Check for updates

Book Review

Journal of Educational and Behavioral Statistics
2016, Vol. 41, No. 4, pp. 451–454
DOI: 10.3102/1076998616631748
© 2016 AERA. http://jebs.aera.net

A Review of *Longitudinal Categorical Data Analysis*

Hsiu-Ting Yu
National Chengchi University

B. C. Sutradhar. *Longitudinal Categorical Data Analysis.* New York, NY: Springer, 2014. 369 pp. $129.00, ISBN 978-1-4939-2136-2.

Introduction

Longitudinal categorical data appear often in a wide range of areas including sociology, psychology, education, medicine, public health, and so on. However, the methods for analyzing this type of data are not commonly taught in statistical and methodology courses in graduate programs of social science. Traditionally, longitudinal categorical data in social science are handled by log-linear models for repeated observations.

A typical textbook on this topic is *Statistical Analysis of Longitudinal Categorical Data in the Social and Behavioral Sciences* by von Eye and Niedermeier (1999). This book also uses the traditional hierarchical log-linear models to analyze this type of data. However, this approach considers time as a nominal fixed covariate defined through dummy variables, and the categorical response variable is usually obtained using the method of dichotomization. These features limit the possibilities of the traditional hierarchical log-linear models in analyzing longitudinal categorical data. Unlike the traditional method, this book by Sutradhar takes the perspective of "correlation between repeated measures." Therefore, this book opens up a new and alternative approach to analyzing longitudinal categorical data.

Longitudinal Categorical Data Analysis is a text that aims to explore alternative methods of modeling longitudinal categorical data. Even though the title of this book contains the term "longitudinal," both data from cross-sectional and longitudinal setups are discussed. Specifically, this book focuses on longitudinal multinomial data analysis by developing various parametric correlation models for repeated multinomial responses. This approach is quite different from the typical approach of using a marginal model for longitudinal data based on a generalized estimating equation approach.

Specific Contents

This book is written in six chapters. These chapters are grouped into two main parts: Chapters 2 through 4 deal with univariate categorical data, and Chapters

5 and 6 handle bivariate categorical data. Both cross-sectional and longitudinal settings are discussed in the different chapters for the univariate and bivariate categorical data. The first chapter gives a general introduction about the background of the models covered in this book. This chapter also points out the proposed multinomial logit models used throughout this book as the new alternative approach from the traditional log-linear models.

In the cross-sectional setup, a comprehensive review of regression models for both univariate and ordinal categorical data analysis is presented in Chapter 2. Starting from the most basic univariate multinomial fixed effect models, a univariate multinomial regression model is introduced, leading to a cumulative logits model for univariate ordinal categorical data. A detailed and specific derivation, likelihood function, estimation equations, as well as several empirical data illustrations are included.

Chapter 3 discusses longitudinal categorical data analysis. The new and unique parametric correlation model is proposed by relating the current and previous multinomial responses. Specifically, the linear dynamic conditional multinomial probability and the multinomial dynamic logit are used to model the dynamic relationships based on conditional probabilities. In Chapter 3, under the longitudinal setting, the covariate free and time-independent (stationary) covariates are discussed. Chapter 4 discusses univariate longitudinal categorical data analysis with time-dependent (nonstationary) covariates.

In Chapter 5, the bivariate correlations are modeled through an individual random effect shared by both response variables in the cross-sectional setting. In addition, a bivariate normal linear conditional multinomial probability model is discussed in this chapter. Instead of a cross-sectional setting, Chapter 6 extends the discussion to a repeated bivariate multinomial model in the longitudinal setting. This is done by combining the dynamic relationships for both multinomial response variables through a random effect shared by both responses from an individual. This may be referred to as the *familial longitudinal multinomial model* (Sutradhar, 2011), with family size two corresponding to two responses from the same individual. Therefore, this chapter may be treated as an extension of the familial longitudinal binary model described in an earlier book from this author (Sutradhar, 2011, Chap. 11).

Values and Limitations

The book is technically rigorous, and a step-by-step derivation of equations is provided for most cases. The author also covers related details for the model under discussion; for example, model specification, likelihood function, and estimation methods (both generalized quasi-likelihood and the exact likelihood approaches are included). In terms of its computational aspects, the author indicates that the developed formulas were computed using Fortran-90. However, there is no general program available (as a supplementary or online resource),

and this may hinder the empirical applications of the proposed models and methods. Therefore, I would consider this book mainly for theoretical and technical discussion, since the applications of these models would require the readers to have reasonably good computing knowledge and skills.

The author indicates that this book is written for graduate students and researchers in statistics and the social sciences, among other applied statistics research areas. The author also suggests that the part of the book related to univariate categorical data under a cross-sectional setup (Chapter 2) and under a longitudinal setup (Chapter 3) is also suitable for undergraduate students. However, I think only students who have a strong background in statistics can digest the material well. Graduate students who do not have systematic training in statistics may find this book somewhat too technically oriented. Although the comprehensive technical details covered are one of the main strengths of this book, at the same time, this book may not suitable for a standard textbook on this topic. Nevertheless, this book contributes by providing a solid and useful approach to analyzing longitudinal categorical data.

Summary and Conclusion

In summary, this is the first book on longitudinal categorical data analysis with parametric correlation models developed based on dynamic relationships among repeated categorical responses. This book is a natural generalization of longitudinal binary data analysis to the multinomial data setup with more than two categories. Thus, unlike the existing books on cross-sectional categorical data analysis using log-linear models, this book uses multinomial probability models both in cross-sectional and longitudinal setups. A theoretical foundation is provided for the analysis of univariate multinomial responses by developing models systematically for the cases with no covariates, as well as categorical covariates, both in cross-sectional and in longitudinal setups. In the longitudinal setup, both stationary and nonstationary covariates are considered. These models have also been extended to the bivariate multinomial setup along with suitable covariates.

References

Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data*. New York, NY: Springer.

von Eye, A., & Niedermeier, K. E. (1999). *Statistical analysis of longitudinal categorical data in the social and behavioral sciences: An introduction with computer illustrations*. Mahwah, NJ: Lawrence Erlbaum.

Author

HSIU-TING YU is an associate professor in the department of psychology at the National Chengchi University. Between 2009-2015, she worked at McGill University in Canada. Her address is No. 64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei, Taiwan. E-mail:

htyu@nccu.edu.tw. Her research interests focus on the development and application of quantitative methods for analyzing data collected in psychological and educational research. She particularly emphasize issues of analyzing data with multilevel nested structure and repeated measurements.

Manuscript received November 16, 2015
Revision received January 13, 2016
Accepted January 16, 2016