Multilevel Analysis
Techniques and Applications

Third Edition

QUANTITATIVE METHODOLOGY SERIES

George A. Marcoulides, Series Editor

This series presents methodological techniques to investigators and students. The goal is to provide an understanding and working knowledge of each method with a minimum of mathematical derivations. Each volume focuses on a specific method (e.g., Factor Analysis, Multilevel Analysis, Structural Equation Modelling).

Proposals are invited from interested authors. Each proposal should consist of: a brief description of the volume's focus and intended market; a table of contents with an outline of each chapter; and a curriculum vitae. Materials may be sent to Dr. George A. Marcoulides, University of California, Riverside, george.marcoulides@ucr.edu.

Published titles

Marcoulides • *Modern Methods for Business Research*

Marcoulides & Moustaki • *Latent Variable and Latent Structure Models*

van der Ark, Croon, & Sijtsma (Eds.) • *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*

Duncan, Duncan, & Strycker • *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications, Second Edition*

Heck & Thomas • *An Introduction to Multilevel Modeling Techniques, Second Edition*

Cardinet, Johnson, & Pini • *Applying Generalizability Theory Using EduG*

Heck, Thomas, & Tabata • *Multilevel and Longitudinal Modeling with IBM SPSS*

Heck, Thomas, & Tabata • *Multilevel Modeling of Categorical Outcomes Using IBM SPSS*

Hox • *Multilevel Analysis: Techniques and Applications, Third Edition*

Multilevel Analysis

Techniques and Applications

Third Edition

Joop J. Hox

Mirjam Moerbeek

Rens van de Schoot

Utrecht University, the Netherlands

Third Edition Published in 2018
by Routledge
711 Third Avenue,
New York, NY 10017
www.psypress.com
www.researchmethodsarena.com

Published in Great Britain
by Routledge
2 Park Square
Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an Informa business

ISBN: 978–1–138-12140-9 (hbk)
ISBN: 978–1–138-12136-2 (pbk)
ISBN: 978–131–556-089-2 (ebk)

www.routledge.com/cw/hox

Contents

Preface

To err is human, to forgive divine;
but to include errors into your design is statistical.
—Leslie Kish

This book is intended as an introduction to multilevel analysis for students and researchers. The term 'multilevel' refers to a hierarchical or nested data structure, usually subjects within organizational groups, but the nesting may also consist of repeated measures within subjects, or respondents within clusters, as in cluster sampling. The expression *multilevel model* is used as a generic term for all models for nested data. *Multilevel analysis* is used to examine relations between variables measured at different levels of the multilevel data structure. This book presents two types of multilevel models in detail: the multilevel regression model and the multilevel structural equation model. Although multilevel analysis is used in many research fields, the examples in this book are mainly from the social and behavioral sciences.

In the past decades, multilevel analysis software has become available that is both powerful and accessible, either as special packages or as part of a general software package. In addition, several handbooks have been published, including the earlier editions of this book. There is a continuing interest in multilevel analysis, as evidenced by the appearance of several reviews and monographs, applications in different fields ranging from psychology and sociology, to education and medicine, a thriving Internet discussion list with more than 1400 subscribers, and a biennial International Multilevel Conference that has been running for more than 20 years. The view of 'multilevel analysis' applying to individuals nested within groups has changed to a view that multilevel models and analysis software offer a very flexible way to model complex data. Thus, multilevel modeling has contributed to the analysis of traditional individuals within groups data, repeated measures and longitudinal data, sociometric modeling, twin studies, meta-analysis and analysis of cluster randomized trials.

This book treats two classes of multilevel models: multilevel regression models, and multilevel structural equation models (MSEM).

Multilevel regression models are essentially a multilevel version of the familiar multiple regression model. As Cohen and Cohen (1983), Pedhazur (1997) and others have shown, the multiple regression model is very versatile. Using dummy coding for categorical variables, it can be used to analyze analysis of variance (ANOVA)-type of models as well as the more usual multiple regression models. Since the multilevel regression model is an extension of the classical multiple regression model, it too can be used in a wide variety of research problems.

Chapter Two of this book contains a basic introduction to the multilevel regression model, also known as the hierarchical linear model, or the random coefficient model. Chapters Three and Four discuss estimation procedures, and a number of important methodological and statistical issues. They also discuss some technical issues that are not specific to multilevel regression analysis, such as centering of predictors and interpreting interactions.

Chapter Five introduces the multilevel regression model for longitudinal data. The model is a straightforward extension of the standard multilevel regression model, but there are some specific complications, such as autocorrelated errors, which are discussed.

Chapter Six treats the generalized linear model for dichotomous data and proportions. When the response (dependent) variable is dichotomous or a proportion, standard regression models should not be used. This chapter discusses the multilevel version of the logistic and the probit regression model.

Chapter Seven extends the generalized linear model introduced in chapter Six to analyze data that are ordered categorical and to data that are counts of events. In the context of counts, it presents models that take an overabundance of zeros into account.

Chapter Eight introduces multilevel modeling of survival or event history data. Survival models are for data where the outcome is the occurrence or nonoccurrence of a certain event, in a certain observation period. If the event has not occurred when the observation period ends, the outcome is said to be censored, since we do not know whether or not the event has taken place after the observation period ended.

Chapter Nine discusses cross-classified models. Some data are multilevel in nature, but do not have a neat hierarchical structure. Examples are longitudinal school research data, where pupils are nested within schools, but may switch to a different school in later measurements, and sociometric choice data. Multilevel models for such cross-classified data can be formulated, and estimated with standard software provided that it can handle restrictions on estimated parameters.

Chapter Ten discusses multilevel regression models for multivariate outcomes. These can also be used to assess the reliability of multilevel measurements.

Chapter Eleven describes a variant of the multilevel regression model that can be used in meta-analysis. It resembles the weighted regression model often recommended for meta-analysis. Using standard multilevel regression procedures, it is a flexible analysis tool, especially when the meta-analysis includes multivariate outcomes.

Chapter Twelve deals with the sample size needed for multilevel modeling, and the problem of estimating the power of an analysis given a specific sample size. An obvious complication in multilevel power analysis is that there are different sample sizes at the distinct levels, which should be taken into account.

Chapter Thirteen discusses the statistical assumptions made and presents some ways to check these. It also discusses more robust estimation methods, such as the profile likelihood method and robust standard errors for establishing confidence intervals, and multilevel bootstrap methods for estimating bias-corrected point-estimates and confidence intervals. This chapter also contains an introduction into Bayesian (MCMC) methods for estimation and inference.

Multilevel structural equation models (MSEM), are a powerful tool for the analysis of multilevel data. Recent versions of structural equation modeling software such as

Lavaan, Lisrel, and Mplus all include at least some multilevel features. The general statistical model for multilevel covariance structure analysis is quite complicated. Chapter Fourteen in this book describes two different approaches to estimation in multilevel confirmatory factor analysis. In addition, it deals with issues of calculating standardized coefficients and goodness-of-fit indices in multilevel structural models. Chapter Fifteen extends this to multilevel path models.

Chapter Sixteen describes structural models for latent curve analysis. This is a SEM approach to analyzing longitudinal data, which is very similar to the multilevel regression models treated in Chapter Five.

This book is intended as an introduction to the world of multilevel analysis. Most of the chapters on multilevel regression analysis should be readable for social and behavioral scientists who have a good general knowledge of analysis of variance and classical multiple regression analysis. Some of these chapters contain material that is more difficult, but these are generally a discussion of specialized problems, which can be skipped at first reading. An example is the chapter on longitudinal models, which contains a prolonged discussion of techniques to model specific structures for the covariances between adjacent time points. This discussion is not needed to understand the essentials of multilevel analysis of longitudinal data, but it may become important when one is actually analyzing such data. The chapters on multilevel structure equation modeling obviously require a strong background in multivariate statistics and some background in structural equation modeling, equivalent to, for example, the material covered in Tabachnick and Fidell's (2013) book on multivariate analysis. On the other hand, in addition to an adequate background in structural equation modeling, the chapters on multilevel structural equation modeling do not require knowledge of advanced mathematical statistics. In all these cases, we have tried to keep the discussion of the more advanced statistical techniques theoretically sound, but non-technical.

In addition to its being an introduction, this book describes many extensions and special applications. As an introduction, it is useable in courses on multilevel modeling in a variety of social and behavioral fields, such as psychology, education, sociology, and business. The various extensions and special applications also make it useful to researchers who work in applied or theoretical research, and to methodologists who have to consult with these researchers. The basic models and examples are discussed in non-technical terms; the emphasis is on understanding the methodological and statistical issues involved in using these models. Some of the extensions and special applications contain discussions that are more technical, either because that is necessary for understanding what the model does, or as a helpful introduction to more advanced treatments in other texts. Thus, in addition to its role as an introduction, the book should be useful as a standard reference for a large variety of applications. The chapters that discuss specialized problems, such as the chapter on cross-classified data, the meta-analysis chapter, and the chapter on advanced issues in estimation and testing, can be skipped entirely if preferred.

New to this edition

One important change compared to the second edition is the introduction of two co-authors. This reflects the expansion of multilevel analysis; the field has become so

broad that it is virtually impossible for a single author to keep up with the new developments, both in statistical theory and in software.

Compared to the second edition, some chapters have changed much, while other chapters have mostly been updated to reflect recent developments in statistical research and software development. One important development is increased use of Bayesian estimation and development of robust Maximum Likelihood estimation. We have chosen not to add a separate chapter on Bayesian estimation; instead, Bayesian estimation is discussed in those places where its use improves estimation. The chapters on multilevel logistic regression and on multilevel ordered regression have been expanded with a better treatment of the linked problems of latent scale and explained variance. In multilevel structural equation modeling (MSEM) the developments have been so fast that the chapters on multilevel confirmatory factor analysis and on multilevel path analysis have been significantly revised, in part by removing discussion of estimation methods that are now clearly outdated. The chapter on sample size and power and the chapter on multilevel survival analysis have been extensively rewritten.

An updated website holds the data sets for all the text examples formatted using the latest versions of SPSS, HLM, MLwiN and Mplus, plus some software introductions with updated screen shots for each of these programs. Most analyses in this book can be carried out by any multilevel regression program, although the majority of the multilevel regression analyses were carried out in HLM and MLwiN. The multilevel SEM analyses all use Mplus. System files and setups using these packages are also available at the website.

Some of the example data are real, while others have been simulated especially for this book. The data sets are quite varied so as to appeal to those in several disciplines, including education, sociology, psychology, family studies, medicine, and nursing, Appendix E describes the various data sets used in this book in detail. Further example data will be added to the website for use in computer labs.

Acknowledgments

We thank Dick Carpenter, Lawrence DeCarlo, Brian Gray, Ellen Hamaker, Don Hedeker, Peter van der Heijden, Herbert Hoijtink, Suzanne Jak, Bernet Sekasanvu Kato, Edith de Leeuw, Cora Maas, George Marcoulides, Cameron McIntosh, Herb Marsh, Allison O'Mara, Ian Plewis, Ken Rowe, Elif Unal, Godfried van den Wittenboer, and Bill Yeaton for their comments on the manuscript of the current book or on earlier editions. Their critical comments still shape this book. We also thank numerous students for the feedback they gave us in our multilevel courses.

We thank our colleagues at the Department of Methodology and Statistics of the Faculty of Social Sciences at Utrecht University for providing us with many discussions and a generally stimulating research environment. Our research has also benefited from the lively discussions by the denizens of the Internet *Multilevel Modeling* and the *Structural Equations Modeling (SEMNET)* discussion lists.

We also express our gratitude to the reviewers that reviewed our proposal for the new edition. They provided valuable feedback on the contents and the structure of the proposed book.

As always, any errors remaining in the book are entirely our own responsibility. We appreciate hearing about them, and will keep a list of errata on the homepage of this book.

Joop Hox
Mirjam Moerbeek
Rens van de Schoot

Utrecht, August 2017