



A Review of “Statistical Models: Theory and Practice”: D. A. Freedman. New York: Cambridge University Press, 2005, 424 pages, \$34.95

Paul Barrett

To cite this article: Paul Barrett (2009) A Review of “Statistical Models: Theory and Practice”: D. A. Freedman. New York: Cambridge University Press, 2005, 424 pages, \$34.95, Structural Equation Modeling: A Multidisciplinary Journal, 16:2, 391-395, DOI: [10.1080/10705510902751390](https://doi.org/10.1080/10705510902751390)

To link to this article: <https://doi.org/10.1080/10705510902751390>



Published online: 15 Apr 2009.



Submit your article to this journal [↗](#)



Article views: 155



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

BOOK REVIEW

Statistical Models: Theory and Practice. D. A. Freedman. New York: Cambridge University Press, 2005, 424 pages, \$34.95.

Reviewed by Paul Barrett
Hogan Assessment Systems

The title of the book, its front and back covers, and its Table of Contents lead the unsuspecting reader into thinking this book is yet another merely competent textbook on statistical models for social scientists, students, and teaching faculty. Don't be fooled, as this book is a veritable Pandora's box. David Freedman has written a masterwork. This is the thinking person's book about statistical models. Unlike the plethora of multivariate statistics and modeling how-to books, this one takes a statistical model, examines its logic in depth, examines its assumptions in exquisite detail, and discusses how to interpret the results of such a model. The writing style is conversational, engaging, and intelligent; the reader becomes immersed in the questioning mind of a master statistician. That is no mean feat of writing.

Chapter 1 is about the role of observation, experiments, and statistical models of data within an investigative science, looking at how moving from simple description of rates and observations to regression-type models influenced the kinds of results and subsequent assertions of causality made by several famous scientists. This chapter is really very interesting, built around four major epidemiological series of studies the results of which produced major medical and social interventions: the health effects of smoking, the utility of mammography for the early detection of breast cancer, the work by John Snow in the 1850s on the cause of cholera, and Yule's late 19th-century work on the causes of poverty in England. The description of Snow's work is particularly fascinating. The text of the chapter is followed by a series of exercises, actually rather interesting probe questions about certain aspects of each study. These exercises are followed by endnotes for the chapter, in which Freedman defines key terminology and provides further clarifying and historical notes on features of the studies he discussed. It's all extremely interesting and unusual. This chapter alone would provide 3 hours of deep discussion on the role of observation, theory, the constitutive properties of quantitative measurement, and deterministic versus stochastic models within science.

Chapter 2 is a good, solid introduction to the linear regression model, discussed in the context of Hooke's Law (or the *spring equation* as defined in Giancoli, 2000, where the

equation defines the stiffness of a particular spring). Freedman also gently introduces the contrast between qualitative and quantitative inference here. Two sets of exercises and extra notes are provided in this chapter, Set A following the discussion of simple linear regression, and Set B following the introduction of the basic multiple regression model. Note that this chapter is just 11 pages in length. It is an introduction to the logic of regression, not a how-to chapter.

Given most of the technical material in the book is written using matrix algebra, chapter 3 is a brief overview of essential matrix algebra for regression models. This is a 12-page chapter, with almost 6 pages devoted to exercises. Chapter 4 is a big chapter on multiple regression, ordinary and generalized least squares, essential equations, parameter estimation procedures, and the logic and interpretation of regression equations. There are five sets of exercises in this chapter, along with 30 discussion questions.

Chapter 5 introduces path and structural models, using a study from Blau and Duncan (1967) on social stratification in the United States as the vehicle to explain many of the procedures and concepts of these methods of analysis. Things get very interesting when Freedman discusses the assumptions that are involved in using these models, especially with regard to how they are used to model causal processes. As Freedman notes, these models do not infer causation from association, but rather assume causation from the outset. About nine pages are devoted to understanding how researchers typically interpret path models, and then he atypically unpacks these typical interpretations in exquisite detail to show how so many of them are dependent on mostly untested assumptions. Freedman does this by holding a hypothetical conversation between two researchers, one trying to explain results to another, with the other asking some rather awkward questions. It is here that Freedman also introduces the concept of a response schedule, which is his preferred way of formalizing the assumed causal relation between one or more input variables and their effects on a measured output variable. Considering regression models in this way forces investigators to really think about what they are trying to assert within a regression and path model, and the extent to which the lack of control over input variable quantities requires some very strong assumptions to be made about how these quantities might be considered causal for some particular output. This issue is taken up further in Freedman (2006). At the end of this chapter is a thought-provoking quotation from Otis Dudley Duncan, the coauthor of the original Blau and Duncan (1967) study discussed in this chapter. Freedman observes that later in life, Duncan's views of statistical modeling in the social sciences became very negative:

Coupled with downright incompetence in statistics, paradoxically we often find the syndrome that I have come to call *statisticism*: the notion that computing is synonymous with doing research, the naïve faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories or the “importance” of the causes of a “dependent” variable. (Duncan, 1984, p. 226)

When reading the complete book quotation, it is clear that Duncan is railing against what would now be considered the status quo in modern social sciences. That word *statisticism* might also be joined here by the word *methodolatry*, the term introduced into the literature by Janesick (1994) and extended into the health and social sciences by Chamberlain (2000) that defines the preoccupation with selecting and defending methods to the exclusion of empirical phenomena

to be explained. Indeed, method becomes the paramount concern of social scientists, and not the better understanding of the phenomenon of interest. This is a point taken up by Freedman as the conclusion of his book.

Chapter 6 is a robust presentation of maximum likelihood estimation. Here Freedman also introduces the concept of a latent variable, although specifically from the viewpoint of statistical estimation theory rather than with any epistemological focus. Here, a paper by Evans and Schwab (1995) on the relative educational success of Catholic over state schools in the United States is used as the target article to exemplify the various concepts developed in the chapter. This paper is also reprinted in full at the back of the book. Mostly probit and logit regression models are discussed. Again, the chapter is replete with exercises, as well as 20 discussion questions and the usual endnotes to the chapter. This will be hard going for many graduate psychology students, and probably represents more of the kind of training required from young statisticians rather than potential psychological scientists.

Chapter 7 is all about the bootstrap and developing estimates of parameters using resampling procedures. Straightforward and interesting in places, I suspect the coverage here would require much more substantive class input by an instructor.

Chapter 8, the final chapter of the book, is all about the use of simultaneous equation models in the social sciences. It begins, though, with a discussion of such models in the context of economic models and instrumental variables. But, in section 8.5, Freedman discusses an example from Rindfuss, Bumpass, and St. John (1980) on the relationship between education and fertility in women's lives. This paper is reprinted near the end of the book. In section 8.7, he goes on to discuss two-stage least squares in the context of Schneider, Teske, Marschall, Mintrom, and Roch (1997), who studied the effects of the choice of school on subsequent social capital (the role of sociability in society at large, which reinforces democratic institutions). This paper is also reprinted at the end of the book. Overall, this chapter is a thoughtful discourse and exposition of the logic, assumptions, and interpretations involved in using such models. But, in section 8.9, "Issues in Statistical Modeling," eight pages are devoted to what is perhaps the most thoughtful and intriguing portion of this book. It is thoughtful because Freedman engages the reader with a series of insights, questions, and answers that cause the reader to consider what we think we are achieving with the use of sophisticated statistical models in social sciences in contrast to their actual contribution to our science. On page 194, Freedman provides a list of data model methodologies that he suggests are the late 20th century's "Philosophers' Stones" (p. 194) of data analysis, the means by which the mere application of a model will turn observational data into a meaningful causal explanatory model. Almost every data model taught by quantitative social science faculties is contained within this list. On page 195, he provides a list of typical responses to the claim that these represent statistical alchemy. This section concludes with a paragraph that summarizes the overarching theme running through this book: Many of those who promote the use of advanced statistical models in the social and behavioral sciences seem unaware of the strong assumptions on which rest the claims of validity of such models. The last sentence of the book is truly profound: "The goal of empirical research is—or should be—to increase our understanding of the phenomena, rather than displaying our mastery of technique" (p. 200). This is, for all intents and purposes, exactly the same message conveyed by the concepts of Duncan's *statisticism* and Janesick's *methodolatry*.

The astute reader of this review might have noticed that the book has 414 pages, yet it apparently ends on page 200. That is so, for the next 66 pages constitute references and

answers to all the exercises given in each chapter. From page 267 to 287, some guidance is provided on using UNIX, and MATLAB, as part of computer labs associated with topics in the book. Pages 289 to 403 are devoted to the reprints of the key articles discussed in various chapters. Clearly this is an unusual book!

As stated in the preface, the intended readership of this book is advanced undergraduate or beginning graduate students in statistics. It is also recommended for students and professionals in the social and health sciences. I am not so sure about its applicability to students in the psychological sciences. Being aware of how research methods are taught in many social science departments, which Gigerenzer (2004) described as a series of “statistical rituals” (p. 587) within his article “Mindless Statistics,” it is difficult to see how the content of this book cannot but undermine the *run-the-numbers* focus of many social science graduate methods courses. It is similar to suggesting that Michell’s (1990) book on the logic of psychological measurement might be introduced as a graduate text into a typical psychometrics class! The content of both books has the capacity to completely undermine the use of carefully ritualized formulaic learning styles within a typical advanced psychology methods class. But, as Erich Lehmann, Emeritus Professor of Statistics at the University of California, Berkeley, notes on the back cover of Freedman’s book, it provides a critical assessment of the logic of the methods, and how their results are interpreted and reported by social scientists. It is not a book devoted to procedural rules setting out how to compute parameter estimates, but more one devoted to the strategies, interpretations, and implications of the use of classes of regression-related methods.

In my view, this is not a textbook that might be used by a class to learn how to analyze regression, path, or structural equation models using package software such as SPSS, SAS, STATISTICA, AMOS, or *Mplus*. Rather, it is best introduced into the latter stages of a methods course, where the critical evaluation of statistical data models as tools for investigative science is to be a significant component. Such a course might also introduce articles by Gigerenzer (2004), Breiman (2001), Schonemann (1984), Feynman (1974), and Michell (2001) for substantive consideration. These diverse authors all ask serious questions about the role of measurement and data models within quantitative psychology, and just how far the results of such modeling exercises can address issues of causality, given the veritable raft of assumptions on which their validity is predicated.

Although Freedman does indicate that the book might be suited to advanced undergraduates, this probably applies more to statistics majors than social science students. My view is that the book does require a fairly strong grounding in basic statistical theory, probably up to and including a graduate course in multivariate statistical methods. As noted earlier, chapter 6 on maximum likelihood estimation would require more formal mathematics and algebra than is usual for social science students. I think the optimum time to introduce this book to students might be when the effort of learning the procedural knowledge associated with multivariate statistical methods is reduced (as after a substantive course component of multivariate methods); then this book might be introduced to students as the final part of a course that requires more of a reasoning component in terms of choosing and applying statistical models to particular kinds of research questions. Alternatively, this book could be used very effectively within graduate courses that involve examining the philosophies of science, measurement, and methodologies, where the core logic, underlying philosophies, and assumptions of various methodologies applied within scientific investigation are the primary focus. One consideration for a lecturer thinking of introducing the book into an existing class is that it is a substantive text, and would

not easily augment a conventional statistical methods textbook, or fit seamlessly into a typical graduate social science methods class. The content does pose serious questions about adducing causality using regression-based methods applied to observational data. Perhaps if the topic of inferring causality forms a major part of any course, then this text would be extremely valuable for its insights and counterarguments to what is probably the status quo concerning would-be causal modeling as exemplified by the use of structural equation modeling within the social sciences. The book encourages thinking and debate. If the message embodied in the quote from Duncan, and in the closing paragraph from Freedman, is accepted as largely valid, then students and staff might find themselves severely questioning the utility of the models they have spent so much time absorbing.

Whether or not Freedman's book might become a routine textbook for quantitative psychology graduate students is a moot point. Perhaps it might find its way into niche methods courses with a significant theory component. However, I would recommend this book be bought by every teacher of research methods. In its own quiet, conversational, and reasoned way, it is a landmark text. When I asked myself what limitations this book might have, it has only one as far as I can see: If you are looking for a statistical cookbook, this one fails miserably. As I said at the beginning of this review, this is the thinking persons' book about statistical models, their assumptions, applications, strengths, and weaknesses.

REFERENCES

- Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*, 199–231.
- Chamberlain, K. (2000). Methodolatry and qualitative health research. *Journal of Health Psychology*, *5*, 285–296.
- Duncan, O. D. (1984). *Notes on social measurement*. New York: Russell Sage.
- Evans, W. N., & Schwab, R. M. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics*, *110*, 941–974.
- Feynman, R. P. (1974). Cargo cult science: Some remarks on science, pseudoscience, and learning how not to fool yourself. *Engineering and Science*, *37*, 10–13.
- Freedman, D. A. (2006). Statistical models for causation: What inferential leverage do they provide? *Evaluation Review*, *30*, 691–713.
- Giancoli, D. C. (2000). *Physics for scientists and engineers* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587–606.
- Janesick, V. J. (1994). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209–219). Thousand Oaks, CA: Sage.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, *36*, 211–217.
- Rindfuss, R. R., Bumpass, L., & St. John, C. (1980). Education and fertility: Implications for the roles women occupy. *American Sociological Review*, *45*(3), 431–447.
- Schneider, M., Teske, P., Marschall, M., Mintrom, M., & Roch, C. (1997). Institutional arrangements and the creation of social capital: The effects of public school choice. *American Political Science Review*, *91*, 82–93.
- Schonemann, P. (1994). Measurement: The reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 149–160). New York: Walter de Gruyter.