Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences

Edward E. Gbur, Walter W. Stroup, Kevin S. McCarter, Susan Durham, Linda J. Young, Mary Christman, Mark West, and Matthew Kramer

Book and Multimedia Publishing Committee
April Ulery, Chair
Warren Dick, ASA Editor-in-Chief
E. Charles Brummer, CSSA Editor-in-Chief
Andrew Sharpley, SSSA Editor-in-Chief
Mary Savin, ASA Representative
Mike Casler, CSSA Representative
David Clay, SSSA Representative
Managing Editor: Lisa Al-Amoodi
# CONTENTS

Foreword vii  
Preface ix  
Authors xi  
Conversion Factors for SI and Non-SI Units xiii  

Chapter 1  
**Introduction**  
1  
1.1 Introduction  
1.2 Generalized Linear Mixed Models  
1.3 Historical Development  
1.4 Objectives of this Book  

Chapter 2  
**Background**  
7  
2.1 Introduction  
2.2 Distributions used in Generalized Linear Modeling  
2.3 Descriptions of the Distributions  
2.4 Likelihood Based Approach to Estimation  
2.5 Variations on Maximum Likelihood Estimation  
2.6 Likelihood Based Approach to Hypothesis Testing  
2.7 Computational Issues  
2.8 Fixed, Random, and Mixed Models  
2.9 The Design–Analysis of Variance–Generalized Linear Mixed Model Connection  
2.10 Conditional versus Marginal Models  
2.11 Software  

Chapter 3  
**Generalized Linear Models**  
35  
3.1 Introduction  
3.2 Inference in Generalized Linear Models  
3.3 Diagnostics and Model Fit  
3.4 Generalized Linear Modeling versus Transformations  

Chapter 4  
**Linear Mixed Models**  
59  
4.1 Introduction  
4.2 Estimation and Inference in Linear Mixed Models  
4.3 Conditional and Marginal Models  
4.4 Split Plot Experiments  
4.5 Experiments Involving Repeated Measures  
4.6 Selection of a Covariance Model  
4.7 A Repeated Measures Example  
4.8 Analysis of Covariance  
4.9 Best Linear Unbiased Prediction
### Chapter 5

**Generalized Linear Mixed Models**

- **5.1 Introduction** 109
- **5.2 Estimation and Inference in Generalized Linear Mixed Models** 110
- **5.3 Conditional and Marginal Models** 111
- **5.4 Three Simple Examples** 125
- **5.5 Over-Dispersion in Generalized Linear Mixed Models** 149
- **5.6 Over-Dispersion from an Incorrectly Specified Distribution** 151
- **5.7 Over-Dispersion from an Incorrect Linear Predictor** 160
- **5.8 Experiments Involving Repeated Measures** 167
- **5.9 Inference Issues for Repeated Measures Generalized Linear Mixed Models** 181
- **5.10 Multinomial Data** 184

### Chapter 6

**More Complex Examples**

- **6.1 Introduction** 199
- **6.2 Repeated Measures in Time and Space** 199
- **6.3 Analysis of a Precision Agriculture Experiment** 210

### Chapter 7

**Designing Experiments**

- **7.1 Introduction** 237
- **7.2 Power and Precision** 238
- **7.3 Power and Precision Analyses for Generalized Linear Mixed Models** 239
- **7.4 Methods of Determining Power and Precision** 241
- **7.5 Implementation of the Probability Distribution Method** 243
- **7.6 A Factorial Experiment with Different Design Options** 250
- **7.7 A Multi-location Experiment with a Binomial Response Variable** 255
- **7.8 A Split Plot Revisited with a Count as the Response Variable** 262
- **7.9 Summary and Conclusions** 268

### Chapter 8

**Parting Thoughts and Future Directions**

- **8.1 The Old Standard Statistical Practice** 271
- **8.2 The New Standard** 272
- **8.3 The Challenge to Adapt** 274

### Index

- **Index** 277
Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences is an excellent resource book for students and professionals alike. This book explains the use of generalized linear mixed models which are applicable to students of agricultural and natural resource sciences. The strength of the book is the available examples and statistical analysis system (SAS) code used for analysis. These “real life” examples provide the reader with the examples needed to understand and use generalized linear mixed models for their own analysis of experimental data. This book, published by the American Society of Agronomy, Crop Science Society of America, and the Soil Science Society of America, will be valuable as its practical nature will help scientists in training as well as practicing scientists. The goal of the three Societies is to provide educational material to advance the profession. This book helps meet this goal.

Chuck Rice, 2011 Soil Science Society of America President
Newell Kitchen, 2011 American Society of Agronomy President
Maria Gallo, 2011 Crop Science Society of America President
The authors of this book are participants in the Multi-state Project NCCC-170 “Research Advances in Agricultural Statistics” under the auspices of the North Central Region Agricultural Experiment Station Directors. Project members are statisticians from land grant universities, USDA-ARS, and industry who are interested in agricultural and natural resource applications of statistics. The project has been in existence since 1991. We consider this book as part of the educational outreach activities of our group. Readers interested in NCCC-170 activities can access the project website through a link on the National Information Management and Support System (NIMSS).

Traditional statistical methods have been developed primarily for normally distributed data. Generalized linear mixed models extend normal theory linear mixed models to include a broad class of distributions, including those commonly used for counts, proportions, and skewed distributions. With the advent of software for implementing generalized linear mixed models, we have found researchers increasingly interested in using these models, but it is “easier said than done.” Our goal is to help those who have worked with linear mixed models to begin moving toward generalized linear mixed models. The benefits and challenges are discussed from a practitioner’s viewpoint. Although some readers will feel confident in fitting these models after having worked through the examples, most will probably use this book to become aware of the potential these models promise and then work with a professional statistician for full implementation, at least for their first few applications.

The original purpose of this book was as an educational outreach effort to the agricultural and natural resources research community. This remains as its primary purpose, but in the process of preparing this work, each of us found it to be a wonderful professional development experience. Each of the authors understood some aspects of generalized linear mixed models well, but no one “knew it all.” By pooling our combined understanding and discussing different perspectives, we each have benefitted greatly. As a consequence, those with whom we consult will benefit from this work as well.

We wish to thank our reviewers Bruce Craig, Michael Guttery, and Margaret Nemeth for their careful reviews and many helpful comments. Jeff Velie constructed many of the graphs that were not automatically generated by SAS (SAS Institute, Cary, NC). Thank you, Jeff. We are grateful to all of the scientists who so willingly and graciously shared their research data with us for use as examples.
Edward Gbur is currently Professor and Director of the Agricultural Statistics Laboratory at the University of Arkansas. Previously he was on the faculty in the Statistics Department at Texas A&M University and was a Mathematical Statistician in the Statistical Research Division at the Census Bureau. He received a Ph.D. in Statistics from The Ohio State University. He is a member and Fellow of the American Statistical Association and a member of the International Biometric Society and the Institute of Mathematical Statistics. His current research interests include experimental design, generalized linear mixed models, stochastic modeling, and agricultural applications of statistics.

Walter Stroup is Professor of Statistics at the University of Nebraska, Lincoln. After receiving his Ph.D. in Statistics from the University of Kentucky in 1979, he joined the Biometry faculty at Nebraska's Institute of Agriculture and Natural Resources. He served as teacher, researcher, and consultant until becoming department chair in 2001. In 2003, Biometry was incorporated into a new Department of Statistics at UNL; Walt served as chair from its founding through 2010. He is co-author of SAS for Mixed Models and SAS for Linear Models. He is a member of the International Biometric Society, American Association for the Advancement of Science, and a member and Fellow of the American Statistical Association. His interests include design of experiments and statistical modeling.

Kevin S. McCarter is a faculty member in the Department of Experimental Statistics at Louisiana State University. He earned the Bachelors degree with majors in Mathematics and Computer Information Systems from Washburn University and the Masters and Ph.D. degrees in Statistics from Kansas State University. He has industry experience as an IT professional in banking, accounting, and health care, and as a biostatistician in the pharmaceutical industry. His dissertation research was in the area of survival analysis. His current research interests include predictive modeling, developing and assessing statistical methodology, and applying generalized linear mixed modeling techniques. He has collaborated with researchers from a wide variety of fields, including agriculture, biology, education, medicine, and psychology.

Susan Durham is a statistical consultant at Utah State University, collaborating with faculty and graduate students in the Ecology Center, Biology Department, and College of Natural Resources. She earned a Bachelors degree in Zoology at Oklahoma State University and a Masters degree in Applied Statistics at Utah State University. Her interests cover the broad range of research problems that have been brought to her as a statistical consultant.
Mary Christman is currently the lead statistical consultant with MCC Statistical Consulting LLC, which provides statistical expertise for environmental and ecological problems. She is also courtesy professor at the University of Florida. She was on the faculty at University of Florida, University of Maryland, and American University after receiving her Ph.D. in statistics from George Washington University. She is a member of several organizations, including the American Statistical Association, the International Environmetrics Society, and the American Association for the Advancement of Science. She received the 2004 Distinguished Achievement Award from the Section on Statistics and the Environment of the American Statistical Association. Her current research interests include linear and non-linear modeling in the presence of correlated error terms, sampling and experimental design, and statistical methodology for ecological and environmental research.

Linda J. Young is Professor of Statistics at the University of Florida. She completed her Ph.D. in Statistics at Oklahoma State University and has previously served on the faculties of Oklahoma State University and the University of Nebraska, Lincoln. Linda has served the profession in a variety of capacities, including President of the Eastern North American Region of the International Biometric Society, Treasurer of the International Biometric Society, Vice-President of the American Statistical Association, and Chair of the Committee of Presidents of Statistical Societies. She has co-authored two books and has more than 100 refereed publications. She is a fellow of the American Association for the Advancement of Science, a fellow of the American Statistical Association, and an elected member of the International Statistical Institute. Her research interests include spatial statistics and statistical modeling.

Mark West is a statistician for the USDA-Agricultural Research Service. He received his Ph.D. in Applied Statistics from the University of Alabama in 1989 and has been a statistical consultant in agriculture research ever since beginning his professional career at Auburn University in 1989. His interests include experimental design, statistical computing, computer intensive methods, and generalized linear mixed models.

Matt Kramer is a statistician in the mid-Atlantic area (Beltsville, MD) of the USDA-Agricultural Research Service, where he has worked since 1999. Prior to that, he spent eight years at the Census Bureau in the Statistical Research Division (time series and small area estimation). He received a Masters and Ph.D. from the University of Tennessee. His interests are in basic biological and ecological statistical applications.
In statistics, a generalized linear mixed model (GLMM) is an extension to the generalized linear model (GLM) in which the linear predictor contains random effects in addition to the usual fixed effects.[1][2][3] They also inherit from GLMs the idea of extending linear mixed models to non-normal data. GLMMs provide a broad range of models for the analysis of grouped data, since the differences between groups can be modelled as a random effect. These models are useful in the analysis of many kinds of data, including longitudinal data.[4]. Fitting a model. Fitting GLMMs via maximum likelihood (as