25 or 60–70 to strands that span the full length of the gene coding region.) Most microarrays in common use can be classified as one-channel or two-channel, depending upon whether one or two measurements are obtained from each spot. This example considers two-channel arrays, which are used to simultaneously measure gene expression from two treated samples.

The samples require careful laboratory preparation before they are applied to the microarrays. Preparation typically consists of extraction of RNA from the samples and then utilization of a reverse transcription reaction that incorporates a fluorescent tag, typically green (Cy3) and red (Cy5). The resulting labeled complementary DNAs (cDNAs) in the samples are “genetic mirror images” of single-stranded DNAs already affixed to the microarray. The samples are mixed and exposed to the microarray to allow the cDNAs to bind to their matches in a process called

hybridization. A laser scanner is then used to “read” the spots at two different frequencies corresponding to the fluorescence of Cy3 and Cy5. Those spots that respond with greater color intensity (green or red, depending on the treatment) indicate genes showing a response (or “gene expression”) to that treatment. The log₂ intensity of the resulting pixel image is typically used as the response variable. Two such intensities are obtained from each spot, corresponding to the two treatments.

Replication in gene expression experiments is obtained in two different ways. First, cDNA can be placed in more than one spot on a given slide. Craig et al. (2001) describe a process using multiple “pins” (a multi-tip pipette that takes cDNA from a source plate and inserts it into the spots) and “dips” (multiple uses of each pin to complete the placement of cDNA on the slide). Second, replication can be achieved by using more than one slide.

A common design for replicate slices, used for experiments with three or more treatments, is a cyclic incomplete block design, known in microarray research as a loop design. This example contains six treatments and six chips using a loop design, described by Table 16.10.

Table 16.10 Loop Design for Microarray Experiment with Six Treatments and Two Dyes

Chip	Green Dye	Red Dye
1	Trt 1	Trt 2
2	Trt 2	Trt 3
3	Trt 3	Trt 4
4	Trt 4	Trt 5
5	Trt 5	Trt 6
6	Trt 6	Trt 1

In this example, there are $g=8$ genes. Each slide is an 8×4 grid of spots, with cDNA from each gene appearing once in each column. For each pair of columns, two pins with two dips each are used to place the cDNA. See Craig et al. (2001) for a more detailed description of the design approach. The data appear as Data Set 16.12, “Microarray,” in Appendix 2, “Data Sets.”

A linear mixed model to describe data from this design is

$$Y_{ijkmn} = \mu + \lambda_i + \tau_j + \delta_k + (\tau\lambda)_{ij} + (\delta\lambda)_{ik} + a_m + (a\lambda)_{im} + d(a)_{mn} + p_r + (ap)_{mr} + e_{ijkmn}$$

where

Y_{ijkmn} is the log intensity of the i^{th} gene, j^{th} treatment, and the k^{th} dye observed on the r^{th} pin and n^{th} dip within the m^{th} array (or slide)

μ is the intercept

λ_i is the i^{th} gene effect

τ_j is the j^{th} treatment effect

δ_k is the k^{th} dye effect

$(\tau\lambda)_{ij}$ and $(\delta\lambda)_{ik}$ are treatment \times gene and dye \times gene interaction

a_m is the m^{th} array (or slide) effect, assumed $iid N(0, \sigma_A^2)$

$(a\lambda)_{im}$ is the chip \times gene effect, assumed $iid N(0, \sigma_{GA}^2)$

$d(a)_{mn}$ is the effect of the n^{th} dip within the m^{th} array (slide), assumed $iid N(0, \sigma_D^2)$

p_r is the effect of the r^{th} pin, assumed $iid N(0, \sigma_P^2)$

$(ap)_{mr}$ is the array \times pin interaction effect, assumed $iid N(0, \sigma_{AP}^2)$

e_{ijkmnr} is random error, assumed $iid N(0, \sigma^2)$

16.12.2 Analysis with PROC GLIMMIX

You can use PROC MIXED or PROC GLIMMIX to analyze these data using the model above. To take advantage of features that make analysis of gene-specific treatment effects and treatment-specific gene expression more convenient, we use PROC GLIMMIX. Use the following statements.

Program

```
proc glimmix data=microarray;
  class array dye trt gene pin dip;
  model log_intensity = dye trt gene dye*gene
    trt*gene pin / ddfm=kr;
  random array array*gene dip(array) pin*array;
  lsmeans trt*gene / slice=gene
    slicediff=gene
    slicedifftype=control('0' '2');
run;
```

The SLICEDIFF option allows you to look at gene-specific treatment effects. The SLICEDIFFTYPE option allow you to define the “0” level of treatment as a control or reference treatment, restricting simple effects for each gene to differences between treatment “0” and each other treatment. Note that you must also give a reference level for GENE as well, because the least-squares means are defined on TRT \times GENE even though SLICEDIFF by TRT is not requested here. If desired, you could add TRT to the SLICEDIFF request.

Output 16.37 shows relevant analysis for these data.

Results

Output 16.37 Selected PROC GLIMMIX Output for Microarray Data

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
array	0.01361	0.01591
array*gene	0.001685	0.001615
dip(array)	0.003328	0.001912

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
array*pin	0.03742	0.01487
Residual	0.01953	0.001702

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
dye	1	263.3	43.98	<.0001
trt	5	267	1.84	0.1053
gene	7	16.78	8.19	0.0002
dye*gene	7	263.3	0.66	0.7071
trt*gene	35	125	1.55	0.0420
pin	3	14.77	0.74	0.5462

Tests of Effect Slices for trt*gene Sliced By gene				
gene	Num DF	Den DF	F Value	Pr > F
1	5	175.1	0.33	0.8962
2	5	182	3.82	0.0026
3	5	180.7	0.48	0.7874
4	5	181.5	2.23	0.0532
5	5	182.1	0.67	0.6461
6	5	179.9	3.29	0.0072
7	5	179.6	1.62	0.1574
8	5	180	0.42	0.8319

Simple Effect Comparisons of trt*gene Least Squares Means By gene							
Simple Effect Level	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
gene 2	1	0	0.1299	0.07843	251	1.66	0.0988
gene 2	2	0	0.09961	0.08804	135.8	1.13	0.2599
gene 2	3	0	0.1830	0.08988	121.2	2.04	0.0440
gene 2	4	0	0.2943	0.08817	140.7	3.34	0.0011
gene 2	5	0	0.3143	0.07858	253.1	4.00	<.0001
gene 3	1	0	-0.06433	0.07819	248.5	-0.82	0.4114
gene 6	1	0	0.005857	0.07841	251.1	0.07	0.9405
gene 6	2	0	0.1379	0.08798	134.2	1.57	0.1194

Simple Effect Comparisons of trt*gene Least Squares Means By gene							
Simple Effect Level	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
gene 6	3	0	0.1634	0.08992	121.9	1.82	0.0717
gene 6	4	0	0.2108	0.08794	134	2.40	0.0179
gene 6	5	0	0.2884	0.07835	249.4	3.68	0.0003

Interpretation

The “Covariance Parameter Estimates” table indicates the magnitude of variation associated with the random effects in the model. The largest sources of random variation are ARRAY \times PIN, ARRAY, and RESIDUAL. The primary test of interest among the “Type III Tests of Fixed Effects” is TRT \times GENE. This tests the null hypothesis of no differential gene expression by treatment (or, alternatively, no difference in treatment effect by gene). You can see that the *p*-value is 0.0420, suggesting some differential gene expression. Looking at the SLICE results, you can see that genes 2 and 6, with *p*-values of 0.0026 and 0.0072, respectively, are the genes affected by treatments. The SLICEDIFF results shown here are restricted to those two genes. For each gene, it appears that differences in gene expression are between treatment 4 and 5 relative to the control (treatment 0), and to a lesser extent treatment 3 versus the control.

Note: For typical microarrays with thousands of genes, the preceding model is not computationally feasible because of the large size of the \mathbf{X} matrix. In this situation, we recommend breaking the model into two parts: those terms that involve GENE and those that do not. First fit the effects that do not involve genes and construct residuals from this fit. This is a kind of “normalization” or centering of the overall data set. Then, in a second stage, use GENE as a BY variable and fit separate mixed models to each gene. For details on this method, refer to Wolfinger et al. (2001) and Gibson and Wolfinger (2004).

Appendix 1: Linear Mixed Model Theory



A1.1 Introduction	734
A1.2 Matrix Notation	734
A1.3 Formulation of the Mixed Model	735
A1.3.1 The General Linear Mixed Model.....	735
A1.3.2 Conditional and Marginal Distributions	736
A1.3.3 Example: Growth Curve with Compound Symmetry	737
A1.3.4 Example: Split-Plot Design	738
A1.4 Estimating Parameters, Predicting Random Effects	742
A1.4.1 Estimating β and Predicting u : The Mixed Model Equations	742
A1.4.2 Random Effects, Ridging, and Shrinking.....	743
A1.4.3 It's All in the Sweep	744
A1.4.4 Maximum Likelihood and Restricted Maximum Likelihood for Covariance Parameters.....	746
A1.5 Statistical Properties	751
A1.6 Model Selection	752
A1.6.1 Model Comparisons via Likelihood Ratio Tests.....	752
A1.6.2 Model Comparisons via Information Criteria.....	753
A1.7 Inference and Test Statistics	754
A1.7.1 Inference about the Covariance Parameters.....	754
A1.7.2 Inference about Fixed and Random Effects.....	755

A1.1 Introduction

This appendix provides an overview of a modern approach to general linear mixed models. This approach simplifies and unifies many common statistical analyses, including those involving repeated measures, random effects, random coefficients, spatial smoothing, and Gaussian regularization methods. The main focus of this appendix is the linear mixed model for normally distributed data with normally distributed random effects. Extensions to generalized linear mixed and nonlinear mixed models are possible if you pursue the route of linearization methods (see Chapters 14 and 15), where a model is approximated by a linear mixed model. Theory for nonlinear mixed models, where fitting relies on the likelihood principle and integral approximations, is not discussed here.

The primary aim of this appendix is to provide you with additional mathematical and statistical background regarding fitting linear mixed models, beyond the material covered in the main chapters of this book. Also, we hope to give you a better understanding of statistical inference using linear mixed models, including prediction and hypothesis testing.

A1.2 Matrix Notation

Suppose you observe n data points y_1, \dots, y_n and you want to explain them using n values for each of p explanatory variables $x_{11}, \dots, x_{1p}, x_{21}, \dots, x_{2p}, \dots, x_{n1}, \dots, x_{np}$. The x_{ij} values may be either regression-type continuous variables or dummy variables indicating class membership. The standard linear model for this setup is

$$\begin{aligned} Y_i &= \sum_{j=1}^p x_{ij}\beta_j + e_i \\ i &= 1, \dots, n \end{aligned} \tag{A1.1}$$

where β_1, \dots, β_p are unknown **fixed effects parameters** to be estimated and the errors e_1, \dots, e_n are unknown independent and identically distributed normal (Gaussian) random variables with mean 0 and variance σ^2 .

Model (A1.1) can be written using vectors and a matrix as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

For convenience, simplicity, and extendability, we write this entire system as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{A1.2}$$

where \mathbf{Y} denotes the vector of responses, \mathbf{X} is the known matrix of x_{ij} 's, $\boldsymbol{\beta}$ is the unknown fixed effects parameter vector, and \mathbf{e} is the unobserved vector of independent and identically distributed Gaussian random errors. When we refer to the responses as random variables, we use uppercase notation, \mathbf{Y} . When we refer to the observed values, lowercase notation is used,

$\mathbf{y} = [y_1, y_2, \dots, y_n]'$. The fact that a random vector \mathbf{U} is normally distributed with mean $\boldsymbol{\mu}$ and variance \mathbf{V} is denoted $\mathbf{U} \sim N(\boldsymbol{\mu}, \mathbf{V})$. So in the general linear model (A1.2) we write $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

In addition to denoting data, random variables, and explanatory variables in the preceding fashion, the subsequent development makes use of basic matrix operators such as transpose ('), inverse ($^{-1}$), generalized inverse (\top), determinant ($|\mathbf{V}|$), and matrix multiplication. Refer to Searle (1982) and Harville (1997) for details on these and other matrix techniques.

A1.3 Formulation of the Mixed Model

A1.3.1 The General Linear Mixed Model

The previous general linear model is certainly a useful one (Searle 1971), and it is the one fitted by the GLM procedure. However, many times the independence distributional assumption about \mathbf{Y} is too restrictive. The mixed model extends the general linear model by allowing elements of \mathbf{Y} to be correlated. You can motivate this extension by allowing a more flexible specification of the covariance matrix of \mathbf{e} —say, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$. This is the approach taken in some models for repeated measures data, temporal data, or spatial data. Or you can motivate the extension as one that permits random effects and random coefficients in the analysis, giving rise to the $\mathbf{Z}\mathbf{u}$ terms in the model, where \mathbf{u} is normal with mean $\mathbf{0}$ and variance \mathbf{G} . The two approaches are sometimes used interchangeably, but there are important differences. We refer to a model using random effects and a \mathbf{Z} matrix as a **conditional** model, whereas models without a \mathbf{Z} matrix that capture complex covariance structure directly through the variance matrix of the errors \mathbf{e} are called **marginal** models. Because the variance matrix of the random effects \mathbf{u} is typically denoted \mathbf{G} and the variance matrix of \mathbf{e} is typically denoted \mathbf{R} , we also refer to G-side and R-side modeling of the variation in the data.

The general linear mixed model can be written as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \\ \mathbf{u} &\sim N(\mathbf{0}, \mathbf{G}) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{R}) \\ \text{Cov}[\mathbf{u}, \mathbf{e}] &= \mathbf{0} \end{aligned} \tag{A1.3}$$

In the general linear model (A1.1), it is possible to relax the normality assumption on the errors and to assume that they just have mean zero and variance σ^2 , leading often to methods that are robust to specific distributional assumptions. We do not pursue this kind of modeling here but rather make an explicit normality assumption accompanying the specification of \mathbf{u} and \mathbf{e} . This assumption is often verified informally in practice via various plots. Also, the assumption that \mathbf{u} and \mathbf{e} are uncorrelated (because of the normality, this implies independence) is a key aspect of the model that avoids parameter confounding.

Parameters of a statistical model are **fixed unknown constants** to be estimated from the data. The parameters in model (A1.3) are thus the fixed effects vector $\boldsymbol{\beta}$ and all unknowns in the covariance matrices \mathbf{G} and \mathbf{R} . The **random effects** \mathbf{u} are *not* parameters, since they are not fixed. Although \mathbf{Z} , like \mathbf{X} , can contain either continuous or dummy variables, the vector \mathbf{u} contains random variables. It is the presence of fixed effects parameters $\boldsymbol{\beta}$ and of G-side random effects \mathbf{u} from which the class of **mixed models** derives its classical name, although we use

“mixed models” in a much broader sense to include general covariance modeling in \mathbf{G} and \mathbf{R} . All unknowns in the \mathbf{G} and \mathbf{R} matrices are collectively referred to as the **covariance parameters** and denoted as $\boldsymbol{\theta}$.

Refer to Henderson (1990) and Searle, Casella, and McCulloch (1992) for historical developments of the mixed model.

A1.3.2 Conditional and Marginal Distributions

Because the right-hand side of the general linear mixed model contains two random vectors, \mathbf{u} and \mathbf{e} , we can consider different distributions. The conditional distribution of $\mathbf{Y}|\mathbf{u}$ and the marginal distribution of \mathbf{Y} are as follows:

$$\begin{aligned}\mathbf{Y}|\mathbf{u} &\sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}) \\ \mathbf{Y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}) \\ \mathbf{V} = \text{Var}[\mathbf{Y}] &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}\end{aligned}$$

Notice that the variance of the conditional distribution is simply $\text{Var}[\mathbf{e}] = \mathbf{R}$. If a model does not have G-side random effects ($\mathbf{u} = \mathbf{0}$, $\mathbf{G} = \mathbf{0}$), the two distributions are the same and \mathbf{R} is also the marginal variance. This is the approach taken in, for example, Chapter 5, where repeated measures are modeled with a REPEATED statement and without a RANDOM statement in PROC MIXED. Such a model is not a mixed model in the classical sense of fixed and random effects. Similarly, the heterogeneous variance models in Chapter 8 and some of the spatial models in Chapter 11 are not mixed models in the classical sense, since variation is not modeled through random effects, but directly through the \mathbf{R} matrix. However, mixed model software such as PROC MIXED can often be used to fit marginal models because the procedure accommodates complex variation in the conditional variance matrix \mathbf{R} .

Details about the process of fitting a mixed model by (restricted) maximum likelihood are given below. Suffice it to say at this point that the fitting process relies on maximizing the **marginal** log likelihood of the data. Two models that differ in their specification for \mathbf{G} , \mathbf{R} , and \mathbf{u} can lead to the same marginal log likelihood. A common example is a conditional model with a single variance component and a marginal model with compound-symmetric covariance structure. This model equivalency can be exploited to fit a model in the most computationally efficient manner. For example, in models without G-side random effects the \mathbf{Z} matrix does not have to be formed and stored. However, the equivalency is one with respect to the marginal distribution alone. You may need random effects in the model to obtain certain inferential quantities such as predicted values using empirical BLUPs or prediction standard errors. Also, the equivalency of the models may hold only under certain additional restriction. For example, the marginal compound-symmetry model and the variance component model are equivalent only if the intra-class correlation is nonnegative, since the variance component must be nonnegative.

A1.3.3 Example: Growth Curve with Compound Symmetry

Suppose you have three growth curve measurements for s individuals, and you want to fit an overall linear trend in time. Your \mathbf{X} matrix is as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

The first column (coded entirely with 1s) fits an intercept, and the second column (coded with times of 1,2,3) fits a slope. Here $n = 3s$ and $p = 2$.

Suppose further that you want to introduce a common correlation among the observations from a single individual, and that correlation is the same for all individuals. One way of setting this up in the general mixed model is to eliminate the \mathbf{Z} and \mathbf{G} matrices and let the \mathbf{R} matrix be block diagonal with blocks corresponding to the individuals, each block having the *compound-symmetry* structure. This structure has two unknown parameters, one modeling a common covariance and the other modeling a residual variance. The form for \mathbf{R} is then as follows:

$$\mathbf{R} = \begin{bmatrix} \sigma_i^2 + \sigma^2 & \sigma_i^2 & \sigma_i^2 \\ \sigma_i^2 & \sigma_i^2 + \sigma^2 & \sigma_i^2 \\ \sigma_i^2 & \sigma_i^2 & \sigma_i^2 + \sigma^2 \\ & \ddots & \\ & & \sigma_i^2 + \sigma^2 & \sigma_i^2 & \sigma_i^2 \\ & & \sigma_i^2 & \sigma_i^2 + \sigma^2 & \sigma_i^2 \\ & & \sigma_i^2 & \sigma_i^2 & \sigma_i^2 + \sigma^2 \end{bmatrix}$$

The PROC MIXED program to fit this model is the following:

```
proc mixed;
  class indiv;
  model y = time;
  repeated / type=cs subject=indiv;
run;
```

Here INDIV is a classification variable indexing individuals. The MODEL statement fits a straight line for TIME; the intercept is fit by default. The REPEATED statement models the \mathbf{R} matrix: TYPE=CS specifies the compound symmetry structure and SUBJECT=INDIV specifies the blocks of \mathbf{R} .

An alternative way of specifying the common intra-individual correlation is to let

$$\mathbf{Z} = \begin{bmatrix} 1 & & & & \\ 1 & & & & \\ 1 & & & & \\ & 1 & & & \\ & 1 & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & 1 & \\ & & & 1 & \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \ddots & \\ & & & \sigma_1^2 \end{bmatrix}$$

and $\mathbf{R} = \sigma^2 \mathbf{I}_n$. \mathbf{Z} has $3s$ rows and s columns, and \mathbf{G} is $s \times s$.

You can set up this model in PROC MIXED in two different but equivalent ways:

```
proc mixed;
  class indiv;
  model y = time;
  random indiv;
run;

proc mixed;
  class indiv;
  model y = time;
  random intercept / subject=indiv;
run;
```

Both of these specifications fit the same model as the previous one using the REPEATED statement; however, the RANDOM specifications constrain the correlation to be nonnegative, whereas the REPEATED specification leaves the correlation unconstrained.

A1.3.4 Example: Split-Plot Design

The split-plot design involves two experimental treatment factors, **A** and **B**, and two different sizes of experimental units to which they are applied (refer to Winer 1971, Snedecor and Cochran 1989, Milliken and Johnson 1992). The levels of **A** are randomly assigned to the larger size of experimental unit, called **whole plots**, whereas the levels of **B** are assigned to the smaller size of experimental unit, the **sub-plots**. The sub-plots are assumed to be nested within the whole plots so that a whole plot consists of a cluster of sub-plots and a level of **A** is applied to the entire cluster.

Such an arrangement is often necessary by the nature of the experiment, the classical example being the application of fertilizer to large plots of land, and different crop varieties planted in subdivisions of the large plots. For this example, fertilizer is the whole-plot factor **A** and variety the sub-plot factor **B**. In addition, we assume the whole plots are arranged in a randomized block design. The appropriate PROC MIXED program for this model is as follows:

```

proc mixed;
  class a b block;
  model y = a|b;
  random block a*block;
run;

```

Here, $\mathbf{R} = \sigma^2 \mathbf{I}_{24}$ and \mathbf{X} , \mathbf{G} , and \mathbf{Z} have the following form (vertical lines delineate columns associated with individual effects):

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & & 1 & 1 & & \\ 1 & 1 & & 1 & 1 & & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & & 1 & 1 & & \\ 1 & 1 & & 1 & 1 & & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \\ 1 & & 1 & 1 & & 1 & \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_B^2 & & & & & & \\ & \sigma_B^2 & & & & & \\ & & \sigma_B^2 & & & & \\ & & & \sigma_B^2 & & & \\ & & & & \sigma_{AB}^2 & & \\ & & & & & \sigma_{AB}^2 & \\ & & & & & & \ddots \\ & & & & & & & \sigma_{AB}^2 \end{bmatrix}$$

where σ_B^2 is the variance component for BLOCK and σ_{AB}^2 is the variance component for A \times BLOCK. Changing the RANDOM statement to

```
random int a / subject=block;
```

fits the same model, but with **Z** and **G** sorted differently:

$$\mathbf{G} = \begin{bmatrix} \sigma_B^2 & & & & \\ & \sigma_{AB}^2 & & & \\ & & \sigma_{AB}^2 & & \\ & & & \ddots & \\ & & & & \sigma_B^2 \\ & & & & & \sigma_{AB}^2 \\ & & & & & & \sigma_{AB}^2 \\ & & & & & & & \sigma_{AB}^2 \end{bmatrix}$$

A1.4 Estimating Parameters, Predicting Random Effects

In order to proceed with statistical inference in a mixed model we need to estimate the unknowns β , G , and R and predict the random variables u . The special structure of the normal distribution for u and e allows us to divide the theoretical development into two steps. First, we consider G and R to be known and show how to obtain solutions for β and u . Second, we inquire how to estimate the unknowns in G and R . (Note that in actual computations, the second step is performed first.) In a subsequent section we address the consequences of working with estimates of G and R rather than known matrices. As demonstrated throughout this book, the fact that statistical inference in mixed models relies on estimated covariance parameters has implications for the quality of standard errors, p -values, and Type I errors.

A1.4.1 Estimating β and Predicting u : The Mixed Model Equations

If θ , the vector of covariance parameter, is known, we can construct the G and R matrices. Let g denote the number of elements in u , and let n be the total sample size (the dimension of y).

Henderson (1950, 1984) motivated estimation of β and prediction of u through a set of least-squares-type estimating equations. Although initially referred to as joint maximum likelihood estimation, the function being maximized is not a true likelihood. This part of mixed model estimation must not be confused with (restricted) maximum likelihood estimation of the covariance parameters θ . It is simply a method to motivate estimating equations for β and u . Commence by formulating the joint distribution of u and e ,

$$f(u, e) = \frac{1}{(2\pi)^{(n+g)/2}} \begin{vmatrix} G & \mathbf{0} \\ \mathbf{0} & R \end{vmatrix}^{-1/2} \exp \left\{ -\frac{1}{2} \begin{bmatrix} u \\ y - X\beta - Zu \end{bmatrix}' \begin{bmatrix} G^{-1} & \mathbf{0} \\ \mathbf{0} & R^{-1} \end{bmatrix} \begin{bmatrix} u \\ y - X\beta - Zu \end{bmatrix} \right\}$$

Maximization of $f(u, e)$ with respect to β and u requires minimization of

$$\begin{aligned} Q &= \begin{bmatrix} u \\ y - X\beta - Zu \end{bmatrix}' \begin{bmatrix} G^{-1} & \mathbf{0} \\ \mathbf{0} & R^{-1} \end{bmatrix} \begin{bmatrix} u \\ y - X\beta - Zu \end{bmatrix} \\ &= u'G^{-1}u + (y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) \end{aligned}$$

where we have taken advantage of the independence of u and e . This leads to the equations

$$\begin{aligned} \partial Q / \partial \beta &= 0 \Leftrightarrow X'R^{-1}X\tilde{\beta} + X'R^{-1}Z\tilde{u} = X'R^{-1}y \\ \partial Q / \partial u &= 0 \Leftrightarrow (Z'R^{-1}Z + G^{-1})\tilde{u} + Z'R^{-1}X\tilde{\beta} = Z'R^{-1}y \end{aligned}$$

After some rearranging, these can be written as

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (\text{A1.4})$$

The equations in (A1.4) are known as Henderson's mixed model equations, or simply as the **mixed model equations**. The solutions are

$$\begin{aligned} \begin{bmatrix} \tilde{\beta} \\ \tilde{u} \end{bmatrix} &= \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}) \end{bmatrix} \end{aligned} \quad (\text{A1.5})$$

Since the right-most term in the parentheses for the random effects solution is the fixed effects estimate, we can also write

$$\tilde{u} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta})$$

A1.4.2 Random Effects, Ridging, and Shrinking

Before going into further details about clever ways of solving this system of equations, it is helpful to ponder the mixed model equations for a moment. The equations in (A1.4) have an almost familiar form, resembling estimating equations in the linear model. Suppose that the model of interest is $\mathbf{Y} = \mathbf{X}\beta_1 + \mathbf{Z}\beta_2 + \mathbf{e}$, where both β_1 and β_2 are fixed effects, and \mathbf{e} has mean $\mathbf{0}$ and variance \mathbf{R} . Then the generalized least-squares equations would be

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

The matrix in the lower-right corner of the coefficient matrix changes from $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ in the all-fixed-effects model to $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$ in the mixed model. When you move an effect from the fixed effects part to the random effects part of the model, you do not achieve the same predictions, because the random effects solutions are not identical to the solutions you would obtain had the same effect been fixed. An interesting way of looking at this comparison is to think of the \mathbf{G}^{-1} matrix as a ridge factor applied to the random effects regression component. Ridge factors are used in standard regression to combat multicollinearity; the ridging has the effect of shrinking the parameter estimates. In the mixed model equations, the "ridge factor" \mathbf{G}^{-1} also creates shrinkage in the random effects solutions—a regression toward the mean. The mixed model equations are sometimes referred to as extended normal equations because of the involvement of \mathbf{G}^{-1} .

It can be helpful to consider extremes. First assume that \mathbf{G} is non-singular and that its eigenvalues are very large. In the case of a diagonal \mathbf{G} matrix this means that the diagonal elements are large, and the random effects have large variance. In this case, \mathbf{G}^{-1} contributes very little to the equations, and \tilde{u} is very close to what it would be if \mathbf{u} actually contained fixed effects parameters. On the other hand, when the eigenvalues of \mathbf{G} are very small, \mathbf{G}^{-1} dominates the equations and \tilde{u} is close to $\mathbf{0}$. If \mathbf{G} is singular, the equations must be modified (Henderson 1984), and the elements of \tilde{u} corresponding to the singular portion of \mathbf{G} equal $\mathbf{0}$. An example of this situation is when a variance component estimate falls on the boundary constraint of $\mathbf{0}$. For intermediate cases, \mathbf{G}^{-1} shrinks the fixed effects estimates of \mathbf{u} toward $\mathbf{0}$ (Robinson 1991).

In nonparametric regression, splines are often fit by adding a penalty criterion to control the magnitude of the spline coefficients and the degree of smoothing. If \mathbf{B} is some matrix formed from a spline basis, and $\boldsymbol{\alpha}$ are the spline coefficients, then a solution can be obtained by, for example,

$$\hat{\boldsymbol{\alpha}} = (\mathbf{B}'\mathbf{B} + \lambda\mathbf{D})^{-1}\mathbf{B}'\mathbf{y}$$

where \mathbf{D} is a symmetric matrix and λ is the smoothing parameter (see, for example, Eilers and Marx 1996). If you put $\lambda\mathbf{D} = \mathbf{G}^{-1}$, the estimation criterion for penalized splines can be cast in terms of mixed model equations (Ruppert, Wand, and Carroll 2003).

A1.4.3 It's All in the Sweep

It is typically very inefficient to compute the solutions for $\boldsymbol{\beta}$ and \mathbf{u} directly according to the last expression in equation (A1.5), because it involves the inverse of the marginal variance matrix. For example, in a variance component model, \mathbf{R} and \mathbf{G} are diagonal matrices, but $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ is not diagonal. The formation and inversion of \mathbf{V} is a computationally costly proposition. A much more elegant way to obtain solutions is by way of a sweep (Goodnight 1979), an operation related to Gauss-Jordan elimination and the Forward Doolittle procedure. To fix ideas, consider the nonnegative definite, symmetric matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}$$

and apply the sweep operator to its leading partition (\mathbf{A}_{11}). This operation, denoted SWEEP($\mathbf{A}, \mathbf{A}_{11}$), yields

$$\text{SWEEP}(\mathbf{A}, \mathbf{A}_{11}) = \begin{bmatrix} \mathbf{A}_{11}^- & \mathbf{A}_{11}^- \mathbf{A}_{12} \\ -\mathbf{A}'_{12} \mathbf{A}_{11}^- & \mathbf{A}_{22} - \mathbf{A}'_{12} \mathbf{A}_{11}^- \mathbf{A}_{12} \end{bmatrix}$$

where \mathbf{A}^- is a generalized inverse (a g2-generalized inverse). You can apply sweep operations in the mixed model problem as follows. First, augment the coefficient matrix on the left-hand side of the mixed model equations in (A1.4) with a “y border.”

$$\mathbf{C}^- = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}$$

$$\mathbf{M} = \left[\begin{array}{cc|c} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \hline \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{array} \right] = \left[\begin{array}{cc|c} \mathbf{C}^- & & \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix} \mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}[\mathbf{X} \quad \mathbf{Z}] & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} & \end{array} \right]$$

The operation SWEEP(\mathbf{M}, \mathbf{C}^-) yields a number of important quantities that are also needed later in computing the (restricted) log likelihood of the data. First, the matrix \mathbf{M} is transformed into

$$\text{SWEEP}(\mathbf{M}, \mathbf{C}^-) = \begin{bmatrix} \mathbf{C} & \begin{bmatrix} \tilde{\beta} \\ \tilde{\mathbf{u}} \end{bmatrix} \\ -[\tilde{\beta} \quad \tilde{\mathbf{u}}]' & \mathbf{r}'\mathbf{V}^{-1}\mathbf{r} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathbf{X}\tilde{\beta} \\ \mathbf{C} &= \begin{bmatrix} \Omega & -\Omega\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\ -\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}\Omega & \mathbf{Q} + \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}\Omega\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \end{bmatrix} \\ \Omega &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ \mathbf{Q} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \end{aligned} \tag{A1.6}$$

The “y border” of \mathbf{M} has been transformed into the solutions for the fixed effects parameter estimates and the solutions for the random effects. In the lower-right corner of the swept matrix we find the weighted residual sum of squares, $\mathbf{r}'\mathbf{V}^{-1}\mathbf{r}$.

The matrix \mathbf{C} , which occupies the swept partition, is the estimated covariance matrix of $[\tilde{\beta} \quad \tilde{\mathbf{u}} - \mathbf{u}]$. Its leading term, Ω , is the variance-covariance matrix of the fixed effects parameter estimates. The term in the lower-right corner of \mathbf{C} is the **prediction** variance matrix of the random effects. Because the \mathbf{u} are random, it is often not appropriate to simply consider $\text{Var}[\tilde{\mathbf{u}}]$ in statistical inference. This would not take into account that the \mathbf{u} are unobservable quantities. Instead, we need to consider the variance of the difference between the solutions and the unknown random effects. Here we end up with the variance of a difference—rather than $\text{Var}[\tilde{\mathbf{u}}]$.

The fact that the same rationale seemingly does not apply to the fixed effects is easily explained. Both variances are in fact mean-squared prediction errors. Suppose we want to predict the quantity U based on some function $f(\mathbf{Y})$, and suppose that U is unbiased in the sense that $E[U] = E[f(\mathbf{Y})]$. The mean-squared prediction error for U based on $f(\mathbf{Y})$ is

$$\text{MSE}[U; f(\mathbf{Y})] = E[(U - f(\mathbf{Y}))^2] = \text{Var}[U - f(\mathbf{Y})] = \text{Var}[U] + \text{Var}[f(\mathbf{Y})] - 2\text{Cov}[U, f(\mathbf{Y})]$$

If the target U is fixed, not random, then the mean-squared prediction error reduces to $\text{Var}[f(\mathbf{Y})]$. This is the case for the fixed effects. For the random effects, the target is \mathbf{u} and we are predicting it with $f(\mathbf{Y}) = \tilde{\mathbf{u}}$. Because $\tilde{\mathbf{u}}$ is unbiased for \mathbf{u} in the sense that both have expectation zero—which relied on the fact that $\tilde{\beta}$ is unbiased for β —the mean-squared prediction error for the random effects has the form of a variance of a difference.

We can also view the linear mixed model setup from a Bayesian perspective, where a prior distribution is assumed for β and u . The prior for the fixed effects parameters is flat, whereas the prior for u is assumed to be normal with mean $\mathbf{0}$ and variance \mathbf{G} . With \mathbf{G} and \mathbf{R} known, the posterior distribution for $[\beta \ u]$ is normal, and the variance and mean correspond to the partitions in the first row of $\text{SWEET}(\mathbf{M}, \mathbf{C})$.

A side product of sweeping, albeit a very important one, is the log determinant of the swept matrix. It is obtained by summing the log pivots (the log of the diagonal element prior to sweeping the row) of the rows that are swept. You can show that $\log |\mathbf{C}|$ equals

$$\begin{aligned}\log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \log |\mathbf{V}| - \log |\mathbf{R}| - \log |\mathbf{G}| \\ = \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \log |\mathbf{Z}'\mathbf{R}\mathbf{Z} + \mathbf{G}^{-1}|\end{aligned}$$

These log determinants are components of the (restricted) log likelihood of the mixed model (see below). It really is “all in the sweep.”

The previous expressions and derivations tacitly assumed that θ is known and hence \mathbf{G} and \mathbf{R} can be formed as needed. We also tacitly assumed that \mathbf{G} and \mathbf{R} are positive definite. Although we are not going to address in this appendix how to handle the case of non-positive definite variance matrices, the issue of not knowing the covariance parameters needs to be addressed. Otherwise, none of the mentioned results are operational. An obvious amendment of the derivations is to use “plug-in estimates”—that is, replace the unknown quantities in the equations with estimates. For example, \mathbf{R} and \mathbf{G} are replaced with $\hat{\mathbf{R}}$ and $\hat{\mathbf{G}}$, which are formed by evaluating their elements based on the estimate $\hat{\theta}$ of θ . The estimator of β and the predictor of u given above are a generalized least squares estimator (GLSE) and a best linear unbiased predictor (BLUP), respectively. When these are computed based on estimated covariance parameters, we refer to them as **estimated** GLS estimator (EGLSE) and as **estimated** (or empirical) BLUP (EBLUP).

A1.4.4 Maximum Likelihood and Restricted Maximum Likelihood for Covariance Parameters

Throughout this text we have invoked various methods for estimating covariance parameters: method of moments (Type 1, 2, 3), minimum variance quadratic unbiased estimation (MIVQUE0), maximum likelihood (ML), restricted maximum likelihood (REML), and pseudo-likelihood (PL). Arguably the most important method of estimating covariance parameters is restricted maximum likelihood estimation, and it is the default of PROC MIXED. (It is also the default of PROC GLIMMIX for mixed models with normally distributed data.) The details on likelihood estimation in linear mixed models that apply to PROC MIXED are given in Wolfinger, Tobias, and Sall (1994). Further details can be found in, e.g., Hartley and Rao (1967), Patterson and Thompson (1971), Harville (1977), Laird and Ware (1982), and Jennrich and Schluchter (1986). We report here the necessary details to tie the optimization of a likelihood function to the solution of the mixed model equations in the previous sections and to exemplify the differences between maximum likelihood and restricted maximum likelihood estimation.

PROC MIXED minimizes twice the negative of the (restricted) log likelihood. Following Wolfinger, Tobias, and Sall (1994), the subscript R is used in the following expressions to distinguish REML estimation from ML estimation.

Maximum Likelihood (ML)

The $-2 \log$ likelihood for model (A1.3) is

$$-2l(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{y}) = \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + c \quad (\text{A1.7})$$

where c is the constant $n \log(2\pi)$. In Section A1.4.1 we have seen that, given the covariance parameter $\boldsymbol{\theta}$, a solution for $\boldsymbol{\beta}$ can be obtained in closed form as

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\mathbf{X}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{X})^{-1} \mathbf{X} \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{y}$$

We have written this solution here as an explicit function of the covariance parameters, to make the dependency apparent. You can substitute $\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ for $\boldsymbol{\beta}$ in equation (A1.7) to obtain a likelihood from which the fixed effects have been profiled. This process reduces the number of parameters for which the optimization needs to be carried out. The only unknowns in the resulting $-2 \log$ likelihood,

$$-2l(\boldsymbol{\theta}; \mathbf{y}) = \log |\mathbf{V}(\boldsymbol{\theta})| + (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})) + c \quad (\text{A1.8})$$

are the covariance parameters.

The minimum of this objective function, $\hat{\boldsymbol{\theta}}$, is the maximum likelihood estimator of $\boldsymbol{\theta}$, provided the elements of $\hat{\boldsymbol{\theta}}$ are in the parameter space. The ML estimator of the fixed effects parameters is then obtained by evaluating $\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})$ at the ML estimate of $\boldsymbol{\theta}$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{y}$$

Restricted Maximum Likelihood (REML)

REML estimation is arguably the most important technique for estimating covariance parameters in mixed models. It is a likelihood-based technique that aims to remedy some of the shortcomings of regular maximum likelihood estimation; it traces back to Thompson (1962) and Patterson and Thompson (1971). ML estimates are very desirable statistically—for example, they tend to be asymptotically efficient estimates and follow asymptotic normal distributions under some mild regularity conditions. However, most of the desirable properties are asymptotic, and require an increasing sample size. For example, many MLEs are asymptotically unbiased but biased for any fixed sample size. The small-sample bias can be easily illustrated with the following simple example.

Suppose that Y_1, \dots, Y_n is a random sample from a $N(\mu, \sigma^2)$ distribution, and suppose that μ and σ^2 are unknown. Twice the negative of the log likelihood of the sample is

$$-2l(\mu, \sigma^2; \mathbf{y}) = n \ln(2\pi) + n \ln(\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$$

It is easy to show that this function is minimized by

$$\hat{\mu} = n^{-1} \sum_{i=1}^n y_i = \bar{y}$$

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (y_i - \hat{\mu})^2$$

the maximum likelihood estimators of the mean and the variance. Whereas the MLE of μ is an unbiased estimator, the MLE of σ^2 is biased,

$$E[\hat{\sigma}^2 - \sigma^2] = -\frac{1}{n}\sigma^2$$

This bias shrinks with the sample size, and it is negative. The MLE of σ^2 underestimates the variance on average. Notice that if we had known the mean μ , then our MLE for the variance would have been

$$\check{\sigma}^2 = n^{-1} \sum_{i=1}^n (y_i - \mu)^2$$

which is an unbiased estimator of σ^2 . So the finite-sample-size “problem” of MLEs for variance or covariance parameters appears to be linked to “not knowing the mean.” More precisely, it is linked to **not accounting for the unknown mean in the estimation**. In the terminology of mixed models, the issue is the accounting of the fixed effects parameters in estimating the covariance parameters. That is the motivation of Patterson and Thompson (1971).

REML estimation can be cast as likelihood estimation for transformed data. Instead of the log likelihood of \mathbf{Y} , we consider the log likelihood of \mathbf{KY} , where the matrix \mathbf{K} is chosen so that $E[\mathbf{KY}] = \mathbf{0}$. \mathbf{K} is also called a matrix of **error contrasts**; hence the alternate name of **residual** log likelihood. If we consider the $-2 \log$ likelihood of \mathbf{KY} , then

$$-2l_R(\boldsymbol{\theta}; \mathbf{Ky}) = \log |\mathbf{KV}(\boldsymbol{\theta})\mathbf{K}'| + \mathbf{y}'\mathbf{K}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{Ky} + c_R$$

and the fixed effects $\boldsymbol{\beta}$ have seemingly dropped from estimation. The process is not quite identical to profiling $\boldsymbol{\beta}$, since the REML objective function by definition is an objective function for the covariance parameters only. For the linear model with $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, Harville (1977) points out that the $(n-p) \times n$ matrix (where p is the rank of \mathbf{X}) constructed from linearly independent rows of $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ leads to the same objective function, which differs by only a constant amount that does not depend on the parameters.

We can use this result to find the REML estimator of σ^2 in the random sample example above. First, we can express the random sampling situation with a simple model,

$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{e} \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

so that $p = 1$, $\mathbf{X} = \mathbf{1}$, and $\mathbf{M} = \mathbf{I} - \mathbf{J}/n$, where \mathbf{J} is an $(n \times n)$ matrix of ones. So we choose as the matrix \mathbf{K} of error contrasts the first $n-1$ rows of \mathbf{M} :

$$\mathbf{K}_{(n-1) \times n} = \begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \dots & \frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1 - \frac{1}{n} \end{bmatrix} \Rightarrow \mathbf{KY} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_{n-1} - \bar{Y} \end{bmatrix}$$

The new “data” vector \mathbf{KY} now consists of the deviations of the first $n-1$ observations from the sample mean. We have seemingly reduced the size of the data from n to $n-1$, but we have not lost any information. If we know the sample mean and $n-1$ observations, the full data set of n observations can be constructed. Also, it does not matter which of the observations is dropped (which row of \mathbf{M} is eliminated); all vectors \mathbf{KY} so constructed are sufficient for σ^2 . Or, in the words of Harville (1977): “the REML estimator does not ignore any information that is actually *used* by the full approach” (ML). To construct the REML log likelihood for this problem, we now follow Schabenberger and Pierce (2002, Ch. 7.4.1) and the details provided therein. First, denote $\mathbf{KY} = \mathbf{U}$ and note that $\text{Var}[\mathbf{U}] = \sigma^2 \mathbf{KK}' = \sigma^2 \mathbf{P}$. The matrix \mathbf{P} has some remarkable properties—for example, $|\mathbf{P}| = 1/n$, $\mathbf{P}^{-1} = \mathbf{I}_{n-1} + \mathbf{J}_{n-1}$. The $-2 \log$ likelihood for \mathbf{U} is

$$-2l_R(\sigma^2; \mathbf{u}) = (n-1) \ln(\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 + c_R$$

and the solution is the REML estimate

$$\hat{\sigma}_R^2 = (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

The REML estimate equals the sample variance and is an unbiased estimator of σ^2 . If you compare $\hat{\sigma}^2$ and $\hat{\sigma}_R^2$, both estimators involve a sum-of-squares term with deviations from the sample mean. The presence of \bar{y} in the formula for the estimators has a different origin, however. In ML estimation, we substitute the MLE for μ , and that happens to be \bar{y} . In REML estimation we have eliminated the unknown mean from the estimation altogether. \bar{y} appears in the formula for $\hat{\sigma}_R^2$ because the quadratic form $\mathbf{u}'\mathbf{P}^{-1}\mathbf{u}$ equals the corrected sum of squares. This underscores the point that—by itself—REML estimation does not provide estimates of the fixed effect parameters. It yields only estimates of the covariance parameters.

For the general linear mean function $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, the process works in the same way, by forming a matrix of error contrast based on the independent rows of the projection matrix $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Since any set of rows will do, the following objective function is common for REML estimation:

$$-2l_R(\boldsymbol{\theta}; \mathbf{y}) = \log |\mathbf{V}(\boldsymbol{\theta})| + \log |\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}| + (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta})) + c_R$$

Dropping the notational dependence on the covariance parameters and replacing the residual vector with \mathbf{r} , we can write

$$-2l_R(\boldsymbol{\theta}; \mathbf{y}) = \log |\mathbf{V}| + \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{r}'\mathbf{V}^{-1}\mathbf{r} + c_R \quad (\text{A1.9})$$

The REML estimate of $\boldsymbol{\theta}$ is the vector $\hat{\boldsymbol{\theta}}_R$ that minimizes equation (A1.9) and whose elements are in the parameter space of $\boldsymbol{\theta}$. The fixed effects estimates are obtained as estimated generalized least squares estimates evaluated at the REML estimate of the covariance parameters

$$\hat{\boldsymbol{\beta}}_R = \left(\mathbf{X}'\mathbf{V}(\hat{\boldsymbol{\theta}}_R)^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}(\hat{\boldsymbol{\theta}}_R)^{-1}\mathbf{y}$$

Connecting the Dots

The interplay of minimizing equation (A1.9) for REML estimation or equation (A1.8) for ML estimation and the mixed model equations is now becoming apparent. The likelihood optimization yields values for the covariance parameters. Given these values we can set up and solve the mixed model equations by sweeping, thereby acquiring the necessary pieces to compute the objective function. The kernel of the -2 restricted log likelihood can be computed from elements of the sweep alone.

Since $\boldsymbol{\theta}$ is involved nonlinearly in equation (A1.9) or (A1.8), the process of estimating the covariance parameters relies on numerical methods. Starting from initial values of the covariance parameters $\boldsymbol{\theta}^0$, subsequent updates are obtained by an optimization algorithm until a convergence criterion is achieved. The MIXED procedure determines initial covariance parameters as MIVQUE0 estimates, unless you provide starting values with the PARMS statement. The optimization method implemented in PROC MIXED is a ridge-stabilized Newton-Raphson algorithm with analytic derivatives. Lindstrom and Bates (1988) provide reasons for preferring Newton-Raphson to the Expectation-Maximum (EM) algorithm described in Dempster, Laird, and Rubin (1977) and Laird, Lange, and Stram (1987), as well as analytical details for implementing a QR-decomposition approach to the problem. Wolfinger, Tobias, and Sall (1994) present the sweep-based algorithms that are implemented in PROC MIXED.

One advantage of using the Newton-Raphson algorithm is that the second derivative matrix of the objective function evaluated at the optima is available upon completion. Denoting this matrix \mathbf{H} , the asymptotic theory of maximum likelihood (refer to Serfling 1980) shows that the matrix $2\mathbf{H}^{-1}$ is an asymptotic variance-covariance matrix of the estimated parameters of \mathbf{G} and \mathbf{R} . Thus, tests and confidence intervals based on asymptotic normality can be obtained. However, these can be unreliable in small samples, especially for parameters such as variance components whose small-sample sampling distributions tend to be skewed to the right.

Instead of ML or REML, you can use the noniterative MIVQUE0 method to estimate \mathbf{G} and \mathbf{R} (Rao 1972, LaMotte 1973, Goodnight 1978, Wolfinger, Tobias, and Sall 1994). However, Swallow and Monahan (1984) present simulation evidence favoring REML and ML over MIVQUE0. MIVQUE0 is recommended primarily for large data sets or for situations where the iterative REML and ML procedures fail to converge.

A1.5 Statistical Properties

If \mathbf{G} and \mathbf{R} are known, $\tilde{\boldsymbol{\beta}}$ is the **best linear unbiased estimator** (BLUE) of $\boldsymbol{\beta}$ and $\tilde{\mathbf{u}}$ is the **best linear unbiased predictor** (BLUP) of \mathbf{u} (Searle 1971, Harville 1988, 1990, Robinson 1991, McLean, Sanders, and Stroup 1991). Here *best* means minimum mean squared error. We established earlier that the covariance matrix of $[\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{u}} - \mathbf{u}]$ is

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^-$$

where the minus sign ($-$) is used to denote a generalized inverse (refer to Searle 1971).

However, \mathbf{G} and \mathbf{R} are usually unknown and are estimated using one of the aforementioned methods. To be more precise, we could write these matrices as functions of the covariance parameters, $\mathbf{G}(\boldsymbol{\theta})$ and $\mathbf{R}(\boldsymbol{\theta})$. Estimates of the covariance parameters $\boldsymbol{\theta}$ are substituted to obtain estimates of the covariance matrices, $\hat{\mathbf{G}} = \mathbf{G}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{R}} = \mathbf{R}(\hat{\boldsymbol{\theta}})$. These, in turn, are substituted into the preceding expressions to obtain the “plug-in” mixed model equations

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

Notice that we now use caret notation (^) instead of tildes (~) to denote an estimator or predictor that depends on the estimated covariance parameters.

The BLUE and BLUP acronyms no longer apply in all respects to the solutions to the mixed model equations,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

$$\hat{\mathbf{u}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

For example, $\hat{\boldsymbol{\beta}}$ is no longer a linear function of the data. The word **empirical** (or estimated) is often added to indicate such a plug-in estimator. The appropriate acronyms thus become **EBLUE** and **EBLUP**. Because plug-in estimates are used to form the mixed model equations, the first partition of the mixed model solution now also depends on the covariance parameter estimates:

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^-$$

is a model-based variance-covariance matrix of $[\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}} - \mathbf{u}]$. This covariance matrix can be exact or approximate, depending on the structure of the model and the data. For example, in many balanced variance component models, it is *the* covariance matrix of $[\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}} - \mathbf{u}]$. Expression

(A1.6) provides formulas for the specific blocks of $\hat{\mathbf{C}}$. See also Henderson (1984) and McLean and Sanders (1988).

As a cautionary note, when $\hat{\mathbf{C}}$ is not exact, it tends to underestimate the true sampling variability of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ because no account is made for the uncertainty in estimating \mathbf{G} and \mathbf{R} . Several inflation factors have been proposed to the variance of linear functions (Kackar and Harville 1984, Kass and Steffey 1989, Prasad and Rao 1990, Harville and Jeske 1992) or covariance matrices (Kenward and Roger 1997). These factors tend to be small for data sets that are fairly well balanced but can be substantial for small unbalanced designs, or in cases with many covariance parameters. The DDFM=KENWARDROGER option in the MODEL statement of PROC MIXED (and PROC GLIMMIX) applies the standard error correction for fixed effects according to Kenward and Roger (1997). The prediction standard errors for random effects are adjusted according to what Harville and Jeske (1992) term the Prasad-Rao estimator. In their notation, this adjustment is $\tilde{m}^@(\hat{\theta})$. Note that the Prasad-Rao-Jeske-Kackar-Harville adjustment for a linear function of the fixed effects is identical to the Kenward and Roger adjustment if the covariance matrix does not have second derivatives (with respect to the covariance parameters). The Kenward-Roger method, however, also provides adjustments to t - and F -statistics and their degrees of freedom for testing hypotheses and constructing confidence intervals (see Section A1.7).

A1.6 Model Selection

The previous sections on estimation and statistical properties of the estimators assume the specification of a mixed model in terms of \mathbf{X} , \mathbf{Z} , \mathbf{G} , and \mathbf{R} . Even though \mathbf{X} and \mathbf{Z} have known elements, their specific form and construction is flexible, and several possibilities may present themselves for a particular data set. Likewise, several different covariance structures for \mathbf{G} and \mathbf{R} might be reasonable.

Space does not permit a thorough discussion of mixed model selection, but a few brief comments and references are in order. First, subject-matter considerations and objectives should certainly be foremost in your mind when selecting models. A few of the numerous useful references include Jennrich and Schluchter (1986), Diggle (1988), Lindsey (1993), Brown and Prescott (1999), and Verbeke and Molenberghs (1997, 2000).

Second, when the data themselves are looked to for guidance, many of the graphical methods and diagnostics appropriate for the general linear model extend to the mixed model setting as well (Christensen, Pearson, and Johnson 1992, Muller and Fetterman 2002).

A1.6.1 Model Comparisons via Likelihood Ratio Tests

Many analysts prefer a statistical test for model comparisons. When models are fit by maximum or restricted maximum likelihood, you can form a likelihood ratio test statistic by comparing $-2(\text{Res Log Likelihood})$ between the full and a reduced model. The important aspect here is that the “reduced” model is in fact nested in the full model. This means that you can attain the reduced model by imposing constraints on the full model, such as by setting covariance parameters to zero. When models are fit by restricted maximum likelihood (REML), the comparisons of models via restricted likelihood ratio tests must be confined to models that are nested with respect to the covariance parameters alone. Two models that have different \mathbf{X} matrices, whether nested or not, cannot be compared based on $-2 \text{ Res Log Likelihood}$. As

shown in Section A1.4.4, REML estimation can be viewed as maximum likelihood estimation for transformed data, \mathbf{KY} , where $E[\mathbf{KY}] = \mathbf{0}$. Changing the fixed effects part of the model changes \mathbf{K} , so that the two models are performing maximum likelihood estimation for two different sets of data.

Recently, increased attention has been given to the problem of likelihood ratio testing under nonstandard conditions. Typically, the nonstandard condition referred to stems from boundary constraints on the parameters, such as when a variance component is tested against zero. These tests are normally carried out as one-sided tests, with hypotheses

$$H_0 : \sigma^2 = 0 \quad H_a : \sigma^2 > 0$$

Under the null hypothesis the parameter falls on the boundary of the parameter space, unless you consider unbounded estimation (NOBOUND option in PROC MIXED) and a two-sided test. The usual asymptotic theory for the likelihood-ratio χ^2 statistic would suggest that the difference of the $-2 \log$ likelihoods is distributed as a χ^2 variable with a single degree of freedom. When tested parameters or nuisance parameters (the parameters not specified in the hypothesis) fall on the boundary of the parameter space, the standard theory does not apply (see, for example, Self and Liang 1987 and Shapiro 1988). In this case, however, the null sampling distribution of the test statistic is a 50:50 mixture of a χ_0^2 and a χ_1^2 variable. As a consequence, you can perform the likelihood ratio test as if the standard conditions apply, and divide the resulting p -value by two. In more complex cases, such as when multiple parameters are tested, or when nuisance parameters fall on the boundary, the correction for the likelihood ratio test is not as straightforward. A number of cases have been explored in Self and Liang (1987) and specifically for variance component models in Stram and Lee (1994, 1995). Silvapulle and Silvapulle (1995) and Verbeke and Molenberghs (2003) consider similar problems with the score test.

A1.6.2 Model Comparisons via Information Criteria

As an alternative to likelihood ratio testing, one can take an information-theoretical perspective and compute a variety of information-based measures. These are usually computed as a penalty term applied to the likelihood function. The most common of these are the likelihood ratio test and Akaike's and Schwarz's criteria (Bozdogan 1987, Wolfinger 1993a). Recently, a small sample corrected version of Akaike's criterion, known as AICC, has received considerable attention (Hurvich and Tsai, 1989, Burnham and Anderson 2002). Despite their varied philosophical origins, the information criteria tend to have penalty terms that are a function of the number of parameters (d) and some measure of sample size (n). In some instances, such as REML estimation in the mixed model, it is not entirely clear how to best determine d and n , because the fixed effects are removed from the estimation. One approach considers the number of covariance parameters and the number of subjects in the analysis (PROC MIXED, PROC GLIMMIX default). Verbeke and Molenberghs (2000, Table 6.7, p. 74) and Vonesh and Chinchilli (1997, p. 263) count the rank of \mathbf{X} in d and determine n as the total sample size minus the rank of \mathbf{X} . You can obtain information criteria computed according to these definitions for REML estimation with the IC option in the GLIMMIX procedure. The most appropriate term for n is probably linked to the effective sample size, taking into account the covariance structure in the data. For example, if the data consist of s subjects whose observations are perfectly correlated, then $n = s$. If the observations among subjects were independent, then $n = \sum_{i=1}^s n_i$. Covariance structures in mixed models yield effective sample sizes between these extremes.

The use of information criteria with REML estimation has the same caveats as likelihood ratio testing based on the $-2 \text{ Res Log Likelihood}$. If two models have different fixed effects—whether they are nested or not—the $-2 \text{ Res Log Likelihoods}$ are not comparable. Consequently, the information criteria based on these $-2 \text{ Res Log Likelihoods}$ are also not comparable. Adding $2 \times p$ to numbers that are not comparable does not create a basis for comparison. This difficulty does not arise with ML estimation.

The lack of comparability in REML estimation has to do with the fact that the underlying data sets are not comparable. The same issue surfaces, but more severely, in generalized linear mixed models and nonlinear mixed models based on pseudo-data (PROC GLIMMIX, %NLINMIX). Because the fitting process consists of repeatedly fitting a linear mixed model constructed from pseudo-data, the likelihoods are not comparable in the models encountered during the model-fitting process. In addition, likelihoods based on pseudo-data from different models are not directly comparable, whether estimation of the linear mixed models relies on a maximum likelihood or restricted maximum likelihood principle. The NLMIXED procedure produces log likelihoods and information criteria for nonlinear or generalized linear mixed models that can be used for model comparisons.

In contrast to standard linear models, especially regression models, the graphical analysis of residuals is not used as frequently in mixed models. Chapter 10 discusses mixed model diagnostics and the various forms of residuals that arise in mixed models (for example, marginal and conditional residuals). You cannot distill the complexity of model–data interaction, and the adequacy of a model in a single summary measure such as the log likelihood or an information criterion. Residual graphics inform you about aspects of the model fit that summary measures cannot address—for example, the presence of outliers, heterogeneity of variance, the need for transformation, and so on.

A1.7 Inference and Test Statistics

A1.7.1 Inference about the Covariance Parameters

For inferences concerning the covariance parameters in your model, you can use likelihood-based statistics. One common such statistic is the **Wald Z**, which is computed as the parameter estimate divided by its estimated asymptotic standard error. The asymptotic standard errors are computed from the inverse of the second derivative matrix of the log likelihood with respect to the covariance parameters. The Wald Z test is valid for large samples, but it can be unreliable for small data sets and for parameters such as variance components that are known to have a skewed or bounded sampling distribution.

A better alternative is the likelihood ratio χ^2 . This test compares two models that are nested with respect to the covariance parameters. The two models are often referred to as the full model and the reduced model. The reduced model is obtained from the full model by imposing one or more constraints on the covariance parameters—a process known as nesting models. The test statistic is the difference of the $-2 \log$ likelihoods between the reduced and the full model. Issues surrounding likelihood ratio testing in mixed models were addressed in A1.6.1—for example, the need for models to have the same **X** matrix when $-2 \text{ Res Log Likelihoods}$ are used to form test statistics and the nonstandard distribution theory when parameters are on the boundary of the parameter space. In order to perform likelihood ratio tests for covariance parameters, you need to run PROC MIXED twice to obtain the -2 (Res) Log Likelihoods under the full and reduced model.

For variance component models, the null hypothesis of no effect can be tested using the fixed effects F -tests (Morgan and Gumpertz 1996, Pantula 1996, Khuri, Mathew, and Sinha 1998). Confidence intervals on variance components can be constructed in a variety of ways (Burdick and Graybill 1992, Burdick, Borror, and Montgomery 2005). Bayesian inference on variance components can be conducted by generating a sample from their posterior distribution as described in Chapter 13.

A1.7.2 Inference about Fixed and Random Effects

For inferences concerning the fixed and random effects parameters in the mixed model, consider estimable linear combinations of the following form:

$$\omega = \mathbf{L} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix}$$

Functions of this form are called predictable functions if the $\boldsymbol{\beta}$ portion of \mathbf{L} satisfies the estimability requirement (Searle 1971). Chapter 6 gives examples of the logic underlying the construction of predictable functions. Such a formulation in terms of a general \mathbf{L} matrix encompasses a wide variety of common inferential procedures such as those employed with Type I and III tests and least-squares means. The CONTRAST and ESTIMATE statements in PROC MIXED and PROC GLIMMIX allow you to specify your own \mathbf{L} matrices. Typically, inference on fixed effects is the focus, and in this case the \mathbf{u} portion of \mathbf{L} is assumed to contain all 0s.

Statistical inferences can be obtained by testing the null hypothesis

$$H : \mathbf{L} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \mathbf{0}$$

or by constructing point and/or interval estimates.

When \mathbf{L} consists of a single row, a general t -statistic can be constructed as follows:

$$t = \frac{\mathbf{L} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix}}{\sqrt{\mathbf{L} \hat{\mathbf{C}} \mathbf{L}'}} \quad (\text{A1.10})$$

See McLean and Sanders (1988) and Stroup (1989) for details. Under the assumed normality of \mathbf{u} and \mathbf{e} , t has an exact t -distribution only for data exhibiting certain types of balance and for some special unbalanced cases. In general, t is only approximately t -distributed, and its degrees of freedom must be estimated. See the DDFM= option in the MODEL statement for a description of the various degrees of freedom methods available in PROC MIXED.

If v denotes the degrees of freedom associated with ω , the associated confidence interval for ω is

$$\mathbf{L} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} \pm t_{v,\alpha/2} \sqrt{\mathbf{L} \hat{\mathbf{C}} \mathbf{L}'}$$

where $t_{v,\alpha/2}$ is the $(1-\alpha/2)100^{\text{th}}$ percentile of the t_v -distribution.

When the rank of \mathbf{L} is greater than 1, the following general F -statistic can be considered:

$$F = \frac{\left[\begin{matrix} \hat{\beta} \\ \hat{u} \end{matrix} \right] \mathbf{L}' (\mathbf{L} \hat{\mathbf{C}} \mathbf{L}')^{-1} \mathbf{L} \left[\begin{matrix} \hat{\beta} \\ \hat{u} \end{matrix} \right]}{\text{rank}(\mathbf{L})} \quad (\text{A1.11})$$

Analogous to t , F in general has an approximate F -distribution, with $\text{rank}(\mathbf{L})$ numerator degrees of freedom and \hat{v} denominator degrees of freedom.

As sample sizes grow large, equation (A1.10) has a normal distribution and equation (A1.11) has a chi-square distribution. You can base statistical inference in PROC MIXED on these asymptotic distributions with the CHISQ option in the MODEL and CONTRAST statements. In the GLIMMIX procedure you can choose DDFM=NONE in the MODEL statement to compute p -values according to the asymptotic distributions (or use the CHISQ options).

The t - and F -statistics allow you to make inferences about fixed effects that account for the variance-covariance model you select. An alternative for large samples is the χ^2 -statistic associated with the likelihood ratio test. This statistic compares two fixed effects models, one a special case of the other. It is computed just as when comparing different covariance models. However, you should not compare likelihoods of models that differ in their fixed effects when the covariance parameters are estimated by REML (see Section A1.6.1); use ML estimation instead. You can also use information-theoretical criteria to compare fixed effects using ML (see Section A1.6.2).

For variance component and random coefficient models, Bayesian inference on the fixed and random effects parameters can be conducted by generating a sample from their posterior distribution as described in Chapter 13.



Appendix 2: Data Sets

Note: This appendix contains partial listings of the data sets used in this book. You will find a complete listing of each data set on the CD-ROM that accompanies the book and on the SAS Press Companion Web site at support.sas.com/companionsites.

A2.2 Randomized Block Designs	759
Data Set 2.2 Bond.....	759
Data Set 2.5 PBIB	759
A2.3 Random Effects Models	759
Data Set 3.2 Mississippi River.....	759
Data Set 3.4 Semiconductor	760
Data Set 3.5 Genetics	760
A2.4 Analyzing Multi-level and Split-Plot Designs	761
Data Set 4.4 Semiconductor Split-Plot Experiment.....	761
Data Set 4.6 Variety-Pesticide Evaluation	761
Data Set 4.7 Mouse Condition–Diet Experiment.....	762
A2.5 Analysis of Repeated Measures Data.....	762
Data Set 5.2 Respiratory Ability	762
Data Set 5.4 Heart Rates	763
A2.6 Best Linear Unbiased Prediction	764
Data Set 6.4 Genetic Evaluation.....	764
Data Set 6.5 Machine Operators	764
Data Set 6.6 Multicenter Trial	764
A2.7 Analysis of Covariance	765
Data Set 7.3 Average Daily Gain	765
Data Set 7.4 Cooking Times	765
Data Set 7.5 Balanced Incomplete Block	765
Data Set 7.6 Unbalanced Incomplete Block.....	766
Data Set 7.7 Teaching Methods I.....	766

Data Set 7.8 Teaching Methods II.....	767
Data Set 7.9 Wafer Types.....	767
A2.8 Random Coefficient Models.....	768
Data Set 8.2 Winter Wheat	768
Data Set 8.3 Mathematics Study	768
Data Set 8.4 Pig Growth	768
A2.9 Heterogeneous Variance Models	769
Data Set 9.2 TV	769
Data Set 9.3 LR	769
Data Set 9.4 Sales	770
Data Set 9.5 Dial	770
Data Set 9.6 Grip Data.....	770
Data Set 9.7 Preetch Data.....	771
A2.10 Mixed Model Diagnostics.....	771
Data Set 10.2 Cultivar-Inoculation Trial.....	771
A2.11 Spatial Variability.....	772
Data Set 11.5 Uniformity Trial	772
Data Set 11.6 Hazardous Waste	772
Data Set 11.7 Alliance Wheat Yield.....	772
Data Set 11.8 C-N Ratios	773
A2.13 Some Bayesian Approaches to Mixed Models.....	773
Data Set 13.6 Beetle Fecundity.....	773
A2.14 Generalized Linear Mixed Models	774
Data Set 14.5 Clinics	774
Data Set 14.6 Seed Mix	774
A2.15 Nonlinear Mixed Models	775
Data Set 15.3 Orange Trees	775
Data Set 15.7 Phenobarbital	775
A2.16 Case Studies.....	776
Data Set 16.2 Milling Corn.....	776
Data Set 16.3 Mayonnaise.....	776
Data Set 16.4 Irrigation.....	776
Data Set 16.5 Nitrogen	777
Data Set 16.6 Sandhills	777
Data Set 16.7 Breaking Strength.....	777
Data Set 16.8 2 ³ Factorial	778
Data Set 16.9 Edge Slope	778
Data Set 16.10 Use Products	778
Data Set 16.11 CD4 Count	779
Data Set 16.12 Microarray.....	779

A2.2 Randomized Block Designs

Data Set 2.2 Bond

Obs	ingot	metal	pres
1	1	n	67.0
2	1	i	71.9
3	1	c	72.2
4	2	n	67.5
5	2	i	68.8
6	2	c	66.4
:	:	:	:
19	7	n	75.6
20	7	i	84.9
21	7	c	69.0

Data Set 2.5 PBIB

Obs	blk	trt1	y1	trt2	y2	trt3	y3	trt4	y4
1	1	15	2.4	9	2.5	1	2.6	13	2.0
2	2	5	2.7	7	2.8	8	2.4	1	2.7
3	3	10	2.6	1	2.8	14	2.4	2	2.4
4	4	15	3.4	11	3.1	2	2.1	3	2.3
:	:	:	:	:	:	:	:	:	:
13	13	5	3.0	9	3.6	11	3.2	12	3.2
14	14	7	3.0	13	2.8	14	2.4	11	2.5
15	15	10	2.4	4	2.5	8	3.2	11	3.1

A2.3 Random Effects Models

Data Set 3.2 Mississippi River

Obs	influent	N2	type
1	1	21	2
2	1	27	2
3	1	29	2
4	1	17	2
5	1	19	2
6	1	12	2
7	1	29	2
8	1	20	2
9	1	20	2
10	2	21	2
11	2	11	2
:	:	:	:

Obs	influent	N2	type
35	6	35	3
36	6	34	3
37	6	30	3

Data Set 3.4 Semiconductor

Obs	source	lot	wafer	site	Thick
1	1	1	1	1	2006
2	1	1	1	2	1999
3	1	1	1	3	2007
4	1	1	2	1	1980
5	1	1	2	2	1988
6	1	1	2	3	1982
7	1	1	3	1	2000
8	1	1	3	2	1998
9	1	1	3	3	2007
10	1	2	1	1	1991
11	1	2	1	2	1990
:	:	:	:	:	:
34	1	4	3	1	1987
35	1	4	3	2	1990
36	1	4	3	3	1995
37	2	5	1	1	2013
38	2	5	1	2	2004
39	2	5	1	3	2009
:	:	:	:	:	:
68	2	8	2	2	1993
69	2	8	2	3	1996
70	2	8	3	1	1990
71	2	8	3	2	1989
72	2	8	3	3	1992

Data Set 3.5 Genetics

Obs	loc	block	fam	Yield
1	1	1	1	268
2	1	2	1	279
3	1	3	1	261
4	1	1	2	242
5	1	2	2	261
6	1	3	2	258
7	1	1	3	242
8	1	2	3	245
9	1	3	3	234
:	:	:	:	:

Obs	loc	block	fam	Yield
55	4	1	4	180
56	4	2	4	195
57	4	3	4	193
58	4	1	5	199
59	4	2	5	183
60	4	3	5	208

A2.4 Analyzing Multi-level and Split-Plot Designs

Data Set 4.4 Semiconductor Split-Plot Experiment

Obs	resistance	et	wafer	pos
1	5.22	1	1	1
2	5.61	1	1	2
3	6.11	1	1	3
4	6.33	1	1	4
5	6.13	1	2	1
6	6.14	1	2	2
7	5.60	1	2	3
8	5.91	1	2	4
9	5.49	1	3	1
10	4.60	1	3	2
11	4.95	1	3	3
12	5.42	1	3	4
⋮	⋮	⋮	⋮	⋮
45	6.05	4	3	1
46	6.15	4	3	2
47	5.55	4	3	3
48	6.13	4	3	4

Data Set 4.6 Variety-Pesticide Evaluation

Obs	block	type	dose	logdose	y
1	1	r	1	0	15.7
2	2	r	1	0	23.1
3	3	r	1	0	15.9
4	4	r	1	0	20.8
5	5	r	1	0	24.5
6	1	r	2	1	25.1
7	2	r	2	1	29.2
8	3	r	2	1	29.7
9	4	r	2	1	28.6

Obs	block	type	dose	logdose	y
10	5	r	2	1	26.6
:	:	:	:	:	:
36	1	s	8	3	22.8
37	2	s	8	3	33.0
38	3	s	8	3	25.2
39	4	s	8	3	27.2
40	5	s	8	3	20.8

Data Set 4.7 Mouse Condition–Diet Experiment

Obs	cage	condition	diet	gain	
1	1		1	normal	58
2	1		1	restrict	58
3	1		1	suppleme	58
4	1		2	normal	54
5	1		2	restrict	46
6	1		2	suppleme	57
:	:		:	:	:
31	6		3	normal	65
32	6		3	restrict	64
33	6		3	suppleme	54
34	6		4	normal	59
35	6		4	restrict	47
36	6		4	suppleme	73

A2.5 Analysis of Repeated Measures Data

Data Set 5.2 Respiratory Ability

Obs	PATIENT	BaseFEV1	FEV11H	FEV12H	FEV13H	FEV14H	FEV15H	FEV16H	FEV17H	FEV18H	Drug
47	224	3.47	4.27	4.50	4.34	4.00	4.11	3.93	3.68	3.77	c
48	232	2.79	4.10	3.85	4.27	4.01	3.78	3.14	3.94	3.69	c
49	201	2.14	2.36	2.36	2.28	2.35	2.31	2.62	2.12	2.42	p
50	202	3.37	3.03	3.02	3.19	2.98	3.01	2.75	2.70	2.84	p
:	:	:	:	:	:	:	:	:	:	:	:
71	224	3.66	3.98	3.77	3.65	3.81	3.77	3.89	3.63	3.74	p
72	232	2.88	3.04	3.00	3.24	3.37	2.69	2.89	2.89	2.76	p

Data Set 5.4 Heart Rates

Obs	patient	drug	basehr	minute	time	HR
1	201	p	92	1	1	76
2	201	p	92	5	5	84
3	201	p	92	15	15	88
4	201	p	92	30	30	96
5	201	p	92	60	60	84
6	202	b	54	1	1	58
7	202	b	54	5	5	60
8	202	b	54	15	15	60
9	202	b	54	30	30	60
10	202	b	54	60	60	64
:	:	:	:	:	:	:
116	232	a	78	1	1	72
117	232	a	78	5	5	72
118	232	a	78	15	15	78
119	232	a	78	30	30	80
120	232	a	78	60	60	68

A2.6 Best Linear Unbiased Prediction

Data Set 6.4 Genetic Evaluation

Obs	sire	dam	adg
1	1	1	2.24
2	1	1	1.85
3	1	2	2.05
4	1	2	2.41
5	2	1	1.99
:	:	:	:
19	5	2	2.58
20	5	2	2.56

Data Set 6.5 Machine Operators

Obs	machine	operator	y
1	1	1	51.43
2	1	1	51.28
3	1	2	50.93
4	1	2	50.75
5	1	3	50.47
6	1	3	50.83
7	2	1	51.91
8	2	1	52.43
9	2	2	52.26
10	2	2	52.33
11	2	3	51.58
12	2	3	51.23

Data Set 6.6 Multicenter Trial

Obs	location	block	treatment	response
1	A	1	3	3.13
2	A	1	4	3.22
3	A	1	2	3.16
4	A	1	1	3.25
5	A	2	2	2.71
6	A	2	1	3.20
7	A	2	3	2.99
8	A	2	4	2.96
9	A	3	1	2.69
:	:	:	:	:
105	I	3	2	2.66

106	I	3	1	2.31
107	I	3	4	2.51
108	I	3	3	2.65

A2.7 Analysis of Covariance**Data Set 7.3** Average Daily Gain

id	blk	trt	adg	iwt
1	1	0	1.03	338
2	1	10	1.54	477
3	1	20	1.82	444
4	1	30	1.86	370
5	2	0	1.31	403
⋮	⋮	⋮	⋮	⋮
29	8	0	0.18	315
30	8	10	0.64	376
31	8	20	0.76	308
32	8	30	0.70	439

Data Set 7.4 Cooking Times

Obs	block	container	amount	time
1	1	A	2	71.36
2	1	A	4	123.33
3	1	A	6	155.59
4	1	A	8	194.90
5	1	B	4	118.64
⋮	⋮	⋮	⋮	⋮
25	3	A	2	96.46
26	3	C	4	145.55

Data Set 7.5 Balanced Incomplete Block

id	blk	trt	y	x	grp
1	1	1	31	20	13
2	1	2	29	18	24
3	1	3	31	11	13
4	2	1	29	37	13
5	2	2	34	37	24
6	2	4	33	39	24
7	3	1	31	29	13
⋮	⋮	⋮	⋮	⋮	⋮

id	blk	trt	y	x	grp
22	8	2	27	13	24
23	8	3	37	39	13
24	8	4	29	21	24

Data Set 7.6 Unbalanced Incomplete Block

id	blk	trt	y	x
1	1	1	0.62	0.078
2	1	2	0.91	0.010
3	2	1	0.41	0.032
4	2	2	0.48	0.050
5	3	1	0.41	0.000
6	3	2	0.49	0.015
7	4	1	0.26	0.010
⋮	⋮	⋮	⋮	⋮
22	11	4	0.27	0.062
23	12	3	0.24	0.058
24	12	4	0.23	0.082

Data Set 7.7 Teaching Methods I

id	met	teacher	gen	student	score	y_ex
1	1		1 f		15	11
2	1		1 f	2	17	11
3	1		1 f	3	16	11
4	1		1 f	4	16	11
5	1		1 m	1	17	11
6	1		1 m	2	16	11
7	1		1 m	3	17	11
8	1		1 m	4	17	11
9	1		2 f	1	18	8
⋮	⋮		⋮ ⋮	⋮	⋮	⋮
89	3		4 f	1	28	6
90	3		4 f	2	27	6
91	3		4 f	3	28	6
92	3		4 f	4	30	6
93	3		4 m	1	25	6
94	3		4 m	2	25	6
95	3		4 m	3	22	6
96	3		4 m	4	25	6

Data Set 7.8 Teaching Methods II

id	met	teacher	gen	iq	score
1	1		f	89	54
2	1		f	105	55
3	1		f	108	54
4	1		f	116	64
5	1		m	95	59
6	1		m	103	58
7	1		m	91	42
8	1		m	82	48
9	1	2	f	83	48
:	:	:	:	:	:
93	3		m	80	68
94	3		m	97	81
95	3		m	114	89
96	3		m	100	87

Data Set 7.9 Wafer Types

id	grp	temp	type	wafer	site	delta	thick
1	1	900	A	1	1	291	1919
2	1	900	A	1	2	295	1919
3	1	900	A	1	3	294	1919
4	1	900	A	2	1	318	2113
5	1	900	A	2	2	315	2113
6	1	900	A	2	3	315	2113
7	1	900	B	1	1	349	1965
:	:	:	:	:	:	:	:
31	1	1100	B	1	1	352	2086
32	1	1100	B	1	2	353	2086
33	1	1100	B	1	3	350	2086
34	1	1100	B	2	1	330	1899
35	1	1100	B	2	2	330	1899
36	1	1100	B	2	3	334	1899
37	2	900	A	1	1	306	1841
:	:	:	:	:	:	:	:
139	4	1100	B	1	1	335	2174
140	4	1100	B	1	2	339	2174
141	4	1100	B	1	3	338	2174
142	4	1100	B	2	1	304	1802
143	4	1100	B	2	2	303	1802
144	4	1100	B	2	3	303	1802

A2.8 Random Coefficient Models

Data Set 8.2 Winter Wheat

Obs	id	variety	yield	moist
1	1	1	41	10
2	2	1	69	57
3	3	1	53	32
4	4	1	66	52
5	5	1	64	47
6	6	1	64	48
7	7	2	49	30
:	:	:	:	:
58	58	10	71	53
59	59	10	67	48
60	60	10	74	59

Data Set 8.3 Mathematics Study

```

data class;
  input size@@;
  retain class 1;
  do i = 1 to size;
    output;
  end;
  drop size i;
  class + 1;
  datalines;
21 29 16 24 20 19
16 8 19 18 14 14
13 19 22 17 11 9
22 18 19 24 30 20
16 21 29 22 18 19
<more datalines...>
run;

data sims;
  input pretot gain @@;
  datalines;
29 2 38 0 31 6 31 6 29 5 23 9 23 7
33 2 30 1 32 3 22 4 29 6 34 2 30 -1
35 1 25 1 22 3 31 1 33 3 31 1 35 2
20 2 18 -3 12 5 9 1 11 -3 12 3 12 -8
18 -6 13 -4 8 4 21 1 4 4 10 1 8 4
12 -6 14 -2 9 2 11 3 11 -5 12 7 12 2
<more datalines...>
run;

```

Data Set 8.4 Pig Growth

Obs	trt	pig	day0	day6	day12	day18	day24	day30
1	1	1	14.0	22.1	27.7	31.8	35.3	32.6
2	1	2	15.2	23.8	32.6	40.0	42.4	44.6
3	1	3	13.9	22.1	28.1	29.8	33.6	32.9
:	:	:	:	:	:	:	:	:
19	1	19	15.7	22.8	31.7	38.1	40.2	40.2
20	1	20	10.9	23.4	30.0	35.6	37.4	38.7
21	2	1	15.1	27.7	35.1	37.2	45.0	41.2
:	:	:	:	:	:	:	:	:
58	3	18	13.7	20.5	28.0	30.1	33.1	35.1
59	3	19	16.0	25.2	31.3	37.7	39.6	43.2
60	3	20	15.8	20.6	29.7	31.2	32.2	33.2

A2.9 Heterogeneous Variance Models**Data Set 9.2** TV

Obs	child	age	sex	time	age_sex
1	1	2	F	12	2F
2	2	3	M	17	3M
3	3	2	M	10	2M
4	4	4	M	24	4M
5	5	2	M	13	2M
:	:	:	:	:	:
25	25	4	F	12	4F
26	26	4	M	49	4M
27	27	2	M	15	2M

Data Set 9.3 LR

Obs	x	y
1	0.1	4
2	0.4	9
3	0.7	6
4	0.8	10
:	:	:
16	3.0	22
17	3.2	34
18	3.3	42

Data Set 9.4 Sales

Obs	region	dist	s1	s2	s3	s4	s5	s6	s7	s8
1	NC	1	42.9	47.4	41.0	46.6	42.9	39.8	43.4	.
2	NC	2	38.7	47.7	42.3	42.8	42.0	41.4	42.7	41.5
3	NC	3	37.6	43.5	37.8	39.0	39.5	42.2	39.0	45.7
4	NE	1	42.4	44.6	36.9	43.8	37.4	43.4	44.4	37.7
5	NE	2	41.5	41.2	41.4	41.7	41.3	41.8	.	.
:	:	:	:	:	:	:	:	:	:	:
27	SW	1	41.3	39.2	38.0	38.0	33.9	.	.	.
28	SW	2	49.2	47.4	45.8	52.1	48.5	.	.	.
29	SW	3	36.9	31.3	32.2	36.5	39.8	30.3	36.6	34.0

Data Set 9.5 Dial

Obs	sub	qb	tmp	ufr	index	type
1	1	200	0.240	0.645	1	1
2	1	200	0.505	20.115	2	1
3	1	200	0.995	38.460	3	1
4	1	200	1.485	44.985	4	1
5	1	200	2.020	51.765	5	1
6	1	200	2.495	46.575	6	1
7	1	200	2.970	40.815	7	1
8	2	200	0.240	3.720	1	1
:	:	:	:	:	:	:
148	22	300	0.250	.	1	2
149	22	300	0.500	.	2	2
150	22	300	1.000	.	3	2
151	22	300	1.500	.	4	2
152	22	300	2.000	.	5	2
153	22	300	2.500	.	6	2
154	22	300	3.000	.	7	2

Data Set 9.6 Grip Data

Obs	subject	x	trt	Gender	time	t	y
1		26	175	1	M	1	161
2		26	175	1	M	2	210
3		26	175	1	M	3	230
:	:	:	:	:	:	:	:
38		79	225	1	M	2	250
39		79	225	1	M	3	235
40		1	120	2	M	1	130
:	:	:	:	:	:	:	:
86		75	265	2	M	2	255
87		75	265	2	M	3	270

Obs	subject	x	trt	Gender	time	t	y
88	2	80	1	F		1	1
89	2	80	1	F		2	2
:	:	:	:	:		:	:
143	70	34	1	F		2	2
144	70	34	1	F		3	3
145	3	60	2	F		1	1
146	3	60	2	F		2	2
:	:	:	:	:		:	:
198	71	104	2	F		3	3
199	72	60	2	F		1	1
200	72	60	2	F		2	2
201	72	60	2	F		3	3
							58

Data Set 9.7 Preetch Data

Obs	expt	wafer	mask	viscos	spin	baketemp	baketime	aperture	expos	develop	etch	y1	y2	y3	y4	y5
1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	2.43	2.52	2.63	2.52	2.50
2	1	2	-1	-1	-1	-1	-1	-1	-1	-1	-1	2.36	2.50	2.62	2.43	2.49
3	2	1	-1	-1	0	-1	0	0	0	0	0	2.76	2.66	2.74	2.60	2.53
4	2	2	-1	-1	0	-1	0	0	0	0	0	2.66	2.73	2.95	2.57	2.64
5	3	1	-1	-1	1	-1	1	1	1	1	1	2.82	2.71	2.78	2.55	2.36
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
32	17	2	1	-1	0	1	-1	1	-1	0	1	3.12	2.97	3.18	3.03	2.95
33	18	1	1	-1	1	1	0	-1	0	1	-1	3.46	3.49	3.50	3.45	3.57

A2.10 Mixed Model Diagnostics**Data Set 10.2 Cultivar-Inoculation Trial**

Obs	block	cult	inoc	drywt
1	1	a	con	27.4
2	1	a	dea	29.7
3	1	a	liv	34.5
4	1	b	con	29.4
5	1	b	dea	32.5
6	1	b	liv	34.4
:	:	:	:	:
22	4	b	con	26.8
23	4	b	dea	28.6
24	4	b	liv	30.7

A2.11 Spatial Variability

Data Set 11.5 Uniformity Trial

Obs	rep	bloc	row	col	yield
1	1	4	1	1	10.5411
2	1	4	1	2	8.5806
3	1	2	1	3	11.2790
:	:	:	:	:	:
15	1	1	4	3	7.2836
16	1	1	4	4	8.0018
17	2	7	5	1	10.3349
:	:	:	:	:	:
62	4	14	8	6	10.5104
63	4	15	8	7	12.6808
64	4	15	8	8	10.4482

Data Set 11.6 Hazardous Waste

Obs	easting	northing	logt	salt	xxx
1	16.442	18.128	-6.02895	28.6	H-1
2	15.670	18.095	-6.20046	26.9	H-2b1
3	16.748	17.339	-5.60886	-0.6	H-3b1
:	:	:	:	:	:
28	16.740	26.145	-6.60227	0.0	WIPP-30
29	16.715	18.402	-6.29637	26.7	ERDA-9
30	24.145	25.825	-6.55349	18.7	AEC-7

Data Set 11.7 Alliance Wheat Yield

Obs	name	entry	plot	rawyld	rep	nloc	y	lat	lng
1	LANCER	1	1101	585	1	4	29.25	4.3	19.2
2	BRULE	2	1102	631	1	4	31.55	4.3	20.4
3	REDLAND	3	1103	701	1	4	35.05	4.3	21.6
:	:	:	:	:	:	:	:	:	:
55	NE87619	55	1155	605	1	4	30.25	17.2	4.8
56	NE87627	56	1156	403	1	4	20.15	17.2	6.0
57	CENTURA	9	2101	556	2	4	27.80	17.2	8.4
:	:	:	:	:	:	:	:	:	:
222	NE84557	13	4154	531	4	4	26.55	47.3	24.0
223	TAM 107	28	4155	512	4	4	25.60	47.3	25.2
224	HOMESTEAD	22	4156	538	4	4	26.90	47.3	26.4

Data Set 11.8 C-N Ratios

Obs	x	y	TN	TC	CN
1	0	35.0	0.0810	0.9225	11.4201
2	0	40.0	0.0790	0.8310	10.5300
3	0	80.0	0.0800	0.8460	10.5750
⋮	⋮	⋮	⋮	⋮	⋮
10	0	295.0	0.0610	0.6410	10.5113
11	0	300.0	0.0580	0.5925	10.2155
12	5	35.0	0.0880	0.9460	10.7640
⋮	⋮	⋮	⋮	⋮	⋮
193	500	247.5	0.0595	0.6925	11.6414
194	500	295.0	0.0695	0.7425	10.6915
195	500	300.0	0.0765	0.8735	11.4545
196	0	5.0	0.1290	.	.
197	0	10.0	0.1120	.	.
198	0	50.0	0.0755	.	.
⋮	⋮	⋮	⋮	⋮	⋮
393	500	215.0	0.0655	.	.
394	500	277.5	0.0640	.	.
395	500	282.5	0.0630	.	.

A2.13 Some Bayesian Approaches to Mixed Models**Data Set 13.6 Beetle Fecundity**

Obs	sire	dam	progeny	value	trait
1	1	1	1	58.0	fec
2	1	1	1	4.6	mass
3	1	1	2	64.0	fec
4	1	1	2	4.0	mass
5	1	1	3	50.0	fec
6	1	1	3	4.1	mass
7	1	1	4	53.0	fec
8	1	1	4	4.2	mass
9	1	1	5	50.0	fec
10	1	1	5	4.0	mass
11	1	2	1	54.0	fec
⋮	⋮	⋮	⋮	⋮	⋮
48	1	5	4	3.5	mass
49	1	5	5	40.0	fec
50	1	5	5	3.7	mass
51	2	1	1	62.0	fec

Obs	sire	dam	progeny	value	trait
52	2	1	1	4.7	mass
53	2	1	2	52.0	fec
:	:	:	:	:	:
1096	24	4	4	3.9	mass
1097	24	4	5	80.0	fec
1098	24	4	5	4.6	mass

A2.14 Generalized Linear Mixed Models

Data Set 14.5 Clinics

Obs	clinic	trt	fav	unfav	nij
1	1	drug	11.0000	25	36
2	1	cntl	10.0000	27	37
3	2	drug	16.0000	4	20
4	2	cntl	22.0000	10	32
:	:	:	:	:	:
15	8	drug	4.0000	2	6
16	8	cntl	6.0000	1	7

Data Set 14.6 Seed Mix

Obs	trt	blk	mix	count	y
1	1	1	1	24	25
2	1	1	2	12	13
3	1	1	3	8	9
4	1	1	4	13	14
5	1	2	1	9	10
:	:	:	:	:	:
110	7	4	2	56	57
111	7	4	3	26	27
112	7	4	4	27	28

A2.15 Nonlinear Mixed Models

Data Set 15.3 Orange Trees

Obs	tree	day	y
1	1	118	30
2	1	484	58
3	1	664	87
4	1	1004	115
5	1	1231	120
6	1	1372	142
7	1	1582	145
8	2	118	33
:	:	:	:
33	5	1231	142
34	5	1372	174
35	5	1582	177

Data Set 15.7 Phenobarbital

Obs	indiv	time	dose	weight	apgar	conc
1	1	0.0	25.0	1.4	7	.
2	1	2.0	0.0	1.4	7	17.3
3	1	12.5	3.5	1.4	7	.
:	:	:	:	:	:	:
11	1	108.5	3.5	1.4	7	.
12	1	112.5	0.0	1.4	7	31.0
13	2	0.0	15.0	1.5	9	.
:	:	:	:	:	:	:
741	59	120.5	3.0	1.1	6	.
742	59	132.3	3.0	1.1	6	.
743	59	144.8	3.0	1.1	6	.
744	59	146.8	0.0	1.1	6	40.2

A2.16 Case Studies

Data Set 16.2 Milling Corn

Obs	batch	a	b	c	d	y	aa
1	1	1	1	1	1	505	1
2	1	1	-1	-1	-1	493	1
3	1	1	-1	1	-1	491	1
4	2	1	1	-1	0	498	1
:	:	:	:	:	:	:	:
28	10	-1	-1	-1	1	494	-1
29	10	-1	1	-1	-1	497	-1
30	10	-1	-1	-1	-1	495	-1

Data Set 16.3 Mayonnaise

Obs	y	a	b	c	p	d	u	h	t
1	95	1	1	1	X6	0	-1.0000	1	-1.0000
2	67	1	1	1	X6	1	-1.0000	1	-1.0000
3	100	1	1	1	X6	0	-0.5854	2	-0.5854
4	70	1	1	1	X6	1	-0.5854	2	-0.5854
5	100	1	1	1	X6	0	1.0000	3	1.0000
6	77	1	1	1	X6	1	1.0000	3	1.0000
7	41	1	-1	-1	X7	0	-1.0000	1	-1.0000
:	:	:	:	:	:	:	:	:	:
106	12	-1	-1	-1	X23	1	-0.5854	2	-0.5854
107	18	-1	-1	-1	X23	0	1.0000	3	1.0000
108	11	-1	-1	-1	X23	1	1.0000	3	1.0000

Data Set 16.4 Irrigation

Obs	eu	bloc	trt	level	y
1	1	1	1	1	43
2	2	1	2	1	45
3	3	1	1	2	41
4	4	1	2	2	45
:	:	:	:	:	:
71	71	4	1	8	40
72	72	4	3	8	38

Data Set 16.5 Nitrogen

Obs	row	col	N	G	y
1	1	1	1	1	20.1
2	1	1	1	2	20.4
3	2	1	2	1	16.2
:	:	:	:	:	:
29	3	5	5	1	17.0
30	3	5	5	2	10.9

Data Set 16.6 Sandhills

Obs	circle	ec	ir	v	y
1	1	1	1	1	30.0
2	1	1	1	2	40.9
3	1	1	2	1	38.9
4	1	1	2	2	38.2
5	1	2	1	1	41.8
:	:	:	:	:	:
22	3	2	1	2	41.0
23	3	2	2	1	33.2
24	3	2	2	2	34.9

Data Set 16.7 Breaking Strength

Obs	soap_pH	num_washes	water_temp	dry_cycle	Breaking_Strength	Elongation	Slope	Energy	pH	w
1	8	10	cold	air	0.11056	16.8780	0.045683	0.67203	8	10
2	8	10	cold	delicate	0.111170	18.5470	0.041152	0.72147	8	10
3	8	10	cold	normal	0.10799	19.1110	0.036804	0.72370	8	10
4	8	10	hot	air	0.10061	19.5392	0.039650	0.68351	8	10
:	:	:	:	:	:	:	:	:	:	:
29	8	50	hot	delicate	0.07928	15.8182	0.027962	0.42389	8	50
30	8	50	hot	normal	0.07511	15.2064	0.027636	0.39166	8	50
31	10	10	cold	air	0.09464	15.3838	0.040991	0.51177	10	10
:	:	:	:	:	:	:	:	:	:	:
58	10	50	hot	air	0.05022	11.3640	0.024804	0.23344	10	50
59	10	50	hot	delicate	0.04091	11.1683	0.020217	0.17355	10	50
60	10	50	hot	normal	0.03404	11.0530	0.018532	0.14136	10	50

Data Set 16.8 2^3 Factorial

Obs	rep	blk	a	b	c	y	abc
1	1	1	-1	-1	-1	117	-1
2	1	1	-1	1	1	130	-1
3	1	1	1	-1	1	122	-1
4	1	1	1	1	-1	113	-1
5	1	2	-1	-1	1	123	1
:	:	:	:	:	:	:	:
46	6	2	-1	1	-1	122	1
47	6	2	1	-1	-1	123	1
48	6	2	1	1	1	124	1

Data Set 16.9 Edge Slope

Obs	rep	blk	a	b	c	edslp
1	1	1	-1	-1	-1	3.6
2	1	1	-1	1	1	4.5
3	1	1	1	-1	1	4.2
4	1	1	1	1	-1	4.0
5	1	2	-1	-1	1	5.1
6	1	2	-1	1	-1	3.9
7	1	2	1	-1	-1	4.5
8	1	2	1	1	1	4.9
9	2	1	-1	-1	-1	5.2
:	:	:	:	:	:	:
38	5	2	-1	1	1	5.5
39	5	2	1	-1	-1	4.6
40	5	2	1	1	1	5.2

Data Set 16.10 Use Products

Obs	blk	seq	person	I1	I2	I3	I4	period	prod	priorprd	time	y
1	1	1	101	0	0	0	0	1	1	0	1	5
2	1	1	101	0	0	0	0	1	1	0	2	4
3	1	1	101	0	0	0	0	1	1	0	3	3
4	1	1	101	1	0	0	0	2	2	1	1	5
5	1	1	101	1	0	0	0	2	2	1	2	5
6	1	1	101	1	0	0	0	2	2	1	3	5
7	1	1	101	0	1	0	0	3	3	2	1	5
8	1	1	101	0	1	0	0	3	3	2	2	4
9	1	1	101	0	1	0	0	3	3	2	3	4
10	1	2	102	0	0	0	0	1	1	0	1	4
:	:	:	:	:	:	:	:	:	:	:	:	:
214	4	6	406	0	1	0	0	3	3	2	1	3

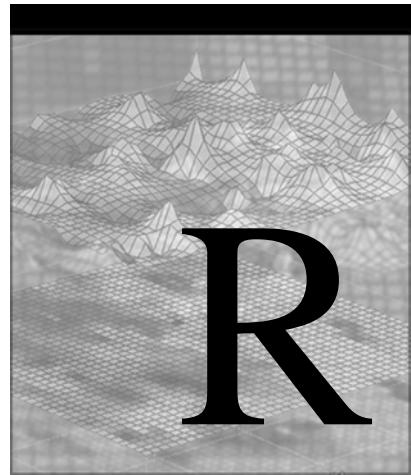
Obs	blk	seq	person	I1	I2	I3	I4	period	prod	priorprd	time	y
215	4	6	406	0	1	0	0	3	3	2	2	3
216	4	6	406	0	1	0	0	3	3	2	3	3

Data Set 16.11 CD4 Count

Obs	randgrp	stratum	unit	v1	v2	v3	v4
1	2	2	2	114	.	71	89
2	1	2	2	40	.	66	21
3	1	2	3	12	13	38	.
4	2	1	3	15	21	7	3
5	1	1	3	53	74	74	45
6	2	1	3	21	.	.	.
:	:	:	:	:	:	:	:
465	1	1	18	10	6	12	.
466	2	2	18	9	10	.	.
467	2	1	18	30	5	.	.

Data Set 16.12 Microarray

Obs	array	dye	trt	row	pin	dip	gene	log_intensity
1	1	g	0	1	1	1	1	0.1907
2	1	g	0	1	1	2	2	-0.0829
3	1	g	0	1	3	3	1	0.3392
4	1	g	0	1	3	4	3	0.3193
5	1	g	0	2	2	1	3	0.2378
:	:	:	:	:	:	:	:	:
31	1	g	0	8	4	3	8	0.4864
32	1	g	0	8	4	4	2	0.3080
33	1	r	1	1	1	1	1	-0.0668
34	1	r	1	1	1	2	2	-0.0327
:	:	:	:	:	:	:	:	:
62	1	r	1	8	2	2	3	-0.0020
63	1	r	1	8	4	3	8	0.2742
64	1	r	1	8	4	4	2	0.6028
65	2	g	1	1	1	1	1	-0.0688
:	:	:	:	:	:	:	:	:
382	6	r	0	8	2	2	5	0.0531
383	6	r	0	8	4	3	7	-0.2809
384	6	r	0	8	4	4	4	-0.2613



References

- Abrams, D. I., A. I. Goldman, C. Launer, J. A. Korvick, J. D. Neaton, L. R. Crane, M. Grodesky, S. Wakefield, K. Muth, S. Kornegay, D. L. Cohn, A. Harris, R. Luskin-Hawk, N. Markowitz, J. H. Sampson, M. Thompson, L. Deyton, and the Terry Beirn Community Programs for Clinical Research on AIDS. 1994. "Comparative Trial of Didanosine and Zalcitabine in Patients with Human Immunodeficiency Virus Infection Who Are Intolerant of or Have Failed Zidovudine Therapy." *New England Journal of Medicine* 330:657–662.
- Aitkin, M. 1987. "Modeling Variance Heterogeneity in Normal Regressing Using GLIM." *Applied Statistics* 36:332–339.
- Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AIC-19:716–723.
- Allard, R. W. 1966. *Principles of Plant Breeding*. New York: John Wiley & Sons.
- Allison, D. B., G. P. Page, T. M. Beasley, and J. W. Edwards. 2006. *DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments*. New York: Chapman and Hall.
- Bates, D., and D. Watts. 1988. *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley & Sons.
- Bayarri, M. J., and J. O. Berger. 2000. "P Values for Composite Null Models." *Journal of the American Statistical Association* 95:1127–1142.
- Beal, S. L., and L. B. Sheiner. 1982. "Estimating Population Kinetics." *CRC Crit. Rev. Biomed. Eng.* 8:195–222.
- Beal, S. L., and L. B. Sheiner. 1988. "Heteroskedastic Nonlinear Regression." *Technometrics* 30:327–338.
- Beal, S. L., and L. B. Sheiner, eds. 1992. *NONMEM User's Guide*. San Francisco: University of California–San Francisco, NONMEM Project Group.
- Beckman, R. J., C. J. Nachtsheim, and D. R. Cook. 1987. "Diagnostics for Mixed-Model Analysis of Variance." *Technometrics* 29:413–426.
- Beitler, P. J., and J. R. Landis. 1985. "A Mixed-Effects Model for Categorical Data." *Biometrics* 41:991–1000.
- Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
- Berger, J. O. 2003. "Could Fisher, Jeffreys, and Neyman Have Agreed on Testing?" (with comments). *Statistical Science* 18:1–32.

- Berger, J. O., and M. Delampady. 1987. "Testing Precise Hypotheses" (with comments). *Statistical Science* 2:317–352.
- Berger, J. O., and T. Sellke. 1987. "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence (with comments)." *Journal of the American Statistical Association* 82:112–139.
- Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.
- Bjerke, Frøydis, Are Halvor Aastviet, W. W. Stroup, Bente Kirkhus, and Tormod Næs. 2004. "Design and Analysis of Storing Experiments: A Case Study." *Journal of Quality Engineering* 16(4): 591–611.
- Bohning, D., E. Dietz, P. Schlattman, L. Mendonca, and U. Kirchner. 1999. "The Zero-Inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology." *Journal of the Royal Statistical Society, Series A*, 162:195–209.
- Box, G. E. P. 1950. "Problems in the Analysis of Growth and Wear Curves." *Biometrics* 6:362–389.
- Box, G. E. P., and D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B*, 26:211–252.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for Experimenters*. New York: John Wiley & Sons.
- Box, G. E. P., and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Wiley Classics Library Edition, published 1992. New York: John Wiley & Sons.
- Bozdogan, H. 1987. "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions." *Psychometrika* 52:345–370.
- Breslow, N. R., and D. G. Clayton. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88:9–25.
- Brown, H., and R. Prescott. 1999. *Applied Mixed Models in Medicine*. New York: John Wiley & Sons.
- Brown, R. L., J. Durbin, and J. M. Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships over Time." *Journal of the Royal Statistical Society, Series B*, 37:149–192.
- Brownie, C., and M. L. Gumpertz. 1997. "Validity of Spatial Analyses of Large Field Trials." *Journal of Agricultural, Biological, and Environmental Statistics* 2(1): 1–23.
- Brownie, C., D. T. Bowman, and J. W. Burton. 1993. "Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods." *Agronomy Journal* 85: 1244–1253.
- Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Burdick, R. K., and F. A. Graybill. 1992. *Confidence Intervals on Variance Components*. New York: Marcel Dekker.
- Burdick, R. K., and F. A. Graybill. 1992. *Confidence Intervals on Variance Components*. New York: Marcel Dekker.
- Burdick, R. K., C. M. Borror, and D. C. Montgomery. 2005. *Design and Analysis of Gauge R & R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*. ASA-SIAM Series on Statistics and Applied Probability. Philadelphia: SIAM; Alexandria, VA: ASA.

- Burnham, K. P., and D. R. Anderson. 1998. *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.
- Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis for Count Data*. Cambridge, UK: Cambridge University Press.
- Carlin, B. P., and T. A. Louis. 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Carroll, R. J., and D. Ruppert. 1988. *Transformation and Weighting in Regression*. London: Chapman & Hall.
- Casella, G., and R. L. Berger. 1987. "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem." *Journal of the American Statistical Association* 82:106–111.
- Castelloe, J. M., and R. G. O'Brien. 2000. "Power and Sample Size Determination for Linear Models." *Proceedings of the Twenty-sixth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute.
- Chilès, J.-P., and P. Delfiner. 1999. *Geostatistics: Modeling Spatial Uncertainty*. New York: John Wiley & Sons.
- Christensen, R., L. M. Pearson, and W. Johnson. 1992. "Case-Deletion Diagnostics for Mixed Models." *Technometrics* 34:38–45.
- Chu, T.-M., B. Weir, and R. D. Wolfinger. 2004. "Comparison of Li-Wong and Mixed Model Approaches to Oligonucleotide Array Data Analysis." *Bioinformatics* 20(4): 500–506.
- Cochran, W. G., and G. M. Cox. 1957. *Experimental Designs*. 2d ed. New York: John Wiley & Sons.
- Cook, R. D. 1977. "Detection of Influential Observations in Linear Regression." *Technometrics* 19:15–18.
- Cook, R. D. 1979. "Influential Observations in Linear Regression." *Journal of the American Statistical Association* 74:169–174.
- Cook, R. D., and S. Weisberg. 1982. *Residuals and Influence in Regression*. New York: Chapman & Hall.
- Craig, B. A., O. Vitek, M. A. Black, M. Tanurdzic, and R. W. Doerge. 2001. "Designing Microarrays." In *Proceedings of the Thirteenth Annual Conference on Applied Statistics in Agriculture*, 159–182. Manhattan: Kansas State University.
- Cressie, N. A. C. 1993. *Statistics for Spatial Data*. Rev. ed. New York: John Wiley & Sons.
- D'Agostini, G. 1998. "Bayesian Reasoning versus Conventional Statistics in High Energy Physics." Invited talk at the XVIII International Workshop on Maximum Entropy and Bayesian Methods, Garching/München (Germany).
- Daniels, M. J., and R. E. Kass. 2001. "Shrinkage Estimators for Covariance Matrices." *Biometrics* 57:1173–1184.
- Davidian, M., and R. A. Gallant. 1993. "The Nonlinear Mixed Effects Model with a Smooth Random Effects Density." *Biometrika* 80:475–488.
- Davidian, M., and D. Giltinan. 1993. "Some General Estimation Methods for Nonlinear Mixed Effects Models." *Journal of Biopharmaceutical Statistics* 3:23–55.
- Davidian, M., and D. M. Giltinan. 1995. *Nonlinear Models for Repeated Measurement Data*. New York: Chapman & Hall.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B*, 39:1–38.

- Diggle, P. J. 1988. "An Approach to the Analysis of Repeated Measures." *Biometrics* 44: 959–971.
- Draper, N. R., and H. Smith. 1981. *Applied Regression Analysis*. 2d ed. New York: John Wiley & Sons.
- Eilers, P. H. C., and B. D. Marx. 1996. "Flexible Smoothing with B-Splines and Penalties." *Statistical Science* 11(2): 89–121.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. Essex, UK: Longman Group Ltd.
- Feng, S., W. D. Wolfinger, T.-M. Chu, G. C. Gibson, and L. A. McGraw. 2006. "Empirical Bayesian Analysis of Variance Component Models for Microarray Data." *Journal of Agricultural, Biological, and Environmental Statistics*. In press.
- Fitzmaurice, G. M., N. L. Laird, and J. H. Ware. 2004. *Applied Longitudinal Analysis*. New York: John Wiley & Sons.
- Fleming, T. R., J. D. Neaton, A. Goldman, D. L. DeMets., C. Launer, J. Korvick, and D. Abrams. 1995. "Insights from Monitoring the CPCRA Didanosine/Zalcitabine Trial. Terry Beirn Community Programs for Clinical Research on AIDS." *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 10, Suppl. 2, S9–18.
- Fry, J. D. 2004. "Estimation of Genetic Variances and Covariances by Restricted Maximum Likelihood Using PROC MIXED." In *Genetic Analysis of Complex Traits Using SAS*, ed. A. Saxton. Cary, NC: SAS Institute Inc.
- Fuller, W. A., and G. E. Battese. 1974. "Estimation of Linear Models with Crossed-Error Structure." *Journal of Econometrics* 2:67–78.
- Galecki, A. T. 1994. "General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis." *Communications in Statistics: Theory and Methods* 23(11): 3105–3119.
- Galecki, A. T., R. D. Wolfinger, O. A. Linares, M. J. Smith, and J. B. Halter. 2004. "Ordinary Differential Equation PK/PD Models Using the SAS Macro NLINMIX." *Journal of Biopharmaceutical Statistics* 14:483–504.
- Gallant, A. R. 1987. *Nonlinear Statistical Models*. New York: John Wiley & Sons.
- Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith. 1990. "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling." *Journal of the American Statistical Association* 85:972–985.
- Gennings, C., V. M. Chinchilli, and W. H. Carter. 1989. "Response Surface Analysis with Correlated Data: A Nonlinear Model Approach." *Journal of the American Statistical Association* 84:805–809.
- Ghosh, M. 1992. Discussion of M. Schervish, *Bayesian Analysis of Linear Models*. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 432–433. Oxford: Oxford University Press.
- Giesbrecht, F. G., and J. C. Burns. 1985. "Two-Stage Analysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results." *Biometrics* 41: 477–486.
- Gigerenzer, G. 2004. "Mindless Statistics." *Journal of Socio-Economics* 33:587–606.
- Gill, J. 2002. *Bayesian Methods for the Social and Behavioral Sciences*. New York: Chapman & Hall.
- Goldstein, H. 1987. *Multilevel Models in Educational and Social Research*. New York: Oxford University Press.

- Good, I. J. 1988. "Statistical Evidence." In volume 8 of *Encyclopedia of Statistical Sciences*, ed. N. L. Johnson and S. Kotz. New York: John Wiley & Sons.
- Good, I. J. 1992. "The Bayes/Non-Bayes Compromise: A Brief Review." *Journal of the American Statistical Association* 87:597–606.
- Goodman, S. N. 1999a. "Toward Evidence-Based Medical Statistics. 1: The *P* Value Fallacy." *Ann. Intern. Med.* 130:995–1004.
- Goodman, S. N. 1999b. "Toward Evidence-Based Medical Statistics. 2: The Bayes Factor." *Ann. Intern. Med.* 130:1005–1013.
- Goodnight, J. H. 1978. SAS Technical Report R-105: *Computing MIVQUE0 Estimates of Variance Components*. Cary, NC: SAS Institute Inc.
- Goodnight, J. H. 1979. "A Tutorial on the Sweep Operator." *The American Statistician* 33: 149–158.
- Grasela, T. H., Jr., and S. M. Donn. 1985. "Neonatal Population Pharmacokinetics for Phenobarbital Derived from Routine Clinical Data." *Dev. Pharmacol. Ther.* 8:374–383.
- Guerin, L., and W. Stroup. 2000. "A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data." In *Proceedings of the Twelfth Annual Conference on Applied Statistics in Agriculture*. Manhattan: Kansas State University.
- Gumpertz, M. L., and S. G. Pantula. 1992. "Nonlinear Regression with Variance Components." *Journal of the American Statistical Association* 87:201–209.
- Guo, X., and B. P. Carlin. 2003. "Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages." *The American Statistician* 58(1): 16–24.
- Hain, M. 2004. "Influence of Laundering and Drying Parameters on PLA Fabrics." Unpublished M.S. thesis (Yiqi Yang, advisor), University of Nebraska.
- Hartley, A. 2004. "The Philosophy of the Law Idea and the Role of the Prescientific in Statistical Inference." *Journal of the ACMS* 1, 1. <http://www.acmsonline.org/Hartley-04.pdf>.
- Hartley, H. O., and J. N. K. Rao. 1967. "Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model." *Biometrika* 54:93–108.
- Harvey, A. C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica* 44:461–465.
- Harville, D. A. 1997. *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.
- Harville, D. A. 1976. "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects." *Annals of Statistics* 2:384–395.
- Harville, D. A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72:320–338.
- Harville, D. A. 1988. "Mixed-Model Methodology: Theoretical Justifications and Future Directions." In *Proceedings of the Statistical Computing Section*, 41–49. New Orleans: American Statistical Association.
- Harville, D. A. 1990. "BLUP (Best Linear Unbiased Prediction), and Beyond." In *Advances in Statistical Methods for Genetic Improvement of Livestock*, 239–276. New York: Springer-Verlag.
- Harville, D. A., and D. R. Jeske. 1992. "Mean Squared Error of Estimation or Prediction under a General Linear Model." *Journal of the American Statistical Association* 87:724–731.
- Helms, R. W. 1992. "Intentionally Incomplete Longitudinal Designs: I. Methodology and Comparison of Some Full Span Designs." *Statistics in Medicine* 11(14–15): 1889–1913.

- Henderson, C. R. 1950. "The Estimation of Genetic Parameters." *The Annals of Mathematical Statistics* 21:309–310.
- Henderson, C. R. 1963. "Selection Index and Expected Genetic Advance." In *Statistical Genetics and Plant Breeding*, ed. W. D. Hanson and H. F. Robinson, 141–163. Washington, DC: National Academy of Sciences and National Research Council, Publication No. 982.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph.
- Henderson, C. R. 1990. "Statistical Method in Animal Improvement: Historical Overview." In *Advances in Statistical Methods for Genetic Improvement of Livestock*, 1–14. New York: Springer-Verlag.
- Hirst, K., G. O. Zerbe, D. W. Boyle, and R. B. Wilkering. 1991. "On Nonlinear Random Effects Models for Repeated Measurements." *Communications in Statistics: Simulation and Computation* 20:463–478.
- Hively, W. 1996. "The Mathematics of Making Up Your Mind." *Discover* 17(5).
- Hoenig, J. M., and D. M. Heisey. 2001. "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *The American Statistician* 55:19–24.
- Houseman, E. A., L. M. Ryan, and B. A. Coull. 2004. "Cholesky Residuals for Assessing Normal Errors in a Linear Model with Correlated Outcomes." *Journal of the American Statistical Association* 99:383–394.
- Hsu, J. C., and B. Nelson. 1998. "Multiple Comparisons in the General Linear Model." *Journal of Computational and Graphical Statistics* 7:23–41.
- Hsu, J. C. 1996. *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall.
- Hsu, J. C., and M. Peruggia. 1994. "Graphical Representation of Tukey's Multiple Comparison Method." *Journal of Computational and Graphical Statistics* 3:143–161.
- Hurvich, C. M., and C.-L. Tsai. 1989. "Regression and Time Series Model Selection in Small Samples." *Biometrika* 76:297–307.
- Huynh, H., and L. S. Feldt. 1970. "Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions." *Journal of the American Statistical Association* 65:1582–1589.
- Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press.
- Jeffreys, H. 1961. *Theory of Probability*. 3d ed. Oxford: Oxford University Press.
- Jennrich, R. I., and M. D. Schluchter. 1986. "Unbalanced Repeated-Measures Models with Structured Covariance Matrices." *Biometrics* 42:805–820.
- Jin, W., R. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel, and G. Gibson. 2001. "Contributions of Sex, Genotype, and Age to Transcriptional Variance in *Drosophila melanogaster*." *Nature Genetics* 29:389–395. http://statgen.ncsu.edu/ggibson/gibson_website/publications.html.
- Johnson, D. E., U. N. Chaudhuri, and E. T. Kanemasu. 1983. "Statistical Analysis of Line-Source Sprinkler Experiments and Other Nonrandomized Experiments Using Multivariate Methods." *Journal of the American Soil Science Society* 47(2): 309–312.
- Johnson, D. H. 1999. "The Insignificance of Statistical Significance Testing." *Journal of Wildlife Management* 63(3): 763–772.
- Johnson, V. 2004. "Bayes Factors Based on Test Statistics." The University of Michigan Department of Biostatistics Working Paper Series, No. 1029. Berkeley, CA: Berkeley Electronic Press.

- Journal of Educational and Behavioral Statistics.* 1995. Special Issue: Hierarchical Linear Models: Problems and Prospects. Summer (20:2): 109–240.
- Journel, A. G., and C. J. Huijbregts. 1978. *Mining Geostatistics*. London: Academic Press.
- Kackar, R. N., and D. A. Harville. 1984. “Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models.” *Journal of the American Statistical Association* 79:853–862.
- Kass, R. E., and A. Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90:773–795.
- Kass, R. E., and D. Steffey. 1989. “Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models).” *Journal of the American Statistical Association* 84:717–726.
- Kenward, M. G., and J. H. Roger. 1997. “Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood.” *Biometrics* 53:983–997.
- Keselman, H. J., J. Algina, R. K. Kowalchuk, and R. D. Wolfinger. 1998. “A Comparison of Two Approaches for Selecting Covariance Structures in the Analysis of Repeated Measures.” *Communications in Statistics: Simulation and Computation* 27(3): 591–604.
- Keselman, H. J., J. Algina, R. K. Kowalchuk, and R. D. Wolfinger. 1999a. “A Comparison of Recent Approaches to the Analysis of Repeated Measurements.” *British Journal of Mathematical and Statistical Psychology* 52:63–78.
- Keselman, H. J., J. Algina, R. K. Kowalchuk, and R. D. Wolfinger. 1999b. “Analysis of Repeated Measurements: A Comparison of Mixed-Model Satterthwaite F Tests and a Nonpooled Adjusted Degrees of Freedom Multivariate Test.” *Communications in Statistics: Theory and Methods* 28(12): 2967–2999.
- Khuri, A. I., and J. A. Cornell. 1996. *Response Surfaces*. 2d ed. New York: Dekker.
- Khuri, A. I., T. Mathew, and B. K. Sinha. 1998. *Statistical Tests for Mixed Linear Models*. New York: John Wiley & Sons.
- Kianifard, F., and W. H. Swallow. 1996. “A Review of the Development and Application of Recursive Residuals in Linear Models.” *Journal of the American Statistical Association* 91:391–400.
- Kowalchuk, R. K., H. J. Keselman, J. Algina, and R. D. Wolfinger. 2004. “The Analysis of Repeated Measurements with Mixed-Model Adjusted F-tests.” *Educational and Psychological Measurement* 64(2): 224–242.
- Kramer, M. 2005. “R² Statistics for Mixed Models.” In *Proceedings of the Kansas State Conference on Applied Statistics in Agriculture*, 2005. Manhattan: Kansas State University.
- Kreft, I. G. G., J. De Leeuw, and R. Van Der Leeden. 1994. “Review of Five Multilevel Analysis Programs: BMDP-5V, GENMOD, HLM, ML3, VARCL.” *The American Statistician* 48:324–335.
- Krige, D. G. 1951. “A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand.” *Journal of the Chemical, Metallurgical, and Mining Society of South Africa* 52:119–139.
- Lahiri, P., ed. 2001. *Model Selection*. Institute of Mathematical Statistics, Lecture Notes – Monograph Series, Volume 38.
- Laird, N. M., and J. H. Ware. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics* 38:963–974.

- Laird, N. M., N. Lange, and D. Stram. 1987. "Maximum Likelihood Computations with Repeated Measures: Application of the EM Algorithm." *Journal of the American Statistical Association* 82:97–105.
- Lambert, D. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics* 34(1):1–14.
- LaMotte, L. R. 1973. "Quadratic Estimation of Variance Components." *Biometrics* 29:311–330.
- Lange, N., B. P. Carlin, and A. E. Gelfand. 1992. "Hierarchical Bayes Models for the Progression of HIV Infection Using Longitudinal CD4 T-cell Numbers" (with discussion). *Journal of the American Statistical Association* 87:615–632.
- Lavine, M., and M. J. Schervish. 1999. "Bayes Factors: What They Are and What They Are Not." *The American Statistician* 53:119–122.
- Lenth, R. V. 2001. "Some Practical Guidelines for Effective Sample Size Determination." *The American Statistician* 55:187–193. <http://www.stat.uiowa.edu/~rlenth/Power/>.
- Liang, K. Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13–22.
- Lindsey, J. K. 1993. *Models for Repeated Measurements*. Oxford: Clarendon Press.
- Lindstrom, M. J., and D. M. Bates. 1988. "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data." *Journal of the American Statistical Association* 83:1014–1022.
- Lindstrom, M. J., and D. M. Bates. 1990. "Nonlinear Mixed Effects Models for Repeated Measures Data." *Biometrics* 46:673–687.
- Littell, R. C., R. J. Freund, and P. C. Spector. 1991. *SAS System for Linear Models*. 3d ed. Cary, NC: SAS Institute Inc.
- Littell, R. C., J. Pendergast, and R. Natarajan. 2000. "Modelling Covariance Structure in the Analysis of Repeated Measures Data." *Statistics in Medicine* 19:1793–1819.
- Littell, R. C., W. W. Stroup, and R. J. Freund. 2002. *SAS for Linear Models*. 4th ed. Cary, NC: SAS Institute Inc.
- Louis, T. A. 1988. "General Methods for Analysing Repeated Measures." *Statistics in Medicine* 7:29–45.
- Matheron, G. 1963. "Principles of Geostatistics." *Economic Geology* 58:1246–1266.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2d ed. New York: Chapman & Hall.
- McLean, R. A., and W. L. Sanders. 1988. "Approximating Degrees of Freedom for Standard Errors in Mixed Linear Models." In *Proceedings of the Statistical Computing Section*, 50–59. New Orleans: American Statistical Association.
- McLean, R. A., W. L. Sanders, and W. W. Stroup. 1991. "A Unified Approach to Mixed Linear Models." *The American Statistician* 45:54–64.
- Mead, R. 1988. *The Design of Experiments*. New York: Cambridge University Press.
- Mendenhall, W., D. D. Wackerly, and R. L. Scheaffer. 1990. *Mathematical Statistics with Applications*. 4th ed. Belmont, CA: Duxbury Press.
- Messina, F. J., and J. D. Fry. 2003. "Environment-Dependent Reversal of a Life History Trade-off in the Seed Beetle *Callosobruchus maculatus*." *J. Evol. Biol.* 16:501–509.
- Milliken, G. A., and D. E. Johnson. 1989. *Analysis of Messy Data*. Volume 2: *Nonreplicated Experiments*. New York: Chapman & Hall.

- Milliken, G. A., and D. E. Johnson. 1992. *Analysis of Messy Data*. Volume 1: *Designed Experiments*. London: Chapman & Hall.
- Milliken, G. A., and D. E. Johnson. 2002. *Analysis of Messy Data*. Volume 3: *Analysis of Covariance*. London: Chapman & Hall.
- Molenberghs, G., and G. Verbeke. 2005. *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Mood, A. M. 1950. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Morgan, J. E., and M. L. Gumpertz. 1996. "Random Effects Models: Testing Whether Variance Components Are Zero." In *Proceedings of the Joint Statistical Meetings*, 118–126. Alexandria, VA: American Statistical Association.
- Muller, K. E., and B. A. Fetterman. 2002. *Regression and ANOVA: An Integrated Approach Using SAS Software*. Cary, NC: SAS Institute.
- Murray, D. M., and R. D. Wolfinger. 1994. "Analysis Issues in the Evaluation of Community Trials: Progress toward Solutions in SAS/STAT MIXED." *Journal of Community Psychology*, CSAP Special Issue:140–154.
- Murray, D. M., P. J. Hannan, R. D. Wolfinger, W. L. Baker, and J. H. Dwyer. 1998. "Analysis of Data from Group-Randomized Trials with Repeat Observations on the Same Groups." *Statistics in Medicine* 17(14): 1581–1600.
- Myers, R. H., and D. C. Montgomery. 2002. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 2d ed. New York: John Wiley & Sons.
- Nelder, J. A., and R. W. M. Wedderburn. 1972. "Generalised Linear Models." *Journal of the Royal Statistical Society, Series A*, 135:370–384.
- Nelson, P. R. 1982. "Exact Critical Points for the Analysis of Means." *Communications in Statistics* 11:699–709.
- Nelson, P. R. 1991. "Numerical Evaluation of Multivariate Normal Integrals with Correlations $\rho_{ij} = -\alpha_i \alpha_j$." *The Frontiers of Statistical Scientific Theory & Industrial Applications*, 97–114.
- Nelson, P. R. 1993. "Additional Uses for the Analysis of Means and Extended Tables of Critical Values." *Technometrics* 35:61–71.
- Ott, E. R. 1967. "Analysis of Means—A Graphical Procedure." *Industrial Quality Control*, 101–109. Reprinted in *Journal of Quality Technology* 15 (1983): 10–18.
- Pantula, S. G. 1996. Discussion of "Likelihood Ratio Tests for Variance Components." In *Proceedings of the Joint Statistical Meetings*, 127–129. Alexandria, VA: American Statistical Association.
- Patel, H. I. 1991. "Analysis of Incomplete Data from a Clinical Trial with Repeated Measurements." *Biometrika* 78:609–619.
- Patterson, H. D., and R. Thompson. 1971. "Recovery of Inter-block Information When Block Sizes Are Unequal." *Biometrika* 58:545–554.
- Pauler, D. K., J. C. Wakefield, and R. E. Kass. 1999. "Bayes Factors for Variance Component Models." *Journal of the American Statistical Association* 94:1242–1253.
- Perkins, J., and D. Wang. 2004. "A Comparison of Bayesian and Frequentist Statistics as Applied in a Simple Repeated Measures Example." *Journal of Modern Applied Statistical Methods* 3(1): 227–233.

- Phadke, M. S., R. N. Kackar, D. V. Speeney, and M. J. Grieco. 1983. "Off-line Quality Control for Integrated Circuit Fabrication Using Experimental Design." *Bell System Technical Journal* 62:1273–1309.
- Pinheiro, J. C., and D. M. Bates. 1995. "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics* 4:12–35.
- Prasad, N. G. N., and J. N. K. Rao. 1990. "The Estimation of Mean Squared Error of Small-Area Estimators." *Journal of the American Statistical Association* 85:163–171.
- Prentice, R. L., and L. P. Zhao. 1991. "Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses." *Biometrics* 47: 825–839.
- Rao, C. R. 1972. "Estimation of Variance and Covariance Components in Linear Models." *Journal of the American Statistical Association* 67:112–115.
- Remmenga, M., G. A. Milliken, D. D. Kratzer, J. W. Schwenke, and H. R. Rolka. 1997. "Estimating the Maximum Effective Dose in a Quantitative Dose-Response Experiment." *Journal of Animal Sciences* 75:2174–2183.
- Robert, C. P. 1994. *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer-Verlag.
- Robinson, G. K. 1991. "That BLUP Is a Good Thing: The Estimation of Random Effects." *Statistical Science* 6:15–51.
- Roe, D. J. 1997. "Comparison of Population Pharmacokinetic Modeling Methods Using Simulated Data: Results from the Population Modeling Workgroup." *Statistics in Medicine* 16:1241–1262.
- Rosa, G. J. M., and R. J. Tempelman. 2004. "Bayesian Mapping Methodology." In *Genetic Analysis of Complex Traits Using SAS*, ed. A. Saxton. Cary, NC: SAS Institute Inc.
- Rosa, G. J. M., J. P. Steibel, and R. J. Tempelman. 2005. "Reassessing Design and Analysis of Two-Colour Microarray Experiments Using Mixed Effects Models." *Comp. Funct. Genom.* 6:123–131.
- Royall, R. M. 1986. "The Effect of Sample Size on the Meaning of Significance Tests." *The American Statistician* 40(4): 313–315. Also: Bailey, K. R. 1987. "Comment on Royall (1986)." *The American Statistician* 41(3): 245–246.
- Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63:581–592.
- Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.
- Rutter, C. M., and R. M. Elashoff. 1994. "Analysis of Longitudinal Data: Random Coefficient Regression Modelling." *Statistics in Medicine* 13:1211–1231.
- SAS Institute Inc. 1990. *SAS/STAT User's Guide, Version 6, Fourth Edition*. Vol. 2. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 1993. *SAS/ETS User's Guide, Version 6, Second Edition*. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 1996. SAS/STAT Software: Changes and Enhancements through Release 6.11. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 2004. *Getting Started with the SAS Power and Sample Size Application*. Cary, NC: SAS Press.
- Satterthwaite, F. E. 1946. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2:110–114.

- Schabenberger, O. 2004. "Mixed Model Influence Diagnostics." *Proceedings of the Twenty-ninth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc. Available in electronic form at <http://www2.sas.com/proceedings/sugi29/toc.html>.
- Schabenberger, O., and C. A. Gotway. 2005. *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: CRC Press/Chapman & Hall.
- Schabenberger, O., and F. J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. Boca Raton, FL: CRC Press.
- Schervish, M. J. 1992. "Bayesian Analysis of Linear Models." In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 419–434 (with discussion). Oxford: Oxford University Press.
- Schervish, M. J. 1996. "P Values: What They Are and What They Are Not." *The American Statistician* 50:203–206.
- Schield, M. 1997. "Interpreting Statistical Confidence." In *Proceedings of the ASA Joint Statistical Meetings Section on Statistical Education*. <http://web.augsburg.edu/~schield/MiloPapers/97ASA.pdf>.
- Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461–464.
- Searle, S. R. 1971. *Linear Models*. New York: John Wiley & Sons.
- Searle, S. R. 1982. *Matrix Algebra Useful for Statisticians*. New York: John Wiley & Sons.
- Searle, S. R., G. Casella, and C. E. McCulloch. 1992. *Variance Components*. New York: John Wiley & Sons.
- Seber, G. A. F., and C. J. Wild. 2003. *Nonlinear Regression*. New York: John Wiley & Sons.
- Self, S. G., and K. Y. Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions." *Journal of the American Statistical Association* 82:605–610.
- Sellke, T., M. J. Bayarri, and J. O. Berger. 2001. "Calibration of p Values for Testing Precise Null Hypotheses." *The American Statistician* 55:62–71.
- Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.
- Shapiro, A. 1988. "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis." *International Statistical Review* 56(1): 49–62.
- Sheiner, L. B., and S. L. Beal. 1985. "Pharmacokinetic Parameter Estimates from Several Least Squares Procedures: Superiority of Extended Least Squares." *Journal of Pharmacokinetics and Biopharmaceutics* 13:185–201.
- Silvapulle, M. J., and P. Silvapulle. 1995. "A Score Test against One-Sided Alternatives." *Journal of the American Statistical Association* 90:342–349.
- Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*. 8th ed. Ames: Iowa State University Press.
- Sokal, R. R., and F. J. Rohlf. 1981. *Biometry*. New York: W. H. Freeman.
- Steel, R. D. G., and J. H. Torrie. 1980. *Principles and Procedures of Statistics*. 2d ed. New York: McGraw-Hill.
- Stram, D. O., and J. W. Lee. 1994. "Variance Components in the Longitudinal Mixed Effects Model." *Biometrics* 50:1171–1177.
- Stram, D. O., and J. W. Lee. 1995. Correction to "Variance Components in the Longitudinal Mixed Effects Model." *Biometrics* 51:1196.

- Strenio, J. F., H. I. Weisberg, and A. S. Bryk. 1983. "Empirical Bayes Estimation of Individual Growth-Curve Parameters and Their Relationship to Covariates." *Biometrics* 39:71–86.
- Stroup, W. W. 1989. "Predictable Functions and Prediction Space in the Mixed Model Procedure." In *Applications of Mixed Models in Agriculture and Related Disciplines*, 39–48. Southern Cooperative Series Bulletin No. 343. Baton Rouge: Louisiana Agricultural Experiment Station.
- Stroup, W. W. 2002. "Power Analysis Based on Spatial Effects Mixed Models: A Tool for Comparing Design and Analysis Strategies in the Presence of Spatial Variability." *Journal of Agricultural, Biological, and Environmental Statistics* 7:491–511.
- Stroup, W. W., P. S. Baenziger, and D. K. Mulitze. 1994. "Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods." *Crop Science* 34:62–66.
- Stoup, W., and R. C. Littell. 2002. "Impact of Variance Component Estimates on Fixed Effect Inference in Unbalanced Mixed Models." In *Proceedings of the Fourteenth Annual Kansas State University Conference in Applied Statistics in Agriculture*, 32–48.
- Swallow, W. H., and J. F. Monahan. 1984. "Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components." *Technometrics* 28:47–57.
- Swinburne, R., ed. 2002. *Bayes Theorem*. Oxford: Oxford University Press.
- Tempelman, R. J. 2006. "Assessing Statistical Precision, Power, and Robustness of Alternative Experimental Designs for Two Colour Microarray Platforms Based on Mixed Effects Models." *Vet. Immunol. Immunopathol.* In press.
- Theil, H. 1971. *Principles of Econometrics*. New York: John Wiley & Sons.
- Thompson, B. 2001. 402 Citations Questioning the Indiscriminate Use of Null Hypothesis Significance Tests in Observational Studies. <http://biology.uark.edu/Coop/Courses/thompson5.html>.
- Thompson, W. A. 1962. "The Problem of Negative Estimates of Variance Components." *Annals of Mathematical Statistics* 33:273–289.
- Tierney, L. 1994. "Markov Chains for Exploring Posterior Distributions" (with discussion). *Annals of Statistics* 22:1701–1762.
- Verbeke, G., and G. Molenberghs. 1997. *Linear Mixed Models in Practice: A SAS Oriented Approach*. Lecture Notes in Statistics, Vol. 126. New York: Springer.
- Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Verbeke, G., and G. Molenberghs. 2003. "The Use of Score Tests for Inference on Variance Components." *Biometrics* 59:254–262.
- Vonesh, E. F., and R. L. Carter. 1992. "Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures." *Biometrics* 48:1–17.
- Vonesh, E. F., and V. M. Chinchilli. 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.
- Vonesh, E. F. 1992. "Nonlinear Models for the Analysis of Longitudinal Data." *Statistics in Medicine* 11:1929–1954.
- Vonesh, E. F. 1996. "A Note on Laplace's Approximation in Nonlinear Mixed Effects Models." *Biometrika* 83:447–452.
- Wedderburn, R. W. M. 1974. "Quasilikelihood Methods, Generalised Linear Models, and the Gauss-Newton Method." *Biometrika* 61:439–447.
- Westfall, P. H. 1997. "Multiple Testing of General Contrasts Using Logical Constraints and Correlations." *Journal of the American Statistical Association* 92:299–306.

- Westfall, P. H. 2002. "From Farms to Pharmaceuticals: Multiple Comparisons Enter the 21st Century." In *Proceedings of the Thirteenth Annual Kansas State University Conference on Applied Statistics in Agriculture*, ed. G. A. Milliken. Manhattan: Department of Statistics, Kansas State University.
- Westfall, P. H., R. D. Tobias, D. Rom, R. D. Wolfinger, and Y. Hochberg. 1999. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute Inc.
- Whittle, P. 1954. "On Stationary Processes in the Plane." *Biometrika* 41:434–449.
- Winer, B. J. 1971. *Statistical Principles in Experimental Design*. 2d ed. New York: McGraw-Hill.
- Wolfinger, R. D. 1993. "Laplace's Approximation for Nonlinear Mixed Models." *Biometrika* 80:791–795.
- Wolfinger, R., and M. O'Connell. 1993. "Generalized Linear Models: A Pseudo-likelihood Approach." *Journal of Statistical Computation and Simulation* 48:233–243.
- Wolfinger, R. D. 1993. "Covariance Structure Selection in General Mixed Models." *Communications in Statistics: Simulation and Computation* 22:1079–1106.
- Wolfinger, R. D. 1996. "Heterogeneous Variance Covariance Structures for Repeated Measures." *Journal of Agricultural, Biological, and Environmental Statistics* 1(2): 205–230.
- Wolfinger, R. D. 1997. "Comment: Experiences with the SAS Macro NLINMIX." *Statistics in Medicine* 16:1258–1259.
- Wolfinger, R. D. 1998. "Tolerance Intervals for Variance Component Models Using Bayesian Simulation." *Journal of Quality Technology* 30:18–32.
- Wolfinger, R. D., and R. E. Kass. 2000. "Non-conjugate Bayesian Analysis of Variance Component Models." *Biometrics* 56:768–774.
- Wolfinger, R. D., and X. Lin. 1997. "Two Taylor-Series Approximation Methods for Nonlinear Mixed Models." *Computational Statistics and Data Analysis* 25:465–490.
- Wolfinger, R. D., and G. L. Rosner. 1996. "Bayesian and Frequentist Analyses of an *in vivo* Experiment in Tumor Hemodynamics." In *Bayesian Biostatistics*, ed. D. Berry and D. Stangl, 389–409. New York: Marcel Dekker.
- Wolfinger, R. D., and R. D. Tobias. 1998. "Estimation of Location, Dispersion, and Random Effects in Robust Design." *Technometrics* 40(1): 62–71.
- Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules. 2001. "Assessing Gene Significance from cDNA Microarray Data via Mixed Models." *Journal of Computational Biology* 8:625–637.
- Wolfinger, R. D., R. D. Tobias, and J. Sall. 1994. "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models." *SIAM Journal on Scientific Computing* 15(6): 1294–1310.
- Wright, S. P., and R. D. Wolfinger. 1997. "Repeated Measures Analysis Using Mixed Models: Some Simulation Results." In *Modeling Longitudinal and Spatially Correlated Data*, ed. T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. D. Wolfinger. Lecture Notes in Statistics, Vol. 122, 147–157. New York: Springer-Verlag.
- Yuh, L., S. Beal, M. Davidian, F. Harrison, A. Hester, K. Kowalski, E. Vonesh, and R. Wolfinger. 1994. "Population Pharmacokinetic/Pharmacodynamic Methodology and Applications: A Bibliography." *Biometrics* 50:566–575.
- Zeger, S. L., K. Y. Liang, and P. S. Albert. 1988. "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics* 44:1049–1060.
- Zimmerman, D. L., and D. A. Harville. 1991. "A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments." *Biometrics* 47:223–239.

Index

A

ADJUST= option, LSMEANS statement 31, 352, 369
adjusted means
 See least-squares means
adjustment (spatial correlation) 439
AIC (Akaike's information criterion) 170, 184-186, 753
 log-linear variance models 410
 within-subject heterogeneity 381-386
AICC (AIC corrected) 170, 184-185, 753
 log-linear variance models 410
 within-subject heterogeneity 381-386
Akaike's information criterion (AIC)
 See AIC
 See AICC
analysis of covariance 244-316
 data sets for this book 765
 one-way treatment structure, balanced
 incomplete block design 272-281
 one-way treatment structure, equal slopes
 model 251-263
 one-way treatment structure, incomplete
 block design 263-272
 one-way treatment structure, unbalanced
 incomplete block design structure 281-285
 split-plot design, covariate on intermediate-size unit 308-315
 split-plot design, covariate on large-size unit 286-297, 507-509
 split-plot design, covariate on small-size unit 297-308
 strategy for (summary) 250
analysis of repeated measures data 198-202, 429-430
 diagnostics for (example) 426-435
analysis of variance (ANOVA) 2-4
 factorial treatment designs 111
 for randomized block designs 20-22
 split-plot design, correlated whole plots 656-659
 two-way, with unequal variances 345-354
 univariate 161
 with spatial covariance 460-471
ANCOVA
 See analysis of covariance

anisotropic covariance structures 440-442
ANOM (analysis of means) 561
ANOVA (analysis of variance) 2-4
 factorial treatment designs 111
 for randomized block designs 20-22
 split-plot design, correlated whole plots 656-659
 two-way, with unequal variances 345-354
 univariate 161
 with spatial covariance 460-471
ANTE(1) models 176-177
AR(1) models 175-176, 181, 707-715
 AR(1)+RE models 177, 181, 182
 ARH(1) models 177, 379, 386
 ARMA(1,1) models 199
 RANDOM and REPEATED statements 185
 unequally spaced repeated measures 199
 within-subject heterogeneity 379
ARRAY statement, NLMIXED procedure 588
AUTOREG procedure 416
autoregressive errors models
 See AR(1) models
AXIS statement 138

B

balanced confounding, incomplete block design structure with 694-699
balanced incomplete block designs (BIBD) 44-45
baseline covariance models 186-187, 717
Bayes factors 502-509
Bayesian approaches to mixed models 497-524
Bayesian information criterion (BIC) 170, 184-185, 753
 log-linear variance models 410
 within-subject heterogeneity 381-386
best linear unbiased estimate (BLUE) 59, 211, 751
best linear unbiased predictors (BLUPs) 8, 52, 206-242, 552, 751
 basic concepts 210-212
 data sets for this book 764
 heterogeneous variances, testing for 369-371
 multi-factor treatment designs 209
 multilocation examples 226-241
 %NLINMIX macro linearization around 581, 625-626, 628

- best linear unbiased predictors (*continued*)
 random coefficient regression models 208
 random effects models 65, 206, 212-219
 spatial covariance structures and 471-478
 two-way mixed models 207, 219-225
- best quadratic unbiased estimate 59
- between-subjects factors 160
 combining within- and between-subject heterogeneity 393-402
 heterogeneous variance models 344
- BIBDs (balanced incomplete block designs) 44-45
- BIC (Bayesian information criterion) 170, 184-185, 753
 log-linear variance models 410
 within-subject heterogeneity 381-386
- binomial data in multi-center trials (example) 527, 542-557
- binomial probability distributions 532, 533
- block effects 18
- block size 46
- blocking factors 5, 18, 95
- BLUE (best linear unbiased estimate) 59, 211, 751
- BLUP option, ESTIMATE statement 552
- BLUPs (best linear unbiased predictors) 8, 52, 206-242, 552, 751
 basic concepts 210-212
 data sets for this book 764
 heterogeneous variances, testing for 369-371
 multi-factor treatment designs 209
 multilocation examples 226-241
 %NLINMIX macro linearization around 581, 625-626, 628
 random coefficient regression models 208
 random effects models 65, 206, 212-219
 spatial covariance structures and 471-478
 two-way mixed models 207, 219-225
- BOUNDS statement, NLMIXED procedure 571
- box plots of residuals 348
- BOXPLOT option (MIXED procedure) 369
- broad inference (population-wide) 208-209, 211-212, 550
 BLUPs, multi-location models 231-232
 BLUPs, two-factor mixed models 222-223
- Brown-Forsythe's test for equality of variances 345
- C**
- CAIC criterion
 log-linear variance models 410
- within-subject heterogeneity 381-386
- canonical link functions 535
 binomial data analysis (example) 542-552
- case studies 638-731
 crossover and repeated measures 699-716
 data sets for 776
 incomplete block design, 2^3 treatment structure 691-699
 microarray example 727-731
 random coefficients models 716-727
 repeated measures, response surface experiment 643-650
 split-plot design, 2^3 treatment structure 684-691
 split-plot design, correlated whole plots 650-659
 split-plot design, response surface experiment 639-643
 split-plot design, unreplicated 674-683
 split-plot design, whole plot as incomplete Latin square 659-667
 strip-split-split-plot experiment 659-667
- center effects 227
- characterization (spatial correlation) 439
- characterization of mean 12
- characterization of probability distribution 12
- Cholesky residuals 416-417
- CL option
 LSMEANS statement 31
 MIXED procedure 22, 65, 321, 327
 RANDOM statement 576
- "Class Level Information" table 24
- classical estimator (empirical semivariograms) 445
- collinearity diagnostics 414
- common slope models 247-249
 combined estimators 258-263
 fitting 255-258, 284-285
 fitting (split-plot design) 301-305
 testing if adequate 253-254, 264, 274-275, 283
- COMPARE procedure 362, 388
- completely randomized designs (CRDs) 3, 96, 98, 100
 MIXED procedure for 101-102
- complex strip-split-split-plot case study 667-674
- %COMPMIX macro 501
- compound symmetry models 152-153, 171-175, 181-183, 185, 737-738
- experiment with repeated measures (case study) 644-647

- unequally spaced repeated measures 199
 - within-subject heterogeneity 379
- comprehensive simulation studies 489-495
- conditional distribution, linear mixed models 736
- conditional hierarchical linear models 58, 75-81, 84
 - GLM procedure for 80-81
 - MIXED procedure for 76-80
- conditional mean 12
- conditional method of incorporating spatial dependency 442-443
- conditional models 735
- conditional probabilities, *p*-values as 500
- conditional residuals 418-423
- confidence intervals
 - binomial data in multi-center trials (example) 549
 - factorial treatment designs 111-113
 - random effects models 59, 73-74
 - variance component estimation 35-37
- confounding, incomplete block design structure with 694-699
- CONTRAST statement, GLM procedure
 - intra-block analysis of PBIB data 49
 - split-plot analyses 135
- CONTRAST statement, MIXED procedure 28, 30-31
 - covariance models 190-194
 - direct regression models 146
 - factorial treatment designs 104-106
 - intra-block analysis of PBIB data 50
 - regression analysis with orthogonal polynomial contrasts 140-143
 - split-plot experiment (example) with GLM 122-128
 - testing equality of regression models 269-272
 - with REPEATED statements 173-174
 - contrasts, factorial treatment designs 104, 106
 - “Contrasts” table 31, 52
 - control treatments in split-plot analyses 124, 127-128
 - CONVG= option, MIXED statement 305, 453, 455
 - CONVH= option, MIXED statement 455
 - Cook’s *D* statistic 417, 424-425, 429-430
 - correlated data in GLMs 540, 546
 - correlated errors 152, 644-647
 - correlated whole plots, split-plot experiment with 650-656
 - count data
 - handling excess of zeros 589-595
 - split-plot design (example) 527, 557-566
 - “Covariance Parameter Estimates” table 25, 27, 37
 - covariance parameters 736
 - covariance structures 160, 174-198
 - baseline covariance models 186-187, 717
 - inference on treatment and time effects 190-198
 - MIXED analysis 187-190
 - nonlinear mixed model methods 628
 - selecting models with graphical tools 177-183
 - selecting models with information criteria 183-186
 - selecting models with power calculations 486-489
 - covariate by treatment interaction models 249
 - COVRATIO statistic 425, 429-430
 - COVTEST option (MIXED procedure) 22, 25, 65, 321, 327
 - COVTRACE statistic 425
 - CRDs (completely randomized designs) 3, 96, 98, 100
 - MIXED procedure for 101-102
 - crossover and repeated measures 699-716
 - CS models
 - See* compound symmetry models
 - CSH models, within-subject heterogeneity 379
 - CVADJUST sub-option, ADJUST= option 32

D

 - D* statistic 417, 424-425, 429-430
 - data misspecification 541
 - data sets, list of 759-779
 - DDFM= option, MODEL statement 84
 - See also* Satterthwaite approximation
 - balanced incomplete block designs (BIBD) 273
 - BLUPs, random effects models 213
 - location-specific inference 238
 - random coefficient models 330
 - split-plot analyses 121
 - zero-valued variance component estimates 149
 - DDFM=KR option
 - See* KR (Kenward-Roger) correction

- degrees of freedom
 covariance models 188
 factorial treatment designs 111-113
 split-plot experiment (example) with GLM
 120-122
 zero-valued variance component estimates
 149
- Delta method 551
- DERIVS= argument, %NLINMIX macro 617
- designed experiments 1-2
- Details for Quantile Simulation" table 33
- deviance 536
- DFFITS statistic 417, 425
- diagnostics for mixed models 413-435
 data sets for this book 771
 influence diagnostics 424-426
 random effects and generalized least squares
 418-424
 unequally spaced repeated measures
 (example) 426-435
- DIFF= option, LSMEANS statement 31, 50,
 122, 123, 311, 671
 control treatments 124, 127-128
 random coefficient models 335
- "Differences of Least Squares Means" table 30,
 33
- Diffograms 34-35, 561
- "Dimensions" table 24, 52
- direct regression models, split-plot analyses
 (example) 143-147
- direct simulation 489-495
- direction-dependent covariances
See anisotropic covariance structures
- dispersion effects parameter 354, 404-405
- DIVISOR option, ESTIMATE statement 105,
 122-123
- E**
- EBLUE (empirical BLUE) 751
- EBLUPs (empirical BLUPs), defined 218, 751
See also BLUPs (best linear unbiased predictors)
- EFFECT= sub-option, INFLUENCE option 427
- effects models 2-5
- EGLS (estimated generalized least squares)
 estimator 418, 423, 746
- empirical Bayes analysis 499
See also random coefficients (RC) models
- empirical BLUPs
See BLUPs (best linear unbiased predictors)
- empirical semivariograms 444-446, 461
- E[MS] (expected mean squares) for randomized block designs 20-21
- equal slopes regression models 247-249
 combined estimators 258-263
 fitting 255-258, 284-285
 fitting (split-plot design) 301-305
 testing if adequate 253-254, 264, 274-275,
 283
- equality of variances, testing 345-350, 371-372
- error 2
 assuming normal distribution of 3-4, 12, 19,
 26, 42
 correlated errors 152, 644-647
 independence of errors 42, 459-460, 702-
 707, 711
 multi-factor treatment designs 106-113
 whole-plot error, GLMMs 565
- error contrasts matrix 748
- error recovery in linear models 416
- estimability, GLM vs. MIXED procedures
 154-156
- estimability criterion, GLM procedure and 156
- estimable functions 29, 110, 211, 537
- ESTIMATE statement, GLIMMIX procedure
 547
 obtaining predicted values 552
 split-plot analyses 129-130
- ESTIMATE statement, GLM procedure 41
 DIVISOR option 105, 122-123
 factorial treatment designs 104-106
 intra-block analysis of PBIB data 49, 52-53
 split-plot analyses 134, 135
- ESTIMATE statement, MIXED procedure 28,
 30-31, 41
 BLUPs, multi-location models 232-233
 BLUPs, random effects models 213-214
 BLUPs, two-factor mixed models 220, 225
 common slope models, fitting 255-258
 direct regression models 145
 DIVISOR option 105, 122-123
 equal slopes hypothesis, testing 274-275
 factorial treatment designs 104-106
 intra-block analysis of PBIB data 50, 52-53
 LSMESTIMATE statement (GLIMMIX
 procedure) vs. 129
 random effects models 65
 split-plot experiment (example) with GLM
 122-128
 unequal slopes models, fitting 276-281, 292-
 297
 with REPEATED statements 173-174

- “Estimated G Matrix” table 336-337
 estimated generalized least squares (EGLS)
 estimator 418, 423, 746
 “Estimated R Correlation Matrix” table 154
 “Estimated R Matrix” table 154, 168, 378
 “Estimates” table 31
 evidence, *p*-values as 500
 EXP option, ESTIMATE statement 547
 EXPAND= argument, %NLINMIX macro 576,
 581, 634
 expected mean squares for randomized block
 designs 20-21
 experiment design, defined 94
 experimental units, defined 95
 exponential covariance models 440, 453
 exponential family of probability distributions
 533-534
 externally studentized residuals 416
- F**
- F*-tests and Wald statistics 25, 536, 537, 539-
 540
 approximating power of 480-483
 classical *p*-values 501
 matrix generalization of 115-117
 FA(1) and FA1(1) models 379
 factor effects, fixed 4-5, 734
 for location (multi-location models) 227-230
 generalized linear models (GLMs) 529-531
 in random effects models 60
 nonlinear models (NLMs) 568
 SAS procedures for 13
 factor effects, random 4-5, 58, 735, 743
 for location (multi-location models) 231-234
 in random effects models 60
 incorporating into linear models 538
 residual and influence diagnostics 418-424
 factorial effects analysis, Levene’s test with
 349-350
 factorial effects model, fitting 288-289, 298-301
 factorial experiments
 See multi-factor treatment designs
 finite-population controlled AIC 170, 184-186,
 753
 log-linear variance models 410
 within-subject heterogeneity 381-386
 first-order ante-dependence models [ANTE(1)]
 176-177
 first-order autoregressive [AR(1)] models
 175-176, 181, 707-715
 AR(1)+RE models 177, 181, 182
- ARH(1) models 177, 379, 386
 ARMA(1,1) models 199
 RANDOM and REPEATED statements
 185
 unequally spaced repeated measures 199
 within-subject heterogeneity 379
 Fisher scoring 86
 “Fit Statistics” table 25
 fitted residuals 414
 fixed effects 4-5, 734
 for location (multi-location models) 227-230
 generalized linear models (GLMs) 529-531
 in random effects models 60
 nonlinear models (NLMs) 568
 SAS procedures for 13
 fixed-treatment experiments 9-10
- G**
- G*-side heterogeneity 344
G-side spatial dependency 442-443
 “Gaussian” isotropic covariance models 441
 general linear models
 See LMs
 generalized least squares 423
 generalized linear mixed models (GLMMs) 13,
 526-566, 735
 See also GLMs (generalized linear models)
 background on GLMs 529-538
 binomial data in multi-center trials (example)
 527, 542-557
 count data in split-plot design (example)
 527, 557-566
 data sets for this book 774
 SAS procedures for 14
 when needed 527-529
 generalized linear models (GLMs) 4, 526,
 529-538
 See also GLMMs (generalized linear mixed
 models)
 expanding to GLMMs 538-542
 fixed effects 529-531
 probability distributions 531-534
 SAS procedures for 13-14
 standard models vs. 537
 GENMOD procedure 13, 15-16
 geostatistics
 See spatial variability
 GLIMMIX procedure 14-16, 414, 526
 canonical link functions, binomial data
 542-552

- GLIMMIX procedure (*continued*)
 comparing simple effects with
 LSMESTIMATE statement 105, 129,
 194
 conditional vs. marginal residuals 419
 covariance models 189-190
 estimation methods in 541
 incomplete block design, 2^3 treatment
 structure 696-699
 interaction plots 139
 least-squares means comparisons 33-35
 microarray example 729-731
 OUTPUT statement 552-553
 PLOTS= option 34-35, 419
 probit link functions, binomial data 554-557
 SLICEDIFF= option, LSMEANS statement
 192-194, 304, 311, 671
 split-plot design 128-130
 split-plot design, 2^3 treatment structure 686-
 691
 split-plot design, correlated whole-plot units
 652-656
 strip-plot design, covariate on intermediate-
 size unit 308-315
 strip-split-split-plot experiment 670-674
GLM procedure 13-14, 15
See also CONTRAST statement, GLM
 procedure
See also ESTIMATE statement, GLM
 procedure
See also RANDOM statement, GLM
 procedure
 BLUPs, random effects models 215-218
 BLUPs, two-factor mixed models 223-225
 conditional hierarchical linear models 80-81
 PBIBDs 45-49, 52
 random effects models, one-way structure
 71-73
 RCBD data 37-42
 split-plot analyses 94, 130-135
 split-plot design, 2^3 treatment structure 690
 split-plot design, whole plot as incomplete
 Latin square 661-667
 testing for homogeneity of variances
 345-350
GLM procedure vs. MIXED procedure 639,
 667, 690-694
 multi-factor treatment designs 130-135,
 154-156
 RCBD data 37-42
 split-plot design, covariate on small-size unit
 305-308
GLMs (generalized linear models) 4, 526, 529-
 538
See also GLMMs (generalized linear mixed
 models)
 expanding to GLMMs 538-542
 fixed effects 529-531
 probability distributions 531-534
 SAS procedures for 13-14
 standard models vs. 537
GLMMs (generalized linear mixed models) 13,
 526-566, 735
See also GLMs (generalized linear models)
 background on GLMs 529-538
 binomial data in multi-center trials (example)
 527, 542-557
 count data in split-plot design (example)
 527, 557-566
 data sets for this book 774
 SAS procedures for 14
 when needed 527-529
Gompertz function 4
 goodness of fit 414, 536
GPLOT procedure 179
 graphical tools to select covariance models
 177-183
GROUP= option
 RANDOM statement 86, 372
 REPEATED statement 350-353, 396
- H**
- hat matrix 415
 heritability, two-way random effects treatment
 structure for 88-91
 heterogeneous variance models 344-411
 between- and within-subject heterogeneity
 393-402
 data sets for this book 769
 log-linear variance models 401-411
 nested models with unequal variances 366-
 374
 simple linear regression models with unequal
 variances 354-365
 two-way ANOVA with unequal variances
 345-354
 within-subject variability 374-392
 HF models, within-subject heterogeneity 379
 hierarchical linear models
See random coefficients (RC) models
HOLD= option, PARMS statement 262-263

- homogeneity of variances, testing for 345-350, 371-372
- HOVTEST= option, MEANS statement 345
- HQIC criterion
- log-linear variance models 410
 - within-subject heterogeneity 381-386
- HTYPE= option, MODEL statement 144, 339
- hurdle models 589-595
- hypothesis testing
- factorial treatment designs 111-113
 - GLMs 536
- I**
- I-I models 380-381, 386
- ID statement, MIXED procedure 398-399
- identity link functions 537
- ILINK option, ESTIMATE statement 547, 552
- incomplete block designs
- GLM vs. MIXED procedures 154-156
 - one-way treatment structures in 263-272
 - three-way interaction as whole-plot comparison 684-691
- incomplete blocks designs 44-56
- GLM procedure for 45-49
 - MIXED procedure 45, 49-56
- independence of errors 42, 459-460, 702-707, 711
- independent increments (I-I) models 380-381, 386
- indirect influence 423
- inference space 8
- broad (population-wide) 208-209, 211-212
 - intermediate 222, 225
 - location-specific 234-241
 - narrow (subject-specific) 208-209, 211-214, 219-220, 225
- influence diagnostics 414, 424-426
- conditional vs. marginal residuals 418-423
 - EGLS estimation and 423-424
- INFLUENCE option, MODEL statement 426, 427
- information criteria to select covariance models 183-186
- log-linear variance models 410
 - within-subject heterogeneity 381-386
- instability of treatments in multilocation trials 237
- inter-block information 259-261
- incomplete block designs with balanced confounding 694-699
- interaction plots 137-138
- interactions, factorial treatment designs 103, 106
- intermediate inference, two-factor mixed models 222, 225
- intra-block information 259-261, 694-699
- intraclass correlation 152
- inverse link functions 535, 551-552
- isotropic covariance structures 440-441
- ITDETAILS option (NLMIXED procedure) 633
- “Iteration History” table 27
- iterative influence analysis 426
- J**
- JOIN sub-option, PLOT=MEANPLOT option 139
- joint survival and longitudinal models 595-607
- K**
- KR (Kenward-Roger) correction 152, 188
- BLUPs, random effects models 213
 - location-specific inference 236, 238
- kriging (spatial prediction) 446, 471-478
- ordinary kriging 471-476
 - universal kriging 476-478
- kriging variance 475
- L**
- lag distance 461
- Laplace’s approximation 570
- Latin square designs 96, 98
- incomplete, whole plot conducted as (case study) 659-667
 - MIXED procedure for 101-102
- LCOMPONENTS option, MODEL statement 501
- LDATA= option, REPEATED statement 380-381
- least-squares means 28, 49
- nonestimable 154-156
 - RCBDs, spatial covariance models vs. 468-471
- “Least Squares Means” table 29
- Levene’s test for equality of variances 345-346, 349, 371-372
- leverage matrix 417
- LIFEREG procedure 13, 599
- likelihood displacement measure 424
- likelihood distance measure 424
- likelihood ratio tests of spatial covariance 450

- linear combinations, factorial treatment designs 104
- linear isotropic covariance models 441
- linear log isotropic covariance models 441
- linear models, generalized
See GLMs
- linear models, generalized mixed
See GLMMs
- linear models (LMs)
See LMs
- linear models, mixed
See LMMs
- linear predictors 530, 550
- linear regression models 3, 42
 testing equality of 269-272
- linear statistical models 2
- linearly unbiased scaled (LUS) estimates 416
- link functions 530, 531, 534-535, 568
 identity link functions 537
 probit models 554-557
- LINK= option, MODEL statement 554
- LMs (linear models) 11, 13, 14, 529-530
See also generalized linear models (GLMs)
- generalized models vs. 537
- influence measures in 417
- random effects and generalized least squares 418-424
- residuals in 415-417
- SAS procedures for 13-14
- LMMs (linear mixed models) 13, 734-756
 estimating parameters and predicting random effects 742-747
 formulation of 735-741
 inference and test statistics 754-756
 information criteria to select models 753
 likelihood ratio tests to select models 752
 matrix notation 734
 pseudo-likelihood approach 538-540, 546
 SAS procedures for 13-14
 statistical properties 747-752
- LOCAL= option, REPEATED statement 355, 360, 443-444
- dispersion effects specification 354, 405
- spatial covariance models, with nugget effect 453
- location-specific inference 234-241
- locations, multiple
See multi-location models
- log-likelihood functions for probability distributions 533
- log-linear variance models 401-411, 534-535
- Power-of-X dispersion, full model 404-407
- Power-of-X dispersion, reduced model 407-408
- UN repeated measures error structure 408-411
- logistic growth curve models 571-587
- %NLINMIX macro for 575-586
- NLMIXED procedure for 571-575
- LOGISTIC procedure 13
- logistic regression (logit) 534-535, 542-552
- logit (logistic regression) 534-535, 542-552
- LSMEANS statement 28-31
 ADJUST= option 31, 352, 369
 CL option 31
 DIFF= option 31, 50, 122-124, 127-128, 311, 335, 671
 SLICE= option 124-127, 134, 137, 190-194, 671
 SLICEDIFF= option 192-194, 304, 311, 671
 STDERR option 41
- LSMEANS statement, GLIMMIX procedure
 ANOM (analysis of means) 561
 Diffogram, requesting 561
 MEANPLOT= option 189
- LSMEANS statement, GLM procedure 156
 BLUPs, random effects models 215, 217-218
 MIXED procedure vs. 41
 split-plot analyses 131, 133
- LSMEANS statement, MIXED procedure
 BLUPs, random effects models 217-218
 common slope models, fitting 284-285
 conditional hierarchical linear models 78
 control treatments 124
 factorial treatment designs 104-106
 GLM procedure vs. 41
 random coefficient models 335, 337
 unequal slopes models, fitting 280, 297
- LSMESTIMATE statement, GLIMMIX procedure 105
 covariance models 194
 split-plot analyses 129
- LUS estimates of residuals 416
- M**
- main effect mean, factorial treatment designs 103, 122-124
- main effects, factorial treatment designs 103, 105
 split-plot experiment (example) with GLM 122-128

- marginal distribution, linear mixed models 736
- marginal likelihood 569-570
- marginal log-likelihood function 538
- marginal mean, factorial treatment designs 103, 122-124
- marginal method of incorporating spatial dependency 442-444
- marginal models 735
- marginal residuals 418-423
- Matérn isotropic covariance models 441
- matrix generalization of *F*-tests 115-117
- matrix notation
 - LMMs (linear mixed models) 734
 - randomized complete block designs 43-44
 - regression model 42
- maximum likelihood (ML) 7, 746-747, 750
 - dispersion effects 407
- MIXED procedure with %NLINMIX macro 626
- Power-of-the-Mean models 359-363
- Power-of-X models 357-358
- random effects models 61-62
 - random effects models, one-way structure 66
- MAXLAGS option, VARIOGRAM procedure 461
- MDFFITS statistic 425, 429
- mean 12
 - ANOM (analysis of means) 561
 - predicting from inverse link function 535
 - unconditional 210
- mean-mean scatter plots (Diffograms) 34-35, 561
- MEANPLOT= option, LSMEANS statement 189
- means models 2
- MEANS statement
 - HOTTEST= option 345
 - WELCH option 345
- method of moments equations 61
- method of moments estimators 61, 63
 - random effects models, one-way structure 66
 - three-level nested linear models 83-84
- METHOD= option, GLIMMIX procedure 541
- METHOD=FIRO option, NLMIXED procedure 570, 571, 633
- METHOD=ML option (MIXED procedure)
 - See* ML (maximum likelihood)
- METHOD=REML option (MIXED procedure)
 - See* REML (restricted maximum likelihood)
- METHOD=TYPE1 option (MIXED procedure) 60
 - conditional hierarchical linear models 78
 - random effects models, one-way structure 70
 - three-level nested linear models 83
- METHOD=TYPE3 option (MIXED procedure) 22, 63
 - conditional hierarchical linear models 78-80
 - split-plot experiment (example) with GLM 118
 - two-way random effects treatment structure 90-91
 - zero-valued variance component estimates 150-154
- microarray example 716-727, 779
- missing data, GLM vs. MIXED procedures 154-156
- misspecification 541
- MIVQUE(0) estimators 750
 - random effects models 61-62
 - random effects models, one-way structure 66
- mixed model case studies 638-731
 - crossover and repeated measures 699-716
 - data sets for 776
 - incomplete block design, 2^3 treatment structure 691-699
 - microarray example 727-731
 - random coefficients models 716-727
 - repeated measures, response surface experiment 643-650
 - split-plot design, 2^3 treatment structure 684-691
 - split-plot design, correlated whole plots 650-659
 - split-plot design, response surface experiment 639-643
 - split-plot design, unreplicated 674-683
 - split-plot design, whole plot as incomplete Latin square 659-667
 - strip-split-split-plot experiment 659-667
- mixed model diagnostics 413-435
 - data sets for this book 771
 - influence diagnostics 424-426
 - random effects and generalized least squares 418-424
 - unequally spaced repeated measures (example) 426-435
- mixed model equations for random effects models 62
- mixed model issues, list of 7-11
- mixed model theory 42-44

- mixed models, defined 6
- mixed models, Bayesian approaches to 497-524
 - Bayes factors and posterior probability
 - approximations 502-507
 - data sets for this book 773
 - p*-values and alternatives 499-502
 - posterior distribution, generating sample with 509-511
 - mixed models, nonlinear
 - See* nonlinear models (NLMs)
- mixed models, power calculations for 479-495
 - comparing spatial designs 486-489
 - constructing power curves 483-486
 - power analysis 480-483
 - simulation 489-495
- mixed models, SAS procedures for 13-16
- mixed models, typology for 11-13
- mixed models with multiple error terms
 - See* multi-factor treatment designs
- mixed models with spatial variability 442-446
 - See also* spatial variability
- MIXED procedure 13, 15
 - See also* specific statement by name
 - as not strictly Bayesian 499
 - built-in variance functions 354
 - incomplete block design, 2³ treatment structure 696-699
 - INFLUENCE option 426, 427
 - interaction plots 137-138
 - METHOD=TYPE1 option 60, 70, 78, 83
 - METHOD=TYPE3 option 22, 63, 78-80, 90-91, 118, 150-154
 - %NLINMIX macro methods 625-626
 - NOBOUND option 150-154, 753
 - NOPROFILE option 70, 196, 467-468
 - ORD option 502
 - random coefficients models 717-727
 - regression analysis with orthogonal polynomial contrasts 140-143
 - SCORING= option 321
 - split-plot design, 2³ treatment structure 686-691
 - split-plot design, response surface experiment 640-643
 - split-plot design, unreplicated 678-683
 - split-plot design, whole plot as incomplete Latin square 661-667
- MIXED procedure, GLM procedure vs. 639, 667, 690, 690-694
- multi-factor treatment designs 130-135, 154-156
- RCBD data 37-42
- MIXED procedure, METHOD=REML option
 - See* REML (restricted maximum likelihood)
- MIXED procedure, for BLUPs
 - multi-location trials, fixed location effects 227-230
 - multi-location trials, location-specific inference 231-241
 - multi-location trials, random location effects 231-234
 - random effects models 213-218
 - two-factor mixed models 220-225
- MIXED procedure, for compound symmetry
 - models 152-153, 171-175, 181-183, 185, 737-738
- experiment with repeated measures (case study) 644-647
- unequally spaced repeated measures 199
- within-subject heterogeneity 379
- MIXED procedure, for covariance models 174-190
 - baseline covariance models 186-187, 717
 - inference on treatment and time effects 190-198
 - selecting, graphical tools for 177-183
 - selecting, information criteria for 183-186
- MIXED procedure, for factorial experiments 101-102
- MIXED procedure, for factorial treatment designs
 - split-plot example (semiconductor) 117-130
 - split-plot example (type \times dose response) 135-147
- MIXED procedure, for heterogeneous variances 86-87, 350-353, 369-371
 - log-linear variance models 402-411
 - nested models with unequal variances 366-374
 - simple linear regression models 354-365
 - testing for homogeneity 346-348
 - within- and between-subject heterogeneity 393-402
 - within-subject variability 374-392
- MIXED procedure, for *p*-values 500-501
 - generating samples from posterior distributions 509-523
 - split-plot design, covariate on large-size unit 507-509
- MIXED procedure, for PBIBDs 45, 49-56
- MIXED procedure, for power calculations 480-483

- MIXED procedure, for random coefficient models 321, 325, 327-330
 repeated measures analysis 335-339
- MIXED procedure, for random effects models 59
 conditional hierarchical linear models 76-80
 estimating variance components 61-62
 method of moments estimators 63
 one-way treatment structure 64-74
 three-level nested linear models 82-87
 two-way random effects treatment structure 88-91
 unequal variance models 86-87
- MIXED procedure, for RCBD data 22-37
- MIXED procedure, for repeated measures experiments 165-171
- MIXED procedure, for spatial variability 438-442
 ANOVA models with spatial covariance 462-466
 estimating ANOVA models 460-471
 estimating regression models 457-460
 estimating spatial covariance 447-457
 kriging 471-478, 472-478
- MIXED procedure, for split-plot experiments covariate on intermediate-size unit (strip-plots) 308-315
 covariate on large-size unit 290-297
 covariate on small-size unit 301-308
- MIXED procedure, for unequal slopes models 265-267
 intra- and inter-block information, obtaining 259-261
- %MixedTPower macro 483-486
- ML (maximum likelihood) 7, 746-747, 750
 dispersion effects 407
- MIXED procedure with %NLINMIX macro 626
 Power-of-the-Mean models 359-363
 Power-of-X models 357-358
 random effects models 61-62
 random effects models, one-way structure 66
- “Model Information” table 24, 27, 52
- MODEL statement 23
 COVTEST option 22, 25, 65, 321, 327
 HTYPE= option 144, 339
 INFLUENCE option 426, 427
 LCOMPONENTS option 501
 NOINT option 23, 59, 145, 252, 265, 284
 ODDSRATIO option 547
 OUTP= option 346, 356
- RESIDUAL option 427
- SCORING= option 86
- SOLUTION (S) option 28, 63, 65, 144, 146, 284, 328
- MODEL statement, DDFM= option 84
See also Satterthwaite approximation
 balanced incomplete block designs (BIBD) 273
- BLUPs, random effects models 213
 location-specific inference 238
 random coefficient models 330
 split-plot analyses 121
 zero-valued variance component estimates 149
- MODEL statement, GLIMMIX procedure 543
- MODEL statement, GLM procedure
 MIXED procedure vs. 40-41
 split-plot analyses 131
- MODEL statement, MIXED procedure
 factorial experiments 101-102
 GLM procedure vs. 40-41
 random coefficient models 327-328, 335-337
 random effects models 59
- MODEL statement, NL MIXED procedure 572
- MODEL= statement, %NLINMIX macro 576
- MODINIT= argument, %NLINMIX macro 616-617
- multi-factor treatment designs 94-157
See also multi-location models
See also split-plot analyses
 BLUP (best linear unbiased predictor) 209, 226-241
 data sets for this book 761
 determining appropriate model 99-102
 effects of interest 102-106
 GLM vs. MIXED procedures 130-135, 154-156
 possible layouts 95-99
 split-plot experiment (example) with GLM 130-135
 split-plot experiment (example) with MIXED 113-130
 standard errors 106-113
 type \times dose response (example) 135-147
 zero-valued variance component estimates 148-154
- multi-location models 8
See also multi-factor treatment designs
 BLUPs 209, 226-241
 BLUPs, fixed location effects 209, 227-230

- multi-location models (*continued*)
BLUPs, random location effects 209, 231-234
location-specific inference 234-241
multierror models
See multi-factor treatment designs
multilevel linear models
See random coefficients (RC) models
multiple comparisons 31-35
multiple error terms
See multi-factor treatment designs
- N**
- naive method of estimating prediction error variance 215
naive standard errors, covariance models 188
narrow inference (subject-specific) 208-209, 211-212
BLUPs, multi-location models 213-214
BLUPs, two-factor mixed models 219-220, 225
natural parameter 534
negative variance, alternative to 152-154
nested models with unequal variances 366-374
nested nonlinear random effects models 587-589
NLIN procedure 13, 15, 466-468
%NLINMIX macro 14, 15, 390
first-order multivariate Taylor series expansion 575-581, 625-627
fitting methods 625-629
linearization around EBLUPs 581, 625-626, 628
logistic growth curve models 575-586
NLMIXED procedure vs. 623-625
one-compartment pharmacokinetic models 616-622
troubleshooting tips 630-633
variance weighting 581-586, 625-626, 628
NLMIXED procedure 14-16, 569-571
joint survival and longitudinal models 595-607
logistic growth curve models 571-575
nested nonlinear random effects models 587-589
%NLINMIX macro vs. 623-625
one-compartment pharmacokinetic models 610-616
Power-of-the-Mean models 364-365
Power-of-X models 358-359
- troubleshooting tips 630-633
zero-inflation and hurdle models 589-595
NLMs (nonlinear models) 4, 13, 568-634
data sets for this book 775
fitting methods of %NLINMIX macro 625-629
joint survival and longitudinal models 595-607
logistic growth curve models 571-587
nested nonlinear random effects models 587-589
NLMIXED procedure vs. %NLINMIX macro 623-625
one-compartment pharmacokinetic models 607-623
SAS procedures for 13
troubleshooting fitting 629-634
zero-inflation and hurdle models 589-595
NLMIXED procedure vs. %NLINMIX macro 623-625
one-compartment pharmacokinetic models 607-623
SAS procedures for 13
troubleshooting fitting 629-634
zero-inflation and hurdle models 589-595
NLMMs (nonlinear mixed models) 13-14
See also NLMs
NLOPTIONS statement, GLIMMIX procedure 189
no-nugget spherical covariance models 447-449
See also nugget effect, models with
NOBLUP option, ESTIMATE statement 552
NOBOUND option, MIXED procedure 150-154, 753
NOILINK option, ESTIMATE statement 552
NOINT option, MODEL statement 23, 59, 145
analysis of covariance 252, 265, 284
NOITER option
MIXED procedure 196, 467-468
PARMS statement 70
non-iterative influence analysis 426
nonestimable means (GLM procedure) 154-156
nonlinear models (NLMs) 4, 13, 568-634
data sets for this book 775
fitting methods of %NLINMIX macro 625-629
joint survival and longitudinal models 595-607
logistic growth curve models 571-587
nested nonlinear random effects models 587-589
NLMIXED procedure vs. %NLINMIX macro 623-625
one-compartment pharmacokinetic models 607-623
SAS procedures for 13
troubleshooting fitting 629-634
zero-inflation and hurdle models 589-595

nonlinear models, mixed

See NLMMs

nonspatial models, comparing with spatial models 451-453

NOPROFILE option (MIXED procedure) 70, 196, 467-468

normal probability distributions 532

normality assumption 3-4, 12, 19, 26, 42

nugget effect, models with 442-444, 453-457

"Null Model Likelihood Ratio Test" table 170, 379

"Number of Observations" table 24

O

O'Brien's test for equality of variances 345

observational studies 1-2

"Odds Ratio Estimates" table 548

ODDSRATIO option, MODEL statement 547

ODS EXCLUDE statement 192

ODS GRAPHICS ON statement 34

ODS HTML statement 34

omitted variables 541

one-compartment pharmacokinetic models 607-623

one-step updates 426

one-way fixed effects treatment structures 245-286

balanced incomplete block design structure 272-281

equal slopes model 251-263

incomplete block designs 263-272

unbalanced incomplete block design structure 281-285

one-way random effects treatment structure 60, 64-74, 320-326

OPTIONS= argument, %NLINMIX macro 634

ORD option (MIXED procedure) 502

ORDER= option (MIXED procedure) 200

ordinary kriging 471-476

ORPOL option, IML procedure 141

orthogonal polynomial analysis 140-143

OUTP= option, MODEL statement 346, 356

OUTPUT statement, GLIMMIX procedure 552-553

over-parameterized models 2-3

overall influence statistic 424

overdispersion 529

Poisson GLMM with (example) 561-566

P

p-values 499-509

pairwise comparisons

See multiple comparisons

parameter estimates, vectors of 424-425, 548

parameterizing covariance matrices with

NLMIXED procedure 630-633

parameterizing variance matrices with

NLMIXED procedure 630-633

PARMS= argument, %NLINMIX macro 576, 617

PARMS statement 196

HOLD= option 262-263

NOITER option 70

spatial covariance in ANOVA models 463, 467

spatial covariance models, no-nugget 447-449

spatial covariance models, with nugget effect 453-457

partial sill 442

partially balanced incomplete block designs (PBIBDs) 44-56

GLM procedure for 45-49

MIXED procedure for 45, 49-56

PBIBDs (partially balanced incomplete block designs) 44-56

GLM procedure for 45-49

MIXED procedure for 45, 49-56

Pearson residuals 417

penalized quasi-likelihood (PQL) 539, 540

perturbation analyses 414, 415

PL (pseudo-likelihood) 538-540, 546

PLOT= option, LSMEANS statement 139, 189

PLOT statement, GPLOT procedure 179, 382

PLOTS= option, GLIMMIX procedure 34-35, 419

plug-in estimators 751

Poisson probability distributions 532, 533

analysis of Poisson GLMMs 557-560

zero-inflation and hurdle models 589-595

population-wide (broad) estimates 208-209, 211-212

positive spatial correlation 438

posterior probability approximations 502-507

split-plot design, covariate on large-size unit 507-509

power calculations for mixed models 479-495

comparing spatial designs 486-489

constructing power curves 483-486

power analysis 480-483

simulation 489-495

power curves 483-486

- power isotropic covariance models 441
Power-of-the-Mean models 354, 359-365
 Power-of-X models vs. 365
 within-subject heterogeneity 386-392
 Power-of-X models 354, 355-359, 365
 log-linear variance models 404-408
 Power-of-the-Mean models vs. 365
 PQL (penalized quasi-likelihood) 539, 540
 practical range of stationary spatial processes
 441
 precision of estimates, influence on 425-426
 PREDICT statement, NLMIXED procedure 359
 predictable functions 211, 539
 PRIOR statement, MIXED procedure 509-510,
 514, 517, 522-523
 probability distributions 531-534
 probit models 554-557
 PROCOPT= argument, %NLINMIX macro
 576, 626
 prospective power calculations
 See power calculations for mixed models
 PSEARCH option, PRIOR statement 514, 517
 pseudo-data 531, 538, 626
 creating with %NLINMIX macro 625-629
 pseudo-likelihood approach 538-540, 546
 PTRANS option, PRIOR statement 514, 517
 PUT statements, NLMIXED procedure 633
- Q**
- QPOINTS= option (NLMIXED procedure) 610
 quadratic random coefficient models
 See random coefficients (RC) models
 quantum random coefficients (RCQ) models
 380, 387
 quasi-likelihood methods 526
- R**
- R option, REPEATED statement 153, 167, 200
 heterogeneous variances 394, 396
 R-side heterogeneity 344, 374-402
 R-side spatial dependency 442-444
 R-squared concepts for mixed models 502
 RANDOM= argument, %NLINMIX macro 624
 random coefficient models 321-326, 328
 random coefficient models, repeated
 measures analysis 333
 RCQ models 387
 TEST option 80
 unequal random effect variances 372-374
 V option 200
 RANDOM statement, GLIMMIX procedure
 546
 whole-plot error 565
 RANDOM statement, GLM procedure 41, 49
 split-plot analyses 131
 split-plot design, covariate on small-size unit
 306-307
 RANDOM statement, MIXED procedure 23, 41
 covariance structures 185
 factorial experiments 101-102
 GROUP= option 86, 372
 log-linear variance models 403-404
 random coefficient models 321-326, 328
 random coefficient models, repeated
 measures analysis 333
 RCQ models 387
 TEST option 80
 unequal random effect variances 372-374
 V option 200
 RANDOM statement, NLMIXED procedure
 572, 587-589
 random variation
 See error
 randomized block designs 6, 18-56
 ANOVA for 20-22

- expected mean squares for 20-21
- incomplete 44-56
- matrix notation for 43-44
- means and variances 19
- randomized complete block designs (RCBDs)
 - 96, 98, 100
 - data sets for this book 759
 - GLM procedure for 37-42
 - MIXED procedure for 22-42, 101-102
 - spatial adjustment on 468-471
- range of stationary spatial processes 441
- RC models 317-341, 380, 387
 - BLUP (best linear unbiased predictor) 208
 - case study 716-727
 - one-way random effects treatment structure 320-326
 - random student effects (example) 326-330
 - repeated measures experiments 330-341
 - within- and between-subject heterogeneity 400-402
- RCBDs
 - See* randomized complete block designs
- RCORR option, REPEATED statement 153, 167
 - heterogeneous variances 394, 396
- RCQ models 380, 387
- recursive (sequential) residuals 416
- REG procedure 13, 15
- regression analysis 3
 - covariance structures 194-198
 - independence of errors 42, 459-460, 702-707, 711
 - matrix notation 42
 - type \times dose response (split-plot example) 140-147
 - with spatial covariance 457-460
- REML (restricted maximum likelihood) 7, 746-750
 - ANOVA models with spatial covariance 462-466
 - conditional hierarchical linear models 76-80
 - covariance structures 195-196
 - factorial treatment designs 111
 - intra-block analysis of PBIB data 52
 - MIXED procedure with %NLINMIX macro 626
 - Power-of-the-Mean models 359-363
 - Power-of-X models 355-357
 - random effects models 61-62
 - random effects models, one-way structure 64-66
- randomized complete block designs 26-27
- spatial covariance models, no-nugget 447-449
- three-level nested linear models 82
- two-way random effects treatment structure 88-89, 91
- zero-valued variance component estimates 150
- repeated measures 9, 160-203
 - basic concepts 160
 - covariance structure, modeling 174-198
 - crossover measures with 699-716
 - data sets for this book 762
 - indirect influence 423
 - mixed model analysis (example) 163-174
 - random coefficient model for (example) 330-341
 - response surface experiment (case study) 643-650
 - statistical model for 161-163
 - types of analyses 161
 - unequally spaced 198-202, 426-435
 - within-subject variability 376
- REPEATED statement
 - GROUP= option 350-353, 396
 - LDATA= option 380-381
 - LOCAL= option 354, 355, 360, 405, 443-444, 453
 - R option 153, 167, 200, 394, 396
 - RCORR option 153, 167, 394, 396
- REPEATED statement, MIXED procedure 153-154
 - ANOVA models with spatial covariance 463
 - CONTRAST and ESTIMATE statements with 173-174
 - covariance structures 185
 - heterogeneous variances 350-353, 376
 - Power-of-the-Mean models 354, 359-365, 386-392
 - random coefficient models 332-341
 - repeated measures experiments 165-171
 - spatial covariance models, no-nugget 447-449
 - spatial covariance models, with nugget effect 453-457
 - spatial dependency 443-444
 - SP(POW) models 200-202
- REPL (restricted pseudo-likelihood) 539
- REPORT sub-option, ADJUST= option 31-32
- residual log likelihood 748

- residual maximum likelihood
See REML
- RESIDUAL option, MODEL statement 427
- residuals in linear models 415-417
 conditional vs. marginal 418-423
- RESPONSE= argument, %NLINMIX macro 624
- response surface experiment, repeated measures 643-650
- response surface experiment, split-plot design 639-643
- restricted maximum likelihood (REML) 7, 746-750
- ANOVA models with spatial covariance 462-466
- conditional hierarchical linear models 76-80
- covariance structures 195-196
- factorial treatment designs 111
- intra-block analysis of PBIB data 52
- MIXED procedure with %NLINMIX macro 626
- Power-of-the-Mean models 359-363
- Power-of-X models 355-357
- random effects models 61-62
- random effects models, one-way structure 64-66
- randomized complete block designs 26-27
- spatial covariance models, no-nugget 447-449
- three-level nested linear models 82
- two-way random effects treatment structure 88-89, 91
- zero-valued variance component estimates 150
- restricted pseudo-likelihood (REPL) 539
- RETAIN statement, %NLINMIX macro 616-617
- ridge factor 743
- run time, NL MIXED procedure 633
- S**
- sample surveys 1-2
- Satterthwaite approximation 63, 73, 84, 544, 642
- Satterthwaite's formula 112
- scale parameters 540, 546
- scaled deviance 536
- Schwarz's Bayesian information criterion
See BIC
- SCORING= option
 MIXED statement 321
 MODEL statement 86
- SEED= option, PRIOR statement 514
- SEED= sub-option, ADJUST= option 31
- semivariograms 444-446
- ANOVA models with spatial covariance 466-468
 empirical 444-446, 461
- sensitivity 479
- sequential (recursive) residuals 416
- set-delegation diagnostics 424
- shrinkage estimation 218-219, 743
- sill
See spatial variability
- simple effects, factorial treatment designs 102-103
- split-plot example (semiconductor) 124-128
- simple linear regression models
 extensions to more complex models 250-251
 one-way fixed effects treatment structures with 245-251
 testing for lack of fit 267-269
 with unequal variances 354-365
- simulation 489-495
- sire means, BLUPs and 218-219
- SLICE= option, LSMEANS statement 124-127, 134, 671
- covariance models 190-194
- split-plot example (type \times dose response) 137
- SLICEBY= sub-option, PLOT=MEANPLOT option 139
- SLICEDIFF= option, LSMEANS statement 192-194, 304, 311, 671
- SLICEDIFFTYPE= option, LSMEANS statement 729
- slopes-equal-to-zero hypothesis 247-248
 testing 251-253, 282
- “Solution for Fixed Effects” table 29
- SOLUTION (S) option, MODEL statement 28, 144, 146
- common slope models 284
- random coefficient models 328
- random effects models 63, 65
- SOLUTION (S) option, RANDOM statement 576
- random coefficient models 322, 327
- random effects models 63, 65
- “Solutions for Fixed Effects” table 145
- SP(GAU) models 199

- SP(POW) models 199-202, 647-650
- SP(SPH) models 199
- spatial correlation models 440-442
- spatial prediction (kriging) 446, 471-478
 - ordinary kriging 471-476
 - universal kriging 476-478
- spatial variability 438-478
 - comparing spatial and nonspatial models 451-453
- data sets for this book 772
- estimating ANOVA models with 460-471
- estimating regression models with 457-460
- estimating spatial covariance 447-457
- kriging (spatial prediction) 446, 471-478
- mixed models with 442-447
- spatial correlation models 440-442
- spherical covariance models 447-449
 - comparing with other models 450-451
- spherical isotropic covariance models 440
- split-block designs 99, 100
 - MIXED procedure for 101-102
 - strip-split-split-plot experiment 667-674
- split-plot analyses 9, 94, 98, 738-741
 - See also* multi-factor treatment designs
 - conditional vs. marginal residuals 418-423
 - correlated whole plots (case study) 650-656
 - count data in (GLMM example) 527, 557-566
 - covariate measured on intermediate-size units 308-315
 - covariate measured on large-size units 286-297, 507-509
 - covariate measured on small-size units 297-308
 - data sets for this book 761
 - example (type \times dose response) 135-147
 - GLM vs. MIXED procedures 130-135, 154-156
 - log-linear variance models 403-404
 - MIXED procedure for 101-102, 113-130
 - repeated measures experiments vs. 160
 - response surface experiment (case study) 639-643
 - unreplicated (case study) 674-683
 - whole plot as incomplete Latin square 659-667
- split-plot experimental units 98
- split-unit designs 97-98, 100
- stability of treatments in multilocation trials 237
- standard errors
 - covariance models 188
- factorial treatment designs 106-113
- GLIMMIX procedure for 548
 - in complex split-plot designs 673-674
- standardized residuals 415
- starting values with NLMIXED procedure 630
- stationary models 440
- statistical models 2-4, 479
- STDERR option, LSMEANS statement 41
- STMTS= argument, %NLINMIX macro 576, 617, 624
- strip-plot designs, covariate on intermediate-size unit 308-315
- strip-split plots 9, 94, 98, 738-741
 - See also* multi-factor treatment designs
 - conditional vs. marginal residuals 418-423
 - correlated whole plots (case study) 650-656
 - count data in (GLMM example) 527, 557-566
 - covariate measured on intermediate-size units 308-315
 - covariate measured on large-size units 286-297, 507-509
 - covariate measured on small-size units 297-308
 - data sets for this book 761
 - example (type \times dose response) 135-147
 - GLM vs. MIXED procedures 130-135, 154-156
 - log-linear variance models 403-404
 - MIXED procedure for 101-102, 113-130
 - repeated measures experiments vs. 160
 - response surface experiment (case study) 639-643
 - unreplicated (case study) 674-683
 - whole plot as incomplete Latin square 659-667
- studentization (linear models) 415-416, 434-435
- study, defined 2
- SUBJECT= argument, %NLINMIX macro 624
- SUB(JECT)= option, RANDOM statement 24, 572
 - random coefficient models 321, 328
- SUB(JECT)= option, REPEATED statement 153, 166, 200, 443-444, 447-448
- subject-specific (narrow) estimates 208-209, 211-212
 - BLUPs, multi-location models 213-214
 - BLUPs, two-factor mixed models 219-220, 225
- sum-to-zero constraints 3

SWC (Schwarz's Bayesian information criterion)
See BIC
sweep operations 744-746
SYMBOL statements 138, 382

T

T-values 30, 549
TEST option
 GLM procedure 131
 RANDOM statement 41, 80
test statistics with covariance models 188
“Tests of Fixed Effects” table 25
three-level nested linear models 81-87
time main effects 160, 190-198
TOEP models 176, 181, 182, 185
 unequally spaced repeated measures 199
TOEPh models 177
Toeplitz models 176, 181, 182, 185
 unequally spaced repeated measures 199
“Transformation for Covariance Parameters”
 table 517
“Transformed Parameter Search” table 517
treatment-by-time interactions 160
treatment main effects 160, 190-198
 spatial adjustment on 468-471
treatment structure/design, defined 94
true positive rate 479
two-level conditional hierarchical linear models
 See conditional hierarchical linear models
two-way ANOVA with unequal variances
 345-354
two-way hierarchical linear models
 See one-way random effects treatment structure
two-way mixed models 44-56
 BLUP (best linear unbiased predictor) 207,
 219-225
two-way random effects treatment structure
 (example) 88-91
TYPE= option
 RANDOM statement 321, 325, 328, 335
 REPEATED statement 166, 171-172, 380,
 443-444, 447-449
type \times dose response (split-plot example)
 135-147
 regression analysis 140-147
“Type 1 Test of Fixed Effects” 144
“Type 3 Analysis of Variance” table 25, 52
 zero-valued variance component estimates
 149

Type II error
See power calculations for mixed models
Type1 estimation method 60
 conditional hierarchical linear models 78
 random effects models, one-way structure 70
 three-level nested linear models 83
Type2 estimation method 22, 61, 63
 conditional hierarchical linear models 78-80
 split-plot experiment (example) with GLM
 118
 two-way random effects treatment structure
 90-91
 zero-valued variance component estimates
 150-154
Type3 estimation method 22, 63
 conditional hierarchical linear models 78-80
 split-plot experiment (example) with GLM
 118
 two-way random effects treatment structure
 90-91
 zero-valued variance component estimates
 150-154

U

UN covariance structures 169, 171-172, 181,
 185
log-linear variance models 408-411
unequally spaced repeated measures 199
within- and between-subject heterogeneity
 394-399
within-subject heterogeneity 386
 within-subject variability 375-379
unbalanced incomplete block design structure
 281-285
unbalanced two-way mixed models 44-56
 BLUP (best linear unbiased predictor) 207,
 219-225
unbounded estimation 753
unconditional expectation 210
unconditional hierarchical linear models 58
unconditional hierarchical linear models, nested
 81-87
unconditional mean 12
unequal slopes models 246, 249
 fitting 265-267, 272-282
 fitting (split-plot design) 290-297
unequal variances 85-87
 nested models with 366-374
 simple linear regression models with
 354-365
two-way ANOVA with 345-354

unequally spaced repeated measures 198-202, 426-435
 univariate analyses of variance 161
 universal kriging 476-478
 unreplicated split-plot design (case study) 674-683
 unstructured covariance models 169, 171-172, 181, 185
 log-linear variance models 408-411
 unequally spaced repeated measures 199
 within- and between-subject heterogeneity 394-399
 within-subject heterogeneity 386
 within-subject variability 375-379

V

V option, RANDOM statement 200
 VALUE= option, SYMBOL statement 382
 variable omissions 541
 variance, negative 152-154
 variance component estimation 7
See also confidence intervals
 factorial treatment designs 111-113
 zero-valued estimates 148-154
 variance components 107
 variance functions 533
 scale parameters 540, 546
 variance heterogeneity
See heterogeneous variance models
 variance homogeneity, testing for 345-350, 371-372
 variance matrices, parameterizing 630-633
 variance structures 535
 variance weighting with %NLINMIX macro 581-586, 625-626, 628
 variances, unequal 85-87
 nested models with 366-374
 simple linear regression models with 354-365
 two-way ANOVA with 345-354
 VARIOGRAM procedure 461
 ANOVA models with spatial covariance 466-468
 vectors of parameter estimates 424-425, 548

W

Wald statistics
See F-tests
 Wald Z test 754
 WEIGHT= argument, %NLINMIX macro 617
 WELCH option, MEANS statement 345

Welch test of equality 345
 whitening residuals 416
 whole-plot error 565
 whole-plot experimental units 98, 100
 correlated, split-plot experiment with 650-656
 within-subjects factors 160, 175
 combining within- and between-subject heterogeneity 393-402
 heterogeneous variance models 344, 374-402

Z

zero-inflation models 589-595
 ZIP (zero-inflation Poisson) models 589-595

Numerics

2x2 factorial experiments, layouts for 95-99
 2^3 treatment structure (case studies)
 data set for 778
 incomplete block design with balanced confounding 691-699
 split-plot design, three-way interaction as whole-plot comparison 684-691

Books Available from SAS Press

Advanced Log-Linear Models Using SAS®
by Daniel Zelterman

Analysis of Clinical Trials Using SAS®: A Practical Guide
by Alex Dmitrienko, Geert Molenberghs, Walter Offen, and Christy Chuang-Stein

Annotate: Simply the Basics
by Art Carpenter

Applied Multivariate Statistics with SAS® Software, Second Edition
by Ravindra Khattree and Dayanand N. Naik

Applied Statistics and the SAS® Programming Language, Fourth Edition
by Ronald P. Cody and Jeffrey K. Smith

An Array of Challenges — Test Your SAS® Skills
by Robert Virgile

Carpenter's Complete Guide to the SAS® Macro Language, Second Edition
by Art Carpenter

The Cartoon Guide to Statistics
by Larry Gonick and Woollcott Smith

Categorical Data Analysis Using the SAS® System, Second Edition
by Maura E. Stokes, Charles S. Davis, and Gary G. Koch

Cody's Data Cleaning Techniques Using SAS® Software
by Ron Cody

Common Statistical Methods for Clinical Research with SAS® Examples, Second Edition
by Glenn A. Walker

The Complete Guide to SAS® Indexes
by Michael A. Raithel

Debugging SAS® Programs: A Handbook of Tools and Techniques
by Michele M. Burlew

Efficiency: Improving the Performance of Your SAS® Applications
by Robert Virgile

The Essential PROC SQL Handbook for SAS® Users
by Katherine Prairie

Fixed Effects Regression Methods for Longitudinal Data Using SAS®
by Paul D. Allison

Genetic Analysis of Complex Traits Using SAS®
Edited by Arnold M. Saxton

A Handbook of Statistical Analyses Using SAS®, Second Edition
by B.S. Everitt and G. Der

Health Care Data and SAS®
by Marge Scerbo, Craig Dickstein, and Alan Wilson

The How-To Book for SAS/GRAPH® Software
by Thomas Miron

In the Know ... SAS® Tips and Techniques From Around the Globe
by Phil Mason

Instant ODS: Style Templates for the Output Delivery System
by Bernadette Johnson

Integrating Results through Meta-Analytic Review Using SAS® Software
by Morgan C. Wang and Brad J. Bushman

Learning SAS® in the Computer Lab, Second Edition
by Rebecca J. Elliott

The Little SAS® Book: A Primer
by Lora D. Delwiche and Susan J. Slaughter

The Little SAS® Book: A Primer, Second Edition
by Lora D. Delwiche and Susan J. Slaughter
(updated to include Version 7 features)

The Little SAS® Book: A Primer, Third Edition
by Lora D. Delwiche and Susan J. Slaughter
(updated to include SAS 9.1 features)

Logistic Regression Using the SAS® System: Theory and Application
by Paul D. Allison

Longitudinal Data and SAS®: A Programmer's Guide
by Ron Cody

Maps Made Easy Using SAS®
by Mike Zdeb

Models for Discrete Data
by Daniel Zelterman

Multiple Comparisons and Multiple Tests Using SAS®

Text and Workbook Set

(books in this set also sold separately)

by **Peter H. Westfall, Randall D. Tobias,
Dror Rom, Russell D. Wolfinger,
and Yosef Hochberg**

Multiple-Plot Displays: Simplified with Macros
by **Perry Watts**

*Multivariate Data Reduction and Discrimination with
SAS® Software*

by **Ravindra Khattree
and Dayanand N. Naik**

Output Delivery System: The Basics
by **Lauren E. Haworth**

Painless Windows: A Handbook for SAS® Users, Third Edition
by **Jodie Gilmore**

(updated to include Version 8 and SAS 9.1 features)

The Power of PROC FORMAT
by **Jonas V. Bilenas**

PROC TABULATE by Example
by **Lauren E. Haworth**

Professional SAS® Programming Shortcuts
by **Rick Aster**

Quick Results with SAS/GRAFH® Software
by **Arthur L. Carpenter
and Charles E. Shipp**

Quick Results with the Output Delivery System
by **Sunil K. Gupta**

Quick Start to Data Analysis with SAS®
by **Frank C. Dilorio
and Kenneth A. Hardy**

Reading External Data Files Using SAS®: Examples Handbook
by **Michele M. Burlew**

*Regression and ANOVA: An Integrated Approach Using
SAS® Software*
by **Keith E. Muller
and Bethel A. Fetterman**

SAS® Applications Programming: A Gentle Introduction
by **Frank C. Dilorio**

SAS® for Forecasting Time Series, Second Edition
by **John C. Brocklebank
and David A. Dickey**

SAS® for Linear Models, Fourth Edition
by **Ramon C. Littell, Walter W. Stroup,
and Rudolf J. Freund**

*SAS® for Monte Carlo Studies: A Guide for Quantitative
Researchers*
by **Xitao Fan, Ákos Felsővályi, Stephen A. Sivo,
and Sean C. Keenan**

SAS® Functions by Example

by **Ron Cody**

SAS® Guide to Report Writing, Second Edition
by **Michele M. Burlew**

SAS® Macro Programming Made Easy
by **Michele M. Burlew**

SAS® Programming by Example
by **Ron Cody
and Ray Pass**

*SAS® Programming for Researchers and Social Scientists,
Second Edition*
by **Paul E. Spector**

SAS® Programming in the Pharmaceutical Industry
by **Jack Shostak**

*SAS® Survival Analysis Techniques for Medical Research,
Second Edition*
by **Alan B. Cantor**

*SAS® System for Elementary Statistical Analysis,
Second Edition*
by **Sandra D. Schlotzhauer
and Ramon C. Littell**

SAS® System for Mixed Models
by **Ramon C. Littell, George A. Milliken, Walter W. Stroup,
and Russell D. Wolfinger**

SAS® System for Regression, Third Edition
by **Rudolf J. Freund
and Ramon C. Littell**

SAS® System for Statistical Graphics, First Edition
by **Michael Friendly**

The SAS® Workbook and Solutions Set
(books in this set also sold separately)
by **Ron Cody**

*Selecting Statistical Techniques for Social Science Data:
A Guide for SAS® Users*
by **Frank M. Andrews, Laura Klem, Patrick M. O'Malley,
Willard L. Rodgers, Kathleen B. Welch,
and Terrence N. Davidson**

Statistical Quality Control Using the SAS® System
by **Dennis W. King**

*A Step-by-Step Approach to Using the SAS® System
for Factor Analysis and Structural Equation Modeling*
by **Larry Hatcher**

*A Step-by-Step Approach to Using SAS® for Univariate and
Multivariate Statistics, Second Edition*
by **Norm O'Rourke, Larry Hatcher,
and Edward J. Stepanski**

*Step-by-Step Basic Statistics Using SAS®: Student Guide
and Exercises*
(books in this set also sold separately)
by **Larry Hatcher**

*Survival Analysis Using the SAS® System:
A Practical Guide*
by **Paul D. Allison**

*Tuning SAS® Applications in the OS/390 and z/OS
Environments, Second Edition*
by **Michael A. Raithel**

*Univariate and Multivariate General Linear Models:
Theory and Applications Using SAS® Software*
by **Neil H. Timm**
and **Tammy A. Mieczkowski**

Using SAS® in Financial Research
by **Ekkehart Boehmer, John Paul Broussard,**
and Juha-Pekka Kallunki

Using the SAS® Windowing Environment: A Quick Tutorial
by **Larry Hatcher**

Visualizing Categorical Data
by **Michael Friendly**

Web Development with SAS® by Example
by **Frederick Pratter**

Your Guide to Survey Research Using the SAS® System
by **Archer Gravely**

JMP® Books

JMP® for Basic Univariate and Multivariate Statistics: A Step-by-Step Guide
by **Ann Lehman, Norm O'Rourke, Larry Hatcher,**
and Edward J. Stepanski

JMP® Start Statistics, Third Edition
by **John Sall, Ann Lehman,**
and Lee Creighton

Regression Using JMP®
by **Rudolf J. Freund, Ramon C. Littell,**
and Lee Creighton