Multilevel Analysis

Techniques and Applications

Second Edition

Joop J. Hox
Utrecht University, The Netherlands

Routledge
Taylor & Francis Group
NEW YORK AND HOVE

Contents

Preface to the Second Edition viii
Series Foreword 000

1. Introduction to Multilevel Analysis 1
   1.1 Aggregation and disaggregation 2
   1.2 Why do we need special multilevel analysis techniques? 4
   1.3 Multilevel theories 7
   1.4 Models described in this book 8

2. The Basic Two-Level Regression Model 11
   2.1 Example 11
   2.2 An extended example 16
   2.3 Inspecting residuals 23
   2.4 Three- and more-level regression models 32
   2.5 A note about notation and software 36

3. Estimation and Hypothesis Testing in Multilevel Regression 40
   3.1 Which estimation method? 40
   3.2 Significance testing and confidence intervals 45
   3.3 Contrasts and constraints 51

4. Some Important Methodological and Statistical Issues 54
   4.1 Analysis strategy 54
   4.2 Centering and standardizing explanatory variables 59
   4.3 Interpreting interactions 63
   4.4 Group mean centering 68
   4.5 How much variance is explained? 69

5. Analyzing Longitudinal Data 79
   5.1 Fixed and varying occasions 80
   5.2 Example with fixed occasions 81
   5.3 Example with varying occasions 93
   5.4 Advantages of multilevel analysis for longitudinal data 98
   5.5 Complex covariance structures 99
   5.6 Statistical issues in longitudinal analysis 104
   5.7 Software issues 111

6. The Multilevel Generalized Linear Model for Dichotomous Data and Proportions 112
6.1 Generalized linear models 112
6.2 Multilevel generalized linear models 117
6.3 Example: Analyzing dichotomous data 121
6.4 Example: Analyzing proportions 123
6.5 The ever changing latent scale: Comparing coefficients and variances 133
6.6 Interpretation and software issues 139

7. The Multilevel Generalized Linear Model for Categorical and Count Data 141
7.1 Ordered categorical data 141
7.2 Count data 151
7.3 The ever changing latent scale, again 157

8. Multilevel Survival Analysis 159
8.1 Survival analysis 159
8.2 Multilevel survival analysis 163
8.3 Multilevel ordinal survival analysis 169

9. Cross-Classified Multilevel Models 171
9.1 Example of cross-classified data: Pupils nested within (primary and secondary schools) 173
9.2 Example of cross-classified data: (Sociometric ratings) in small groups 177
9.3 Statistical and computational issues 185

10. Multivariate Multilevel Regression Models 188
10.1 The multivariate model 189
10.2 Example of multivariate multilevel analysis: Multiple response variables 192
10.3 Example of multivariate multilevel analysis: Measuring group characteristics 197

11. The Multilevel Approach to Meta-Analysis 205
11.1 Meta-analysis and multilevel modeling 205
11.2 The variance-known model 207
11.3 Example and comparison with classical meta-analysis 211
11.4 Correcting for artifacts 217
11.5 Multivariate meta-analysis 221
11.6 Statistical and software issues 228
Appendix 230

12. Sample Sizes and Power Analysis in Multilevel Regression 233
12.1 Sample size and accuracy of estimates 233
12.2 Estimating power in multilevel regression designs 237

13. Advanced Issues in Estimation and Testing 257
13.1 The profile likelihood method 259
13.2 Robust standard errors 260
13.3 Multilevel bootstrapping 264
13.4 Bayesian estimation methods 271

14. Multilevel Factor Models 288
14.1 The within and between approach 290
14.2 Full maximum likelihood estimation 297
14.3 An example of multilevel factor analysis 299
14.4 Standardizing estimates in multilevel structural equation modeling 305
14.5 Goodness of fit in multilevel structural equation modeling 306
14.6 Notation and software 309

15. Multilevel Path Models 312
15.1 Example of a multilevel path analysis 312
15.2 Statistical and software issues in multilevel factor and path models 320
Appendix 323

16. Latent Curve Models 325
16.1 Example of latent curve modeling 328
16.2 A comparison of multilevel regression analysis and latent curve modeling 335

References 337
Appendix A: Data and Stories 352
Appendix B: Aggregating and Disaggregating 360
Appendix C: Recording Categorical Data 363
Appendix D: Constructing Orthogonal Polynomials 366
Author index 369
Subject index 000

CHAPTER CONTENTS

This book treats two classes of multilevel models: multilevel regression models, and multilevel models for covariance structures.

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 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. 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 estimate models that resemble confirmative factor analysis, and to assess the reliability of multilevel measurements. A different approach to multilevel confirmative factor analysis is treated in chapter Thirteen.
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 treats some advanced methods of estimation and assessing significance. It discusses the profile likelihood method, 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 models for covariance structures, or multilevel structural equation models (SEM), are a powerful tool for the analysis of multilevel data. Recent versions of structural equation modeling software such as Eqs, Lisrel, 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 both a simplified statistical model proposed by Muthén (1990, 1994), and more recent developments. It explains how multilevel confirmatory factor models can be estimated with either conventional SEM software or using specialized programs. In addition, it deals with issues of calculating standardized coefficients and goodness-of-fit indices in multilevel structural models. Chapter Fifteen extends this to 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 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 this is 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 (2007) book. Conversely, 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, I have tried to keep the discussion of the more advanced statistical techniques theoretically sound, but non-technical.

Many of the techniques and their specific software implementations discussed in this book are the subject of active statistical and methodological research. In other words: both the statistical techniques and the software tools are evolving rapidly. As a result, increasing numbers of researchers will apply increasingly advanced models to their data. Of course, researchers still need to understand the models and techniques that they use. Therefore, in addition to being an introduction to multilevel analysis, this book aims to let the reader become acquainted with some advanced modeling techniques that might be used, such as bootstrapping and Bayesian estimation methods. At the time of writing, these are specialist tools, and certainly not part of the standard analysis toolkit. But they are developing rapidly, and are likely to become more popular in applied research as well.