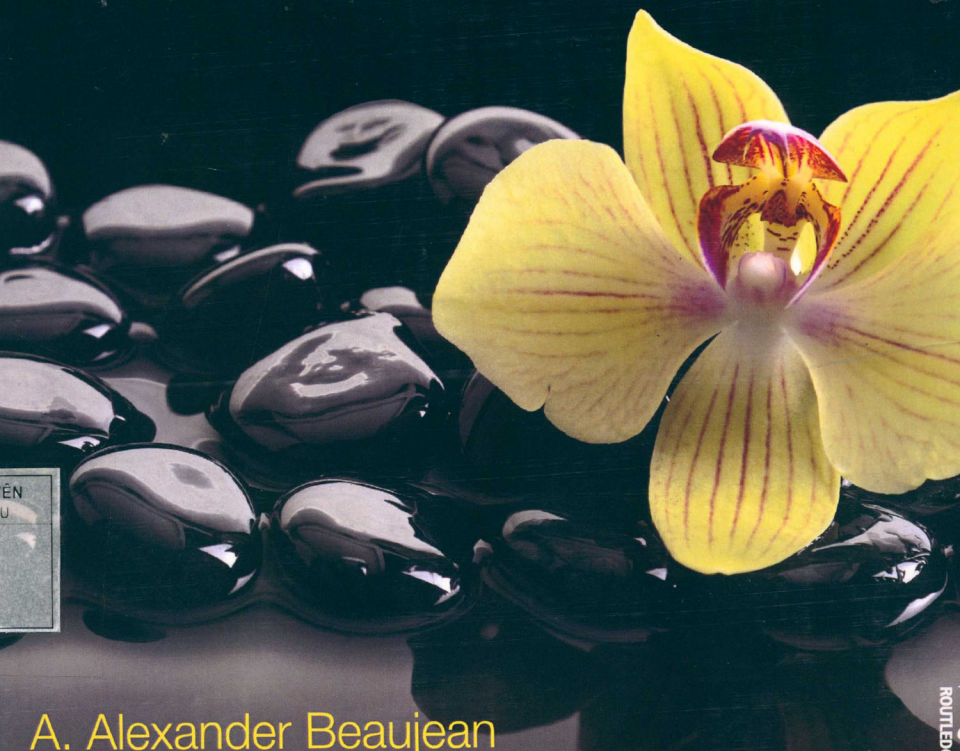
CK.0000074044

Latent Variable Modeling Using R

A Step-by-Step Guide



A. Alexander Beaujean

ROUTLEDGE

Latent Variable Modeling Using R

A Step-by-Step Guide

A. Alexander Beaujean

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2014
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
27 Church Road, Hove, East Sussex BN3 2FA

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

© 2014 Taylor & Francis

The right of A. Alexander Beaujean to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

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
A catalog record has been requested.

ISBN: 978-1-84872-698-7 (hbk)
ISBN: 978-1-84872-699-4 (pbk)
ISBN: 978-1-315-86978-0 (ebk)

Typeset in Latin Modern Roman
by A. Alexander Beaujean

Contents

Author Biography vii

Preface viii

1 Introduction to R 1

1.1 Background . . . 1
1.2 Hints for Using R . . . 18
1.3 Summary . . . 18
1.4 Exercises . . . 18
1.5 References & Further Readings . . . 20

2 Path Models and Analysis 21

2.1 Background . . . 21
2.2 Using R For Path Analysis . . . 27
2.3 Example: Path Analysis using `lavaan` . . . 29
2.4 Indirect Effect . . . 30
2.5 Summary . . . 32
2.6 Writing the Results . . . 32
2.7 Exercises . . . 34
2.8 References & Further Readings . . . 36

3 Basic Latent Variable Models 37

3.1 Background . . . 37
3.2 Latent Variable Models . . . 38
3.3 Example: Latent Variable Model with One Latent Variable . . . 42
3.4 Example: Structural Equation Model . . . 50
3.5 Summary . . . 51
3.6 Writing the Results . . . 51
3.7 Exercises . . . 52
3.8 References & Further Readings . . . 55

4 Latent Variable Models with Multiple Groups 56

4.1 Background . . . 56
4.2 Invariance . . . 56
4.3 Group Equality Constraints . . . 61
4.4 Example: Invariance . . . 62
4.5 Using Labels for Parameter Constraints . . . 70
4.6 Example: Genetically Informative Design . . . 71
4.7 Summary . . . 74
4.8 Writing the Results . . . 75
4.9 Exercises . . . 75
4.10 References & Further Readings . . . 78

5 Models with Multiple Time Periods 79

5.1 Background . . . 79

5.2 Example: Latent Curve Model . . . 80
5.3 Latent Curve Model Extensions . . . 84
5.4 Summary . . . 88
5.5 Writing the Results . . . 88
5.6 Exercises . . . 89
5.7 References & Further Readings . . . 92

6 Models with Dichotomous Indicator Variables 93
6.1 Background . . . 93
6.2 Example: Dichotomous Indicator Variables . . . 104
6.3 Summary . . . 109
6.4 Writing the Results . . . 110
6.5 Exercises . . . 111
6.6 References & Further Readings . . . 112

7 Models with Missing Data 114
7.1 Background . . . 114
7.2 Analyzing Data With Missing Values . . . 117
7.3 Example: Missing Data . . . 121
7.4 Summary . . . 128
7.5 Writing the Results . . . 128
7.6 Exercises . . . 128
7.7 References & Further Readings . . . 130

8 Sample Size Planning 131
8.1 Background . . . 131
8.2 Summary . . . 142
8.3 Writing the Results . . . 142
8.4 Exercises . . . 143
8.5 References & Further Readings . . . 144

9 Hierarchical Latent Variable Models 145
9.1 Background . . . 145
9.2 Summary . . . 151
9.3 Writing the Results . . . 151
9.4 Exercises . . . 151
9.5 References & Further Readings . . . 152

Appendix A Measures of Model Fit 153

Appendix B Additional R Latent Variable Model Packages 167

Appendix C Exercise Answers 171

Glossary 190

Author Index 195

Subject Index 198

R Function Index 202

R Package Index 204

R Dataset Index 205

Author Biography

A. Alexander Beaujean received PhDs in School Psychology and Educational Psychology from the University of Missouri. His research interests are in individual differences, especially their measurement and influence on life outcomes. He is currently an associate professor at Baylor University in the Educational Psychology Department, where he teaches courses on psychological assessment, educational and psychological measurement, and multiple regression. His scholarship has won awards from the American Academy of Health Behavior, American Psychological Association, Mensa, and the Society for Applied Multivariate Research.

Preface

The use of latent variable models has seen a tremendous amount of growth in the past 30 years across a variety of academic disciplines, including the sciences, clinical professions, business, and even the humanities. Part of the reason for this growth is the increasing availability of software to estimate these models' parameters. Traditionally, most of this software has either been too expensive or too complicated for anyone without access to the resources of a large business or university. This trend is rapidly changing, however, and there are now free programs that can conduct a latent variable analysis with only a modicum of knowledge about statistical programming.

This book is designed to introduce **R**, a free statistical program, and show how to use it for latent variable modeling. Thus, the book's two aims are to help readers:

1. understand the basics of the **R** language, and
2. use **R** to analyze a variety of useful latent variable models.

To achieve these aims, this book has some distinctive features that I highlight below.

Path Model Approach to Latent Variable Modeling

Based on teaching graduate students in education, psychology, and related disciplines, I have found that using path models tends to be an effective way to help the novice learn about latent variable models. Consequently, after introducing the **R** program in Chapter 1, I then introduce path models in Chapter 2 and continue to use these models throughout the book. While relying only on path models comes at the price of excluding their matrix representations, it comes with the benefit of increasing the readers' facility of using a model-based approach to translate their research hypotheses into data analysis–an important tool for both students and professionals.

Because of my emphasis on path models throughout the book, I mostly use the **R** package `lavaan` (and packages that work with `lavaan`) to fit the latent variable models. I purposefully did this as `lavaan` uses a path model approach to specify latent variable models. Thus, the chapter text and the **R** syntax complement each other.

Real World Perspective

Having worked with scholars from many disciplines, I know that data are not always well behaved and the syntax to analyze such data are not always easy to find. Consequently, the majority of the examples I use in this text come from published work that represent real data scholars have analyzed. This data comes from a variety of disciplines including education, medicine, psychology, and sociology.

Modern Methods

Because **R** is open-source software, it is continually being updated and improved. Thus, it can use modern techniques to analyze data. While I incorporate this modernity throughout

the book, it is particularly highlighted in the last four chapters as they contain topics that are not readily available from some other latent variable programs. For example, in Chapter 7 I discuss missing data, and demonstrate methods to determine missing data patterns as well as modern methods of handling missing data–including the use of auxiliary variables. Likewise, in Chapter 8 I demonstrate how to use Monte Carlo methods to determine the sample size needed for a prospective study.

Intended Audience

This book can be used as a supplementary text alongside a more theoretical textbook in graduate courses on latent variable modeling. In addition, this book can also be used as a supplementary text in graduate or advanced undergraduate courses that survey latent variable models or courses that review LVMs such as item response theory, measurement, or multivariate statistics taught in a variety of disciplines such as psychology, education, human development, business, economics, and other social and health sciences. Third, professionals and researchers already using latent variable models, but unfamiliar with **R**, will find this book a useful tool for learning some important features of the **R** language.

I used examples from a variety of disciplines to make the context accessible to readers from many different backgrounds, such as business, economics, education, health sciences, human development, psychology, and social science. As the only prerequisite for the text is some familiarity with statistical concepts, both **R** novices and experts should find the text accessible.

Learning Tools

There are some key features in this text to help readers use its material.

Chapter Structure

Every chapter except the first follows the same structure. They all start with some background information, then I work through one or two examples in step-by-step detail, explicitly showing **R** syntax needed for the analyses and interpreting the output. I end each chapter describing how to write the results from that chapter's content for use in a report or publication, as well as providing practice exercises and references/suggested readings. Some of the exercises follow directly from the in-text examples, while others are designed to extend the chapter's content. Most of the exercises require only the use of sample statistics to fit the latent variable model, which I provide in the book. For the exercises that require raw data, I have the files on the book's website at `http://blogs.baylor.edu/rlatentvariable`.

Glossary and Indexes

At the end of the book there are two reader-centered items. The first is a glossary of terms that are likely new and unfamiliar to the latent variable modeling novice. The second are the indices. In addition to the author and subject indices, I also placed three **R** indexes. The first one contains **R** functions, while the second and third contain **R** packages and datasets, respectively. I separated these out purposefully so that the readers do not have to scour the entire index if they forget a **R** function, package, or dataset name.

Text Formatting

This is a hint!

- In the margins I periodically place hints, suggestions, and information that I have found useful. These notes are designed to help readers as they write the **R** syntax for their own models as well as understand some of the complexities involved with latent variable models.

Term

- Every time I introduce a key term, I use **boldface** and place the term in the margin. This should help readers find the areas of interest quickly when they use the book to create their own latent variable models. These terms are then defined in the end of text glossary.

`example.function()`

- Every time I discuss a **R** function or package, I use a `truetype font`. I attach parentheses to the **R** functions [e.g., `example.function()`], and place the name in the margin anytime I introduce a new function or go into substantial detail about it. This will help readers find the these functions quickly when using the book to write their own **R** syntax and analyze their own data.
- I placed all my **R** syntax in a gray box on the page, with resulting output given in the same gray box with two pound symbols ## on the left.

```
R syntax

## Results
```

Book Contents

In Chapter 1, I introduce the **R** program, and discuss how to acquire it, input/import data, and execute some simple functions. The subsequent chapters follow a sequence found in many latent variable textbooks. Chapter 2 introduces path models, while Chapter 3 extends the path models to include latent variables. In Chapter 4 I discuss how to analyze a latent variable model with data from more than one group (including twin data), while in Chapter 5 I discuss how to analyze a latent variable model with data from more than one time period.

The last four chapters are unique for an applied latent variable modeling book. In Chapter 6, I discuss how to handle dichotomous variables, using both the traditional latent variable model perspective as well as an item response theory (IRT) perspective. Further, using a worked example, I show to convert the results from one type of analysis to the other. I devote the entirety of Chapter 7 to fitting a latent variable model with missing data. I discuss types of missing data, methods to determine missing data patterns, and modern methods of handling missing data–including the use of auxiliary variables.

In Chapter 8 I demonstrate how to determine a study's sample size using Monte Carlo simulation. This is not the typical method most textbooks discuss concerning sample size planning, but I chose to focus on this method as it can be used with a wide range of statistical models as well as account for missing data. In the last chapter, Chapter 9, I focus on latent variable models with different levels (i.e., hierarchical models). I include fitting both higher-order models as well as bi-factor models.

After the last chapter, I placed three appendices. Appendix A is about measures of model fit. I do not emphasize the use of any particular model fit index in the book, but in this ap-

pendix I present a variety of common fit indices, including their formulae and interpretation. The second appendix covers a different area. Throughout this book, I mostly use the `lavaan` package. There are other **R** packages that will fit latent variable models, but it has been my experience that it is confusing to learn multiple programs concurrently, as there is a tendency to mix the syntax. Thus, in Appendix B, I provide syntax for other **R** latent variable models packages for readers wishing know how they compare to `lavaan`. Appendix C contains answers (mostly **R** syntax) for each chapter's exercises, although I do suggest trying the exercises yourself before looking at the answers!

While I included as much content as I could, due to space considerations I had to exclude two au courant areas in latent variable modeling. The first area concerns models with a categorical latent variable (i.e., latent class, latent profile). There are **R** packages available for their estimation (e.g., `poLCA`, `mclust`) and the interested reader should read their documentation for more information. The second area is Bayesian estimation. With the integration of winBUGS and JAGS with **R** (e.g., `R2WinBUGS`, `R2jags`), Bayesian estimation of latent variable model is more accessible to **R** users than ever before. Using Bayesian estimation, however, requires much more information about the process of parameter estimation than I provide in this text.

Website

There is a companion website for this book at `http://blogs.baylor.edu/rlatentvariable`. It includes raw data files, **R** syntax for the book examples in a copy-and-paste format, links to related websites with helpful information about **R** and latent variable models, as well as supplemental chapters on creating latent variable model diagrams, LISREL notation, and bootstrapping.

Acknowledgments

I am indebted to many individuals for their help with this book. In particular, I want to thank the individuals who have provided feedback on previous drafts of this text: Danielle Fearon (Baylor University), Darrell Hull (University of North Texas), Grant Morgan (Baylor University), Sonia Parker (Baylor University), Terrill Saxon (Baylor University), Yanyan Sheng (Southern Illinois University-Carbondale), Kara Styck (University of Texas-San Antonio), Phil Wood (University of Missouri), as well as all the students in my latent variable and multiple regression courses.

I also wish to thank the people at Routledge/Taylor & Francis, especially Senior Editor Debra Riegert. While I am responsible for any errors remaining in the text, the book is much better as a result of their input.

I wish to thank Yves Rosseel and Sunthud Pornprasertmanit for answering my questions about their **R** packages, and Mori Jamshidian for the advanced material concerning the `MissMech` package. In addition, thanks to the Law School Admissions Council for allowing me to use some example *Figure Classification* items in the text, and to Craig Enders for allowing the use of his *Eating Attitudes Test* data.