

Review of Statistics with Stata (Updated for Version 7)

Jeremy Freese
University of Wisconsin Madison
jfreese@ssc.wisc.edu

Abstract. The new book by Hamilton (2002) is reviewed.

Keywords: gn0003

1 Introduction

They say that if you build a better mousetrap, the world will beat a path to your door. Then again, because the world will only know how to use the lesser mousetraps that existed prior to your invention, somebody should really be ready with a second door, as the world will also beat a path looking for guidance on how to harness all the murine-catching power of your new apparatus. When Stata was new, its users were primarily researchers who were already familiar with other general-use statistical packages but who saw in Stata perhaps a more elegant and powerful means of capturing the ever-elusive order in data. Since Version 2, Lawrence Hamilton's *Statistics with Stata* has been a prominent guide to showing people what this new mousetrap could do. Of course, Stata strains my extended metaphor by being a better mousetrap that also evolves and improves with each new release, but Hamilton has been up to the task of keeping his guide current. Duxbury Press has now published the fourth edition of *Statistics with Stata*, with revisions for use with Version 7.

Hamilton describes the intent of *Statistics with Stata* as a bibliobridge between books that focus on the teaching of statistics and Stata's own extensive documentation. In other words, *Statistics with Stata* presumes that if one needs an introduction to statistical methods, one will pick up an appropriate textbook, and that if one needs detailed information about a particular Stata command, one will reach for the appropriate Reference Manual. Accordingly, the book is intended mainly to serve an audience who already know how to analyze data or are being taught how to do so, and it offers to show these readers how to do what they want to do in Stata.

This account of its *raison d'être*, however, sells the book a bit short. Given at least the typical training in statistics for the average researcher, I think there is much one can learn about data analytic methods from careful reading of *Statistics with Stata*. It has no pretensions of being a comprehensive or fully systematic consideration of data analytic issues, but it does say many useful things about a wide variety of models. A central virtue of the book as a teaching companion is that it shows the reader a substantial breadth of things that can be done in the course of an analysis; consequently, it not only teaches, but also encourages the student to spend more time with the textbooks

and the Stata manuals to gain a more complete understanding of the craft of analysis. Meanwhile, even veteran users of Stata will pick up some new (or forgotten) information on what the package can do by at least skimming its pages. In other words, if the book is a bridge, it is a bridge that promotes much exploration of both of the lands that it joins.

Statistics with Stata also spans a considerable range of material, especially for a book as ultimately modest in size as it is. After an introduction that focuses primarily on informing the reader of various sources of further help and information about Stata, the book starts off with chapters that introduce the reader to Stata's capabilities for managing data, making graphs, and working with various kinds of summary statistics, exploratory plots, and tables. Then it moves into the work of model estimation by presenting a chapter on ANOVA before tackling the linear regression model. Hamilton devotes chapters to the basics of estimating a regression model, to regression diagnostics, to nonlinear regression, and to robust and quantile regression. Hamilton next provides four chapters on models for other kinds of data: logistic regression (binary, ordered, and multinomial), survival analysis, factor analysis, and time-series analysis. Finally, the book concludes with an ambitious introduction to programming, including parts devoted to Stata's facilities for matrix algebra, bootstrapping, and Monte Carlo simulations.

Yet even this accounting of its contents is incomplete, as Hamilton manages to squeeze still more topics into places where their inclusion may strike the reader as surprising. For example, tests of nonnormality and the ladder of powers are discussed in the chapter on summary statistics and tables, robust standard errors are covered in the chapter on robust regression, and generalized linear models are covered in the chapter on survival analysis. Each chapter begins with a section of "example commands"—lines of Stata code followed by brief descriptions of what the line does—before moving into its rapid consideration of the topics collected in each chapter. Just flipping through these example commands from chapter to chapter is a good way for a new user to get an immediate sense of how simply one can get Stata to do quite sophisticated things.

For all that it covers, the treatment of any individual model or command is obviously brief. However, Hamilton has a real gift for conciseness, and I think the book is remarkable for all that it does manage to convey given its high scope-to-size ratio. Of course, a reviewer who picks up the book already very familiar with Stata is at a considerable advantage in apprehending brief articulations compared to a student who is just learning the program, and one worry about *Statistics with Stata* is how well it could be followed by someone relatively new to Stata. My suspicion is that Hamilton's efforts at clarity do work well with keeping less experienced readers onboard, and I have heard many positive reports on earlier editions of the book from students.

At the same time, given all that the book seeks to accomplish, some of the topics do seem like they are draining space that might have been better used on giving more attention to issues that are likely to be more central to the needs of the book's primary users. To give one example, I would question whether the rarely-used `areg` command needed to be included in the discussion of using dummy variables in regression. An entire chapter of the book that does seem a bit quixotic given its audience is the one on

programming, in that I believe that a reader who is advanced enough to follow the assembled short discussions and actually write their own programs afterwards is probably someone who would be fine with and more satisfied by just using Stata's Programming manual. As a different matter, the graphs in the graphics section are not appealingly enough drawn to win many converts to Stata relative to its competitors, although this perhaps points more to the need for Stata to consummate planned improvements in this part of the package than to anything for which Hamilton can be blamed.

These points, however, are quibbles amidst a much greater admiration for the service that *Statistics with Stata* provides. The book provides an introduction to the mechanics and breadth of Stata that is at once valuable as a guide for graduate students learning data analysis, a primer for experienced researchers starting with Stata, and an edifying reference for those familiar with the package. I already recommend it without reservation to anyone looking to learn more about Stata, and I hope very much that it continues to be updated as newer versions of the package are released.

2 References

Hamilton, L. C. 2002. *Statistics with Stata: Updated for Version 7*. Pacific Grove, CA: Duxbury Press.

About the Author

Jeremy Freese is Assistant Professor of Sociology at the University of Wisconsin–Madison.