The Stata Journal (2004)
4, Number 2, pp. 216–219

Review of Statistics with Stata (Updated for Version 8) by Hamilton

Richard Williams
Department of Sociology, University of Notre Dame
Richard.A.Williams.5@nd.edu

Abstract. The new book by Hamilton (2004) is reviewed.

Keywords: gn0012, introductory, teaching

1 Introduction

I will begin with some wise advice that was given to me years ago and that I have often passed on to others. If you are going to buy only one book on Stata, Lawrence C. Hamilton's *Statistics with Stata* (hereafter SWS) is the one you should go with. Not only is it one of the best books on Stata that you will find, it is one of the best introductory books using any statistical language, period. It is one of those extremely rare offerings that has value both for undergraduates who have never had a statistics course in their lives and for senior scholars who have spent their careers doing empirical work.

I have seen few texts that cover the range of topics that Hamilton does. He begins with the fundamentals of Stata, moves on to basic descriptive statistics and graphical techniques, and then addresses the traditional topics of a one-semester US undergraduate course, such as ANOVA and OLS regression. But, unlike the vast majority of introductory texts, he does not stop there. He goes on to cover such advanced topics as factor analysis, regression diagnostics, nonlinear relationships, time-series analysis, robust regression, logistic regression, multinomial and ordered logit models, panel studies, survival and event-count models, and programming with Stata. At my institution, SWS gets used both in our undergraduate statistics course and in our two-semester graduate sequence, and it works very well in both.

2 Strengths and suggestions

With these platitudes having been duly noted, it is also important to acknowledge that, if Hamilton's is the one Stata book that everyone should own, it is partly because there are so few alternatives. If I wish to teach an introductory statistics course using SPSS, I probably have at least a dozen texts to choose from. If I want to teach that course using Stata, Hamilton wins almost by default. As I read SWS, I repeatedly found myself thinking of things I would like to see done differently or additionally. However, no one text can do all things; many of the comments that follow are therefore not criticisms of the book but rather observations on the kinds of other texts I wish were available.

Of course, since such books often are available using other software, if you are looking for a textbook and are not wed to the choice of a specific program, you may want to consider whether, despite all its merits, SWS is the best choice.

First, as noted, one of the most striking features of Hamilton's book is its remarkable scope. This breadth does not come without tradeoffs, however. Those who teach introductory undergraduate courses might well wish for a text that covers fewer topics, but more in depth. Hamilton's book can be a bit overwhelming at times, and probably at least a third of the material will go unassigned in most introductory courses.

At the same time, for advanced courses, I would like to see a book that covers even more than Hamilton does. Hamilton has chosen a good selection of topics, but there are many other choices that would be equally reasonable. While time series gets a whole chapter, panel studies get only a few pages. There is no discussion of Stata's powerful `svy` routines for handling survey data. Public domain, user-written routines get almost no mention. This is perhaps understandable, given how such routines are not officially supported by Stata and may change considerably and unpredictably across time, but it would still be useful if at least some examples were given of the many excellent programs that are out there.

Second, Hamilton has done an impressive job of compiling examples from a wide range of fields. Indeed, his discussion of the 1986 space shuttle Challenger disaster, showing that it could have been determined pre-launch that Challenger was virtually certain to encounter major problems, is one of the most fascinating and powerful examples I have ever seen in an introductory text.

But, as good as Hamilton's examples are, instructors who feel that students learn statistics best by seeing relevant examples from their own fields of study should realize that Hamilton does not provide them with the kind of support that many other books do. Particularly for an introductory course where many of the students will never run a statistical package again, the benefits of using a superior program like Stata and a fine general book like Hamilton's will have to be weighed against the advantages of using a text that focuses on examples from the course's discipline. The market may not be big enough, but if it is, a *Statistics with Stata for the Social Sciences* or (perhaps more realistically) companion guides oriented toward different disciplines might be of great interest to many teachers.

More generally, there was one other aspect of Hamilton's examples that I found disappointing: none of them employed the large-scale survey datasets that are so common in the social sciences and some other fields. For that matter, there were no large datasets, period: the largest sample size for any of the nonsimulated datasets distributed with SWS has only 271 cases (a three-variable study of ice in Greenland!). Besides leaving out important substantive areas, the failure to include any large datasets limits the opportunity to show how Stata's powerful graphics techniques could be used with them. My experience has been that techniques that are very illuminating when working with a few hundred cases often become rather useless when working with a few hundred thousand. Hamilton cannot cover every substantive area, but he could do better in addressing the kind of large-sample issues common across several fields of study.

Third, SWS is not a stand-alone statistics book, nor does it claim to be. The fact that it is not a stand-alone text, however, certainly does not mean that one cannot learn a great deal about statistics from it. I do not claim to be an expert on every method Hamilton discusses, but I generally found that he made it clear what each technique was intended for and the basics of how to use it. In the areas where I feel I do have some expertise, SWS repeatedly gave me new insights. Particularly illuminating were the frequent discussions of diagnostic procedures and of visual and nonvisual ways of inspecting data and making interpretation of results easier. Those who sometimes feel overwhelmed by the complicated mathematics underlying some techniques will appreciate how Hamilton makes their practical uses clear.

There are, of course, numerous stand-alone statistics books that include extensive examples using other software, and such a text that provided examples using Stata would have great appeal to me. But the main thing I would ask of future editions of SWS is that they be even more thorough in providing intuitive explanations of methods. For example, I doubt that many will gain much understanding from a section that begins "Rotation further simplifies factor structure" (page 322) and then lists, without substantive explanation, what the rotation options are. Here, and in a few other places, a few additional paragraphs could help raise the entire book to the levels of clarity which characterize most of it.

Fourth, most of my comments have been addressed to those who are thinking of using SWS as a textbook or as a means of learning Stata for themselves. But there is another potential audience for this book, which is the fifth edition of a text that has been around since 1990: owners of earlier versions will naturally wonder whether this is an indispensable upgrade or a minor revision that they can do without.

Certainly, there is overlap and repetition from earlier versions. But there are also major changes, the most critical of which are discussions of Stata 8 graphics. Not only is there a separate chapter on graphics, the entire book makes good use of graphical techniques throughout. As Hamilton notes in his preface, also new are discussions of panel data, robust standard errors, and cluster analysis. Hamilton has also continued to freshen his examples, leading to improved discussions in some places. Perhaps surprisingly, Hamilton says little about Stata 8's new menu structure, but personally I think he has chosen wisely to spend most of his time on the command-line interface rather than on pages of menu pictures that convey relatively little information. (A more puzzling omission is the lack of any mention of the `findit` command, which may now be the most popular way of searching for Stata materials and programs on the Internet.) In short, Stata 8 was a major revision, and this edition of SWS is, too. I suggest that owners of previous editions do what I did: get a new copy of this relatively inexpensive book for yourself and let your students or colleagues borrow your old one.

3 Conclusion

In closing, I repeat my earlier advice: if you are only going to buy one book on Stata, then Hamilton's text is where you should spend your money. Not only is it one of the

best Stata books you will find, it is one of the best books in its genre, period. Both beginners and advanced users can benefit from this book.

But to say that it is the best book available is not to say that it could not be better or that little would be gained by having other choices. I personally would like to see greater attention to the substantive issues encountered in the social sciences, discussions of the special kinds of problems encountered with large datasets—a text that was a little less broad for undergraduates and even broader and more in depth for graduates. Future editions of SWS could do some of these things, but no single book could do it all. Hopefully, other writers will strive to meet the high standards set by Hamilton and try to fill some of these gaps.

4 References

Hamilton, L. C. 2004. *Statistics with Stata (Updated for Version 8)*. Pacific Grove, CA: Duxbury Press.

About the Author

Richard Williams is Associate Professor and a former Chairman of the Department of Sociology at the University of Notre Dame. His teaching and research interests include methods and statistics, demography, urban sociology, social psychology, fertility, econometrics, categorical data analysis, event history analysis, and structural equation modeling. He recently began teaching Stata in his courses and is a frequent contributor to Statalist.