

Chapman & Hall/CRC  
Statistics in the Social and Behavioral Sciences Series

# Multilevel Modeling Using R

W. Holmes Finch  
Jocelyn E. Bolin  
Ken Kelley



CRC Press  
Taylor & Francis Group



## Statistics

A powerful tool for analyzing nested designs in a variety of fields, multilevel/hierarchical modeling allows researchers to account for data collected at multiple levels. **Multilevel Modeling Using R** provides you with a helpful guide to conducting multilevel data modeling using the R software environment.

After reviewing standard linear models, the authors present the basics of multilevel models and explain how to fit these models using R. They then show how to employ multilevel modeling with longitudinal data and demonstrate the valuable graphical options in R. The book also describes models for categorical dependent variables in both single level and multilevel data. The book concludes with Bayesian fitting of multilevel models. For those new to R, the appendix provides an introduction to this system that covers basic R knowledge necessary to run the models in the book.

## Features

- Shows how to properly model data structures to avoid incorrect parameter and standard error estimates
- Explains how multilevel models provide insights into your data that otherwise might not be detected
- Illustrates helpful graphical options in R appropriate for multilevel data
- Presents models for categorical dependent variables in single level and multilevel contexts
- Discusses multilevel modeling within the Bayesian framework
- Offers an introduction to R in the appendix for R novices
- Uses various R packages to conduct the analyses and interpret the results, with the code available online

Through the R code and detailed explanations provided, this book gives you the tools to launch your own investigations in multilevel modeling and gain insight into your research.



**CRC Press**  
Taylor & Francis Group  
an informa business

[www.crcpress.com](http://www.crcpress.com)





# Mutitilysel Modeling Usin8 R

Finch • Bolin • Kelley



Chapman & Hall/CRC  
Statistics in the Social and Behavioral Sciences Series

# Multilevel Modeling Using R

W. Holmes Finch

Ball State University  
Muncie, Indiana, USA

Jocelyn E. Bolin

Ball State University  
Muncie, Indiana, USA

Ken Kelley

University of Notre Dame  
Notre Dame, Indiana, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

First issued in hardback 2017

© 2014 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an informa business

No claim to original U.S. Government works

ISBN-13: 978-1-4665-1585-7 (pbk)  
ISBN-13: 978-1-1384-6933-4 (hbk)

This book contains information obtained from authentic and highly regarded sources. While all reasonable efforts have been made to publish reliable data and information, neither the author[s] nor the publisher can accept any legal responsibility or liability for any errors or omissions that may be made. The publishers wish to make clear that any views or opinions expressed in this book by individual editors, authors or contributors are personal to them and do not necessarily reflect the views/opinions of the publishers. The information or guidance contained in this book is intended for use by medical, scientific or health-care professionals and is provided strictly as a supplement to the medical or other professional's own judgement, their knowledge of the patient's medical history, relevant manufacturer's instructions and the appropriate best practice guidelines. Because of the rapid advances in medical science, any information or advice on dosages, procedures or diagnoses should be independently verified. The reader is strongly urged to consult the relevant national drug formulary and the drug companies' and device or material manufacturers' printed instructions, and their websites, before administering or utilizing any of the drugs, devices or materials mentioned in this book. This book does not indicate whether a particular treatment is appropriate or suitable for a particular individual. Ultimately it is the sole responsibility of the medical professional to make his or her own professional judgements, so as to advise and treat patients appropriately. The authors and publishers have also attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

---

Library of Congress Cataloging-in-Publication Data

---

Finch, W. Holmes (William Holmes)  
Multilevel modeling using R / W. Holmes Finch, Jocelyn E. Bolin, Ken Kelley.  
pages cm -- (Chapman & Hall/CRC statistics in the social and behavioral sciences ; 16)  
Summary: "This book presents the theory and practice of major multilevel modeling techniques in a variety of contexts using R as the software tool. It describes the applications and extensions of multilevel modeling with special emphasis on the use of R to conduct the analyses and interpret the resulting output. The book is designed for researchers, data analysts, and upper-level undergraduate and graduate students taking a course on multilevel modeling or statistical modeling." -- Provided by publisher.  
Includes bibliographical references and index.  
ISBN 978-1-4665-1585-7 (paperback)  
1. Social sciences--Statistical methods. 2. Multivariate analysis. 3. R (Computer program language) I. Title.

HA31.35.F56 2014  
005.5'5--dc23

2014006663

---

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

# Multilevel Modeling Using R

Chapman & Hall/CRC  
**Statistics in the Social and Behavioral Sciences Series**

**Series Editors**

**Jeff Gill**

Washington University, USA

**Steven Heeringa**

University of Michigan, USA

**Wim van der Linden**

CTB/McGraw-Hill, USA

**J. Scott Long**

Indiana University, USA

**Tom Snijders**

Oxford University, UK  
University of Groningen, NL

**Aims and scope**

Large and complex datasets are becoming prevalent in the social and behavioral sciences and statistical methods are crucial for the analysis and interpretation of such data. This series aims to capture new developments in statistical methodology with particular relevance to applications in the social and behavioral sciences. It seeks to promote appropriate use of statistical, econometric and psychometric methods in these applied sciences by publishing a broad range of reference works, textbooks and handbooks.

The scope of the series is wide, including applications of statistical methodology in sociology, psychology, economics, education, marketing research, political science, criminology, public policy, demography, survey methodology and official statistics. The titles included in the series are designed to appeal to applied statisticians, as well as students, researchers and practitioners from the above disciplines. The inclusion of real examples and case studies is therefore essential.

## Published Titles

### **Analyzing Spatial Models of Choice and Judgment with R**

*David A. Armstrong II, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal*

### **Analysis of Multivariate Social Science Data, Second Edition**

*David J. Bartholomew, Fiona Steele, Irini Moustaki, and Jane I. Galbraith*

### **Latent Markov Models for Longitudinal Data**

*Francesco Bartolucci, Alessio Farcomeni, and Fulvia Pennoni*

### **Statistical Test Theory for the Behavioral Sciences**

*Dato N. M. de Gruijter and Leo J. Th. van der Kamp*

### **Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences**

*Brian S. Everitt*

### **Multilevel Modeling Using R**

*W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley*

### **Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition**

*Jeff Gill*

### **Multiple Correspondence Analysis and Related Methods**

*Michael Greenacre and Jorg Blasius*

### **Applied Survey Data Analysis**

*Steven G. Heeringa, Brady T. West, and Patricia A. Berglund*

### **Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists**

*Herbert Hoijtink*

### **Foundations of Factor Analysis, Second Edition**

*Stanley A. Mulaik*

### **Linear Causal Modeling with Structural Equations**

*Stanley A. Mulaik*

### **Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis**

*Leslie Rutkowski, Matthias von Davier, and David Rutkowski*

### **Generalized Linear Models for Categorical and Continuous Limited Dependent Variables**

*Michael Smithson and Edgar C. Merkle*

### **Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys**

*Guo-Liang Tian and Man-Lai Tang*

### **Computerized Multistage Testing: Theory and Applications**

*Duanli Yan, Alina A. von Davier, and Charles Lewis*





---

# *Preface*

---

The goal of this book is to provide you, the reader, with a comprehensive resource for the conduct of multilevel modeling using the R software package. Multilevel modeling, sometimes referred to as hierarchical modeling, is a powerful tool that allows a researcher to account for data collected at multiple levels. For example, an educational researcher may gather test scores and measures of socioeconomic status (SES) for students who attend a number of different schools. The students would be considered level-1 sampling units, and the schools would be referred to as level-2 units.

Ignoring the structure inherent in this type of data collection can, as we discuss in Chapter 2, lead to incorrect parameter and standard error estimates. In addition to modeling the data structure correctly, we will see in the following chapters that the use of multilevel models can also provide insights into the nature of relationships in our data that might otherwise not be detected.

After reviewing standard linear models in Chapter 1, we will turn our attention to the basics of multilevel models in Chapter 2, before learning how to fit these models using the R software package in Chapters 3 and 4. Chapter 5 focuses on the use of multilevel modeling in the case of longitudinal data, and Chapter 6 demonstrates the very useful graphical options available in R, particularly those most appropriate for multilevel data. Chapters 7 and 8 describe models for categorical dependent variables, first for single-level data, and then in the multilevel context. Finally, we conclude in Chapter 9 with Bayesian fitting of multilevel models. The datasets featured in this book are available at the website [www.mlminr.com](http://www.mlminr.com).

We hope that you find this book to be helpful as you work with multilevel data. Our goal is to provide you with a guidebook that will serve as the launching point for your own investigations in multilevel modeling. The R code and discussion of its interpretation contained in this text should provide you with the tools necessary to gain insights into your own research, in whatever field it may be. We appreciate your taking the time to read our work and hope that you find it as enjoyable and informative to read as it was for us to write.





---

## About the Authors

---

**W. Holmes Finch** is a professor in the Department of Educational Psychology at Ball State University where he has been since 2003. He earned a PhD from the University of South Carolina in 2002. Dr. Finch teaches courses in factor analysis, structural equation modeling, categorical data analysis, regression, multivariate statistics, and measurement to graduate students in psychology and education. His research interests are in the areas of multilevel models, latent variable modeling, methods of prediction and classification, and non-parametric multivariate statistics. Holmes is also an Accredited Professional Statistician (PStat®).

**Jocelyn E. Bolin** earned a PhD in educational psychology from Indiana University Bloomington in 2009. Her dissertation consisted of a comparison of statistical classification analyses under situations of training data misclassification. She is an assistant professor in the Department of Educational Psychology at Ball State University, where she has been since 2010. Dr. Bolin teaches courses on introductory and intermediate statistics, multiple regression analysis, and multilevel modeling for graduate students in social science disciplines. Her research interests include statistical methods for classification and clustering and use of multilevel modeling in the social sciences. She is a member of the American Psychological Association, the American Educational Research Association, and the American Statistical Association and is also an Accredited Professional Statistician (PStat®).

**Ken Kelley** is the Viola D. Hank Associate Professor of Management in the Mendoza College of Business at the University of Notre Dame. Dr. Kelley's research involves the development, improvement, and evaluation of quantitative methods, especially as they relate to statistical and measurement issues in applied research. Dr. Kelley's most notable contributions have been on research design, especially with regard to sample size planning. Dr. Kelley is the developer of the MBESS package for the R statistical language and environment. He is also an Accredited Professional Statistician (PStat®) and associate editor of *Psychological Methods*.



---

# Contents

---

Preface.....xi

About the Authors ..... xiii

**1. Linear Models..... 1**

1.1 Simple Linear Regression .....2

1.1.1 Estimating Regression Models with Ordinary Least Squares.....2

1.2 Distributional Assumptions Underlying Regression .....3

1.3 Coefficient of Determination.....4

1.4 Inference for Regression Parameters.....5

1.5 Multiple Regression.....7

1.6 Example of Simple Manual Linear Regression.....9

1.7 Regression in R.....12

1.7.1 Interaction Terms in Regression .....14

1.7.2 Categorical Independent Variables .....15

1.7.3 Checking Regression Assumptions with R .....18

Summary .....21

**2. Introduction to Multilevel Data Structure .....23**

2.1 Nested Data and Cluster Sampling Designs.....23

2.2 Intraclass Correlation .....24

2.3 Pitfalls of Ignoring Multilevel Data Structure .....28

2.4 Multilevel Linear Models.....29

2.4.1 Random Intercept .....29

2.4.2 Random Slopes.....31

2.4.3 Centering.....34

2.5 Basics of Parameter Estimation with MLMs .....35

2.5.1 Maximum Likelihood Estimation.....35

2.5.2 Restricted Maximum Likelihood Estimation.....36

2.6 Assumptions Underlying MLMs.....36

2.7 Overview of Two-Level MLMs .....37

2.8 Overview of Three-Level MLMs .....38

2.9 Overview of Longitudinal Designs and Their Relationship to MLMs .....40

Summary .....40

**3. Fitting Two-Level Models in R.....43**

3.1 Packages and Functions for Multilevel Modeling in R .....43

3.2 The nlme Package.....44

3.2.1 Simple (Intercept Only) Multilevel Models Using nlme.....44

3.2.2 Random Coefficient Models Using nlme.....49



- 3.2.3 Interactions and Cross-Level Interactions Using nlme....52
    - 3.2.4 Centering Predictors.....54
  - 3.3 The lme4 Package.....55
    - 3.3.1 Random Intercept Models Using lme4 .....55
    - 3.3.2 Random Coefficient Models Using lme4 .....59
  - 3.4 Additional Options.....61
    - 3.4.1 Parameter Estimation Method.....61
    - 3.4.2 Estimation Controls.....62
    - 3.4.3 Chi Square Test for Comparing Model Fit .....62
    - 3.4.4 Confidence Intervals for Parameter Estimates .....63
  - Summary.....64
- 4. Models of Three and More Levels .....67
  - 4.1 The nlme Package.....68
    - 4.1.1 Simple Three-Level Models.....68
    - 4.1.2 Simple Models with More Than Three Levels .....74
    - 4.1.3 Random Coefficient Models with Three or More Levels....76
  - 4.2 lme4 for Three and More Levels .....80
  - Summary.....85
- 5. Longitudinal Data Analysis Using Multilevel Models .....87
  - 5.1 Multilevel Longitudinal Framework.....87
  - 5.2 Person Period Data Structure.....88
  - 5.3 Fitting Longitudinal Models Using nlme and lme4 Packages....90
  - 5.4 Changing Covariance Structures of Longitudinal Models .....96
  - 5.5 Benefits of Using Multilevel Modeling for Longitudinal Analysis.....99
  - Summary.....100
- 6. Graphing Data in Multilevel Contexts.....103
  - 6.1 Plots for Linear Models .....107
  - 6.2 Plotting Nested Data .....111
  - 6.3 Using the lattice Package.....112
    - 6.3.1 dotplot.....112
    - 6.3.2 xyplot.....117
  - Summary.....121
- 7. Brief Introduction to Generalized Linear Models.....123
  - 7.1 Logistic Regression Model for Dichotomous Outcome Variable...124
  - 7.2 Logistic Regression Model for Ordinal Outcome Variable.....128
  - 7.3 Multinomial Logistic Regression.....131
  - 7.4 Models for Count Data.....134
    - 7.4.1 Poisson Regression .....134
    - 7.4.2 Models for Overdispersed Count Data .....136
  - Summary.....139

<b>8. Multilevel Generalized Linear Models.....</b>	<b>141</b>
8.1 Multilevel Generalized Linear Model for Dichotomous Outcome Variable.....	141
8.1.1 Random Intercept Logistic Regression.....	142
8.1.2 Random Coefficient Logistic Regression.....	144
8.2 Inclusion of Additional Level 1 and Level 2 Effects to MLRM.....	145
8.3 Fitting Multilevel Dichotomous Logistic Regression Using lme4.....	147
8.4 MGLM for Ordinal Outcome Variable.....	151
8.4.1 Random Intercept Logistic Regression.....	151
8.5 MGLM for Count Data .....	154
8.5.1 Random Intercept Poisson Regression .....	154
8.5.2 Random Coefficient Poisson Regression .....	156
8.5.3 Inclusion of Additional Level 2 Effects in Multilevel Poisson Regression Model.....	157
8.6 Fitting Multilevel Poisson Regression Using lme4.....	162
Summary .....	166
<b>9. Bayesian Multilevel Modeling .....</b>	<b>167</b>
9.1 MCMC Estimation .....	168
9.2 MCMCglmm for Normally Distributed Response Variable .....	170
9.3 Including Level 2 Predictors with MCMCglmm .....	177
9.4 User-Defined Priors .....	183
9.5 MCMCglmm for Dichotomous Dependent Variable .....	186
9.6 MCMCglmm for Count Dependent Variable .....	189
Summary .....	196
<b>Appendix: Introduction to R .....</b>	<b>199</b>
<b>References .....</b>	<b>207</b>
<b>Index .....</b>	<b>209</b>

