The Stata Journal

Editors

H. Joseph Newton
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

Nicholas J. Cox
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

Christopher F. Baum, Boston College
Nathaniel Beck, New York University
Rino Bellocco, Karolinska Institutet, Sweden, and University of Milano-Bicocca, Italy
Maarten L. Buis, University of Konstanz, Germany
A. Colin Cameron, University of California–Davis
Mario A. Cleves, University of Arkansas for Medical Sciences
Michael Crowther, University of Leicester, UK
William D. Dupont, Vanderbilt University
Philip Ender, University of California–Los Angeles
James Hardin, University of South Carolina
Ben Jann, University of Bern, Switzerland
Stephen Jenkins, London School of Economics and Political Science
Ulrich Kohler, University of Potsdam, Germany
Frauke Kreuter, Univ. of Maryland–College Park
Peter A. Lachenbruch, Oregon State University
Stanley Lemeshow, Ohio State University
J. Scott Long, Indiana University
Roger Newson, Imperial College, London
Austin Nichols, Abt Associates, Washington, DC
Marcello Pagano, Harvard School of Public Health
Sophia Rabe-Hesketh, Univ. of California–Berkeley
J. Patrick Royston, MRC CTU at UCL, London, UK
Mark E. Schaffer, Heriot-Watt Univ., Edinburgh
Philippe Van Kerm, LISER, Luxembourg
Vincenzo Verardi, Université Libre de Bruxelles, Belgium
Ian White, MRC CTU at UCL, London, UK
Richard A. Williams, University of Notre Dame
Jeffrey Wooldridge, Michigan State University

Stata Press Editorial Manager

Lisa Gilmore

Stata Press Copy Editors

Adam Crawley, David Culwell, and Deirdre Skaggs

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-782-8272, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$124	1-year subscription	$154
2-year subscription	$224	2-year subscription	$284
3-year subscription	$310	3-year subscription	$400
1-year student subscription	$ 89	1-year student subscription	$119
1-year institutional subscription	$375	1-year institutional subscription	$405
2-year institutional subscription	$679	2-year institutional subscription	$739
3-year institutional subscription	$935	3-year institutional subscription	$1,025
Electronic only		**Electronic only**	
1-year subscription	$ 89	1-year subscription	$ 89
2-year subscription	$162	2-year subscription	$162
3-year subscription	$229	3-year subscription	$229
1-year student subscription	$ 62	1-year student subscription	$ 62

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

Copyright © 2018 by StataCorp LLC

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LLC. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, fileservers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LLC.

The Stata Journal (2018)
18, Number 2, pp. 485–488

Review of Tenko Raykov and George Marcoulides's A Course in Item Response Theory and Modeling with Stata

Ariel Linden
Linden Consulting Group
San Francisco, CA
alinden@lindenconsulting.org

Abstract. In this article, I review *A Course in Item Response Theory and Modeling with Stata* by Tenko Raykov and George A. Marcoulides (2018 [Stata Press]).

Keywords: gn0076, book review, item response theory, survey development, measurement, instrument, construct, latent variable, Stata

1 Introduction

Item response theory (IRT) is foundational to instrument development and therefore not typically covered in general statistical training. However, IRT is potentially applicable to a variety of measurement problems, making it a valuable methodology for a broader audience. Since the introduction of the `irt` command in Stata 14, Stata users have been able to implement IRT within a familiar software environment. But a fair amount of training is required to ensure that the analytic process is conducted appropriately. *A Course in Item Response Theory and Modeling with Stata*, by Tenko Raykov and George A. Marcoulides, provides both sufficient background theory in IRT and guidance for implementing the procedures in Stata, thereby allowing those with limited or no prior experience in IRT to perform these analyses with confidence.

2 A brief, nontechnical description of IRT

IRT is a statistical methodology for conducting latent-variable modeling in which the responses to items on an instrument are assumed to be explained by one or more latent (unobserved) variables (also referred to as constructs, traits, abilities, etc.). IRT emphasizes the probability of a response for each item as being a function of the level of the latent variable and item characteristics. In the case of binary scored items, the response probability is typically expressed using the logistic function (referred to in IRT as the item characteristic curve [ICC]), in which the probability of a "correct" response (with "correct" representing a "yes" in a yes or no response or a "1" in a 1 or 0 response, etc.) is on the y axis and is plotted against levels of the underlying latent variable θ (theta). The investigator is typically interested in assessing the position of the latent variable where the probability of a "correct" response for an item is 0.50.

© 2018 StataCorp LLC

The further to the left on the horizontal axis of the ICC that this intersection occurs, the stronger the inclination to choose the "correct" response over the "incorrect" response, and vice versa. In the jargon of IRT, an item is considered easier if a "correct" response is obtained at a lower range of the latent variable relative to other items and is considered more difficult if a "correct" response is obtained at a higher range of the latent variable relative to the other items (correspondingly, the model parameter that provides this estimate is called an "item difficulty parameter"). After fitting an IRT model (presumably after determining that it fits the data better than other available models), the investigator may inspect the item information function to see how much information (and where along the continuum) each item contributes about the latent variable. Moreover, the investigator may inspect the test information function—which summarizes the behavior of all items in a single curve—to ensure that the instrument differentiates best between respondents in the desired range of the latent variable. For example, a medical researcher may be interested in developing an instrument to detect early warning signs of impending dementia. Such an instrument would have its best differentiation capabilities in the lower range of scores of the latent trait. Conversely, a researcher developing an instrument to measure perceived pain may desire to have its best differentiation potential in the higher range of scores to ensure that a patient experiences a relatively high level of pain before receiving a powerful narcotic. Finally, an investigator will want to assess whether any item exhibits differential item functioning; that is, whether individuals belonging to different subgroups, but with the same level of the latent variable, have a different probability of responding "correctly" to the item. Although the preceding description of IRT focuses on instruments using binary scored items, IRT can also be used to develop instruments with polytomous items (which may be ordinal or nominal) or instruments with a mix of item types (called hybrid models). Fortunately, the various functions used for developing instruments with binary scored items are readily applied in these more complex models as well.

3 Overview

The first four chapters in this book lay the conceptual groundwork for IRT, including definitions (chapter 1), introduction of the logistic and normal ogive functions as a foundation of IRT (chapter 2), the connection between IRT and classical test theory and factor analysis (chapter 3), and the links between generalized linear models, logistic regression, nonlinear factor analysis, and IRT (chapter 4).

Until chapter 11, the emphasis is on instruments with binary scored items. As such, chapter 5 describes the IRT models used in the binary case. These include the one-, two-, and three-parameter logistic models (which Stata refers to as 1pl, 2pl, and 3pl, respectively). The base model is a 2pl model in which the two parameters estimated for each item are a (called an "item discrimination parameter", which represents the steepness of the ICC at the point where the probability of a "correct" response is 0.50) and b (called an "item difficulty parameter", which represents the point on the latent variable scale where the probability of a "correct" response is 0.50). A 1pl model (sometimes referred to as a Rasch model) is a special case of the 2pl model in which the

item discrimination parameter is assumed equal across all items. A 3pl model introduces an additional parameter g to the 2pl model to account for the possibility of guessing.

Chapter 6 provides a comprehensive tutorial, using empirical data, on how to conduct an IRT analysis for binary scored items using Stata's `irt` command. This involves choosing the best-fitting logistic model (from the three models described in chapter 5) based on information criteria (that is, the likelihood-ratio test, Akaike information criterion, and Bayesian information criterion), transforming the predicted latent scores into T-scores, and plotting the associated ICC.

Chapter 7 describes in greater detail the statistical methods underlying IRT model-estimation procedures that use the maximum likelihood method and briefly discusses how these models can account for missing data under the missing at random assumption.

Chapter 8 introduces additional IRT functions and curves such as the item information function, test information function, and test characteristic curve.

Chapter 9 provides a general framework for instrument development that uses the functions and curves described in the previous chapter.

Chapter 10 is devoted to assessing differential item functioning; that is, whether there is evidence that individuals belonging to different subgroups, but with the same level of the latent variable, have a different probability of responding "correctly" to the item. In Stata, differential item functioning can be assessed using two official commands, `diflogistic` and `difmh`. However, the authors also provide helpful code using `gsem` to perform the more complicated Benjamin–Hochberg multiple-testing procedure to assess differential item functioning.

Chapter 11 shifts from binary scored items to instruments with ordered or nominally scored items and introduces hybrid instruments that may contain a combination of item types. Several new models are introduced in this chapter to handle these types of items, including the nominal response model, the graded response model, the partial credit model, the generalized partial credit model, and the rating scale model. As in the case of models for binary scored items, model selection here is also aided by the Akaike information criterion and Bayesian information criterion indices. The functions or curves that support models for polytomously scored items are generalizations of those for binary scored items and thus are interpreted in a comparable fashion—with one notable exception: whereas the ICC plots the probability of a "correct" response against levels of the underlying latent variable for binary scored items, the category response function is used in polytomous scored items to plot the probability for response in that category as a function of the latent variable. In other words, each category within an item is modeled separately in a category response function, and each item in the instrument is modeled in turn. This chapter also introduces hybrid models, which are fit in Stata using the `irt hybrid` command. In a hybrid model, the user specifies which models (chosen from the array of models available) should be used for specific items in the overall set.

Chapter 12 introduces the reader to multiple item response theory (MIRT) models, which are models used for instruments that appear to reveal more than a single latent

variable based on the response patterns of respondents. After briefly describing the underlying theory and statistical properties of MIRT, the authors provide an empirical example using the `gsem` command (the `irt` command does not currently contain commands for performing MIRT). An important feature of the MIRT modeling process is assessing whether an MIRT model actually fits the data better than a unidimensional model. As before, this assessment is aided by Akaike information criterion and Bayesian information criterion indices.

4 Comments

This book accomplishes what it proposes to do in providing an introductory to intermediate level discussion of IRT and demonstrates its implementation using Stata. The book is well organized and strikes the right balance between theory, statistics, and application. Although instructors will likely find it suitable as a course textbook, individuals who prefer independent study will have no problem mastering the material on their own.

I would change very little about this book, and my suggestions all relate to additions that the authors could consider for future editions. First, in chapter 9, readers would benefit from at least one worked example of instrument construction and development using information functions. This topic is too important to leave the exposition in the theoretical. Second, in chapter 11, it would be valuable to demonstrate scoring (and score transformations) for instruments with polytomous and hybrid items. It is not intuitive if scoring of these instrument types can be performed by replicating the procedure presented in section 6.7 for instruments with binary scored items or if a different procedure must be followed. Finally, chapter 12 currently presents the case in which a unidimensional IRT model is shown to fit the data better than a multidimensional IRT model. It would be instructional to describe a case where the MIRT is considered a more appropriate model for the data and to illustrate how the related functions and curves are generated and interpreted.

In summary, I strongly recommend this book for both students of an introductory course in instrument development and for more seasoned researchers interested in conducting IRT analyses in Stata who may not have been exposed to IRT as part of their statistical training.

5 Reference

Raykov, T., and G. A. Marcoulides. 2018. *A Course in Item Response Theory and Modeling with Stata*. College Station, TX: Stata Press.

About the author

Ariel Linden is a health services researcher specializing in the evaluation of healthcare interventions. He is both an independent consultant and a research scientist in the Department of Medicine at the University of California, San Francisco.