



Journal of Statistical Software

June 2014, Volume 58, Book Review 2.

<http://www.jstatsoft.org/>

Reviewer: Joshua F. Wiley
University of California, Los Angeles

Growth Curve Analysis and Visualization Using R

Daniel Mirman
Chapman & Hall/CRC, Boca Raton, 2014.
ISBN 9781466584327. 188 pp. USD 79.95 (P).
<http://www.danmirman.org/>

The book *Growth Curve Analysis and Visualization Using R* provides an up-to-date, practical introduction to visualizing and modeling time course and multilevel data. It is particularly well-suited to applied researchers in the fields of cognitive science, neuroscience, and linguistics. Virtually no familiarity with R is required (although it helps). The book does assume a solid understanding of multiple regression, but does not assume prior knowledge or experience working with time course data. Detailed code examples are given using **lme4** for linear and logistic growth curve models and **ggplot2** for graphing.

The book can be divided into three sections. Each chapter includes a brief overview, the main content (including code, output, and graphs), and a few exercises designed to help readers practice graphing and modeling in R.

Chapters 1–2 introduce the concepts of time course (repeated measures) data, and motivate growth curve analysis. Basic methods such as *t* tests and repeated measures ANOVA are discussed to highlight what they *cannot* do and what growth curve analysis or multilevel modeling contributes. The chapters also provide gentle introductions to using **ggplot2** to graph data and **lme4** for growth curve analysis, with code examples, descriptions of the syntax and some basic exercises.

Chapters 3–5 cover the application of standard and orthogonal polynomials to model non-linear growth and give ample examples from research on word learning experiments where the functional form of time is non-linear. How to interpret output including the fixed effects and variance components from the **lmer** function in the **lme4** package are also discussed. These chapters also provide examples of how to present results in tables, graphs and written text. An example of how to graphically compare full and reduced models of the same data (Figure 4.5) was particularly helpful. Finally, Chapter 5 covers how to implement different coding schemes (treatment and sum) for categorical predictors and how to use the **glht** function from the **multcomp** package to test specific hypotheses or conduct multiple comparisons.

These chapters were well-written overall; however, the coverage is limited. For example, in discussing how to obtain *p* values for specific parameters, only Markov chain Monte Carlo

sampling, likelihood ratio tests, and normal theory approximations are discussed. The normal theory approximation is suggested, given that the number of observations is much larger than the number of parameters. Whereas this may be true in cognitive sciences where each participant completes dozens of trials, in other contexts, such as biostatistics where a small number of patients with a rare disease participate in a randomized trial with repeated follow-up measures, only a few participants and a few number of observations per person may be available. Relatedly, the author describes the principle of “keeping it maximal” — that random effect structures should be as complex as indicated by the study method or design. Again, this advice applies most clearly in the context of random effects based on the design of an experimental study than it does to observational time course data.

Chapters 6–7 are more advanced and cover how to conduct logistic growth curve analysis (Chapter 6) and how to use the estimate of the random effects from a growth curve analysis as variables in another analysis (Chapter 7). Logistic growth curve analysis is briefly introduced, and results from modeling the same data with a linear, logistic, or empirical logit with weights model are compared. Although some graphs of model results are shown, converting results into odds ratios or average effects on the probability metric are not covered. Chapter 7 shows how to extract the best linear unbiased predictions (BLUPs) and use these to calculate individual effect sizes and also to relate to other variables (e.g., are individuals’ random slopes from separate growth models correlated).

The final chapter is a “cookbook” with seven complete examples with R code showing how to visualize the data, analyze and compare models, graph the results, write the methods, and write the results. This chapter can serve as a nice guide both conceptually and practically with steps for analysis and code examples. For the most part, the examples stand alone so researchers already familiar with growth curve analysis could simply use the last chapter as a reference.

For a book on time course data, notably lacking was discussion of residual structures. There is no discussion of autocorrelation or how to relax the assumption that residuals are independently and identically distributed (such as using the **nlme** package). Readers interested in learning how to do this may also consider the book: *Linear Mixed-Effects Models Using R* by Galecki and Burzykowski (2013), for a recent introductory book with extensive details on residual structures. Although the most common types of growth curve analyses (linear and binary logistic) were covered, models for other types of data (e.g., ordered and count outcomes) were not covered. Missing data were not addressed. The code presented for constructing orthogonal polynomials do not generalize outside of balanced designs.

In summary, this is a good introductory book for applied researchers, particularly from the fields of cognitive sciences and linguistics who have time course or repeated measures data. The writing is clear and easy to follow, without jargon, and assumes little background in R or statistics. Due to the focused nature of the book, topics including missing data, residual structures, and analysis of outcomes beyond continuous or binary were not covered. Therefore, researchers from other fields (e.g., biomedical, ecology) where these topics are important may find the content less applicable.

References

Galecki A, Burzykowski T (2013). *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. Springer-Verlag, New York.

Reviewer:

Joshua F. Wiley
University of California, Los Angeles
Department of Psychology
Los Angeles, CA, 90095, United States of America
E-mail: wileyj@ucla.edu
URL: <http://joshuawiley.com/>