REVIEWS

Edited by **Jeffrey Nunemacher**
Mathematics and Computer Science, Ohio Wesleyan University, Delaware, OH 43015

Statistical Models: Theory and Practice (Revised Edition). By David A. Freedman. Cambridge University Press, New York, New York, 2009, xiv + 442 pp., ISBN 978-0-521-74385-3, $40.

Reviewed by **Rebecca F. Goldin**

Statistics can be remarkably counterintuitive. Consider, for example, the mathematician who moves from one city to another and raises the Intelligence Quotient (IQ) of both cities. Sounds strange at first glance, but if the first city's average IQ is much higher than the second one and the mathematician's IQ is somewhere in between, it will naturally be true.

Our inborn statistical intuition doesn't always serve us well, but quantitative reasoning has become critical in everyday life and throughout science. Data are simply everywhere, and we can't make any sense of them without statistical models. For this reason alone, it is particularly important that a statistics course give students a deep understanding of the underlying concepts. For those who really want this depth as well as the statistical formulas, *Statistical Models: Theory and Practice* by David A. Freedman is one of the most refreshing books that a student of statistics could find.

Written for an advanced undergraduate or beginning graduate student in statistics, Freedman's book focuses on central statistical techniques used in actual studies. Mathematicians who are not experts in statistics will enjoy the meaningful presentation of a topic that has deep reverberations throughout society. The ideas are illustrated through published papers examining controversial and current social and medical issues, from the success (or not) of Catholic schools to the benefit (or not) of hormone replacement therapy. The papers are reprinted in their entirety at the end of the book, which makes the exercise of critiquing the methods quite hands-on and illustrative. Unlike many books filled with formulas for students learning a difficult topic, *Statistical Models* aims high for comprehension. Not only does the book explain the construction of standard statistical models, but it also highlights the assumptions underlying the models.

This understanding is critical because statistical conclusions can have social and political consequences. There are ongoing debates about the validity of certain statistical models. A good example of this is how comparisons among public, charter, and private schools continue to spur debate and even legislation about school choice, vouchers, and evaluation criteria for "success."

The public is distrustful of, yet vulnerable to, statistical conclusions announced in the popular press. Only a few years ago, hormone replacement therapy (HRT) was thought to reduce cardiovascular disease, based on large observational studies [**1**]. Not long after, it was discovered that HRT possibly increases cardiovascular disease [**2**]. From a nonstatistician's point of view, statistics may obscure and confuse the facts. Statistical statements should not be trusted, as any "finding" may be contradicted, with new, improved statistics to back it up, within ten years.

doi:10.4169/000298910X521733

Statistical models have a precarious role in social science studies of various sorts, as well as a strained relationship at times to medical research. Models provide a basis for the discovery and "proof" (by overwhelming evidence) of associations between observations, and they often point to causal relationships; none can be so clear as that between smoking and lung cancer. But they also rely on assumptions, and these assumptions can easily be obscured by the popularity of a technique, or rather, the eagerness of researchers to do an analysis. Despite ignorance of the assumptions or through intentional disregard for them, social science researchers frequently use inappropriate models. This book teaches as much about the assumptions as it does about the techniques. For this feature alone, *Statistical Models* offers the opportunity to teach —and learn—statistics in a way that squarely connects theory to practice.

The U.S. educational system's current emphasis on "measurable" results has brought about a new focus on, and requirement for, statistics. It is common now to advocate evidence-based medical care, and even social programs are required to show their efficacy through designed experiments and statistical analyses. Yet as we sort through the change in culture behind mandates to prove scientific and social claims, we often neglect to ask whether our models are appropriate. Even people trained in statistical methods will not always tease out the assumption that disturbance terms in a model are "independent and identically distributed"—an assumption that may not fit the circumstance in which the model is used. In the same vein, *Statistical Models* gives a short historical account of the skepticism surrounding the use of statistics in the social sciences. Statistical violations abound, from ignoring large numbers of parameters in economics, to multiple testing, to assuming different individuals respond equally to input variables.

But though we as a society would like to measure things, we are actually rather poorly trained to do the measuring. This is not a comment about our poor ranking on math tests in international comparisons, but rather about the state of the art of statistical modeling. Freedman points to the many philosophers' stones in modeling; these are tools that social science modelers use to magically turn data into conclusions about social phenomena, without regard to the underlying assumptions. The list is over 60 items long. A subsequent list of modelers' responses to criticism is sobering, if partly tongue in cheek. It includes comments such as "The biases will cancel" and "You can't prove the assumptions are wrong." Freedman summarizes a large body of literature in which critics argue that statistics has little to offer to political science (and, in particular, that rational choice theory has not advanced knowledge in empirical studies of politics). Even in the few circumstances in which the models and results of social phenomenon can be held up to close scrutiny, he notes serious problems with the models. As he puts it, "Taking assumptions for granted is what makes statistical techniques into philosophers' stones."

In the spirit of his own wisdom, Freedman details the assumptions behind the models he describes in this book. In a discussion of multiple regression, for example, Freedman provides a careful though brief description of the matrix notation and how it can be unfolded into several regression models. He then highlights the underlying assumptions connecting the data to the model. Some of them are phrased in abstract language (such as that the entries of the random error term in a multiple regression model are independent and identically distributed). But others he presents through examples, both hypothetical and actual, to get to the heart of issues arising especially in the social sciences. Importantly, he notes that the models are irrelevant when the assumptions are not met, and that it can be difficult to verify the assumptions.

A pointed discussion of the pitfalls of multiple testing places blame squarely on the shoulders of the many investigators who examine their data before deciding how to

analyze it, or neglect to report the statistical tests that may not have supported their theses. In a section damningly entitled "Data Snooping," Freedman not only notes the obvious fact that multiple testing is more likely to lead to a spurious result, but also takes the reader through the mathematical details of a computer-generated set of random data—and how it can nonetheless provide some "dazzling t-statistics." He emphasizes the need for replicating studies, or at least cross-validating them (using only some of the data to generate the model, and testing it on the rest of the data) to see if a model is a good fit.

The perpetual skeptic repeats the mantra "correlation is not the same as causation" *ad infinitum*. And yet, herein lies the conundrum. How can we do effective research, especially social science research, without occasionally making causal claims? As Freedman puts it, "Statistical techniques can indeed be rigorous—given their assumptions. But the assumptions are usually imposed on the data by the analyst: this is not a rigorous process."

Freedman tackles the question of causality in social science head-on. He accompanies the reader through some of the difficult and subtle statistical methods used to tease out cause, such as path models. In a beautiful description of a hypothetical scientific discussion of a fairly simple path model, Freedman explains the assumptions that are causal, and those that are statistical. A causal assumption, for example, might be that a summer boot camp in mathematics would improve scores on a placement exam in direct proportion to the number of hours spent in the boot camp. A statistical assumption would be that the effect of the number of hours for one individual would not have an influence on the score of another person—this assumption may be reasonable for test scores in mathematics, but less so for exposure to and subsequent development of infectious diseases, he points out. The inability to validate the causal assumptions, as opposed to the statistical ones, is the root of what he calls "iffiness" especially in the social sciences. The blunt truth is that "path models do not infer causality from association. Instead they *assume* causation through response schedules, and—using additional statistical assumptions—estimate causal effects from observational data."

Even the best social science studies have some strengths and some weaknesses. Perhaps for this reason, Freedman is reluctant to put much faith in social science research. Yet he does acknowledge that some of this work is done better than other work. He frames his criticism of this research in light of the importance of recognizing the limitations of any model, especially if one wants to employ it to examine a complicated question. Full disclosure is extremely important, as is a justification (perhaps using qualitative research) for the assumptions made. It is beyond the scope of the book to take on the benefits of qualitative research and how it can inform the assumptions used in quantitative research. Unfortunately, this neglect could make a reader think that no social science research has merit.

Statistical Models has several advantageous features for learners of statistics in addition to those mentioned earlier. Somewhat atypical exercises and questions follow each section. Plenty of questions with numeric answers are posed requiring the use of the just-developed methods. Many additional questions get to the heart of the meaning underlying the methods. In the first section, for example, a student may be asked whether a particular term of an equation is a random variable or a parameter. Discussion questions abound, including the kind of qualitative subjective-response questions that statistics books tend to shun, such as "Is this a good idea?" The exercises have a wonderful mix of concrete, standard statistical problems involving computations, and thought-provoking exercises in the more artful aspects of practicing statistics. While some of the questions may be open-ended, they have good, correct answers, most of which are in the back of the text, providing for more learning opportunities.

Examples of social science research are discussed throughout the book, and these article reprints make evaluating the worthiness of the conclusions a valuable and open exercise. Freedman rips apart a paper finding a causal relationship between a woman's education level (the purported cause) and the age at which she first becomes a mother. The study's claim is that education level has a causal impact on when a woman has her first child. Freedman takes the reader through the statistical techniques used to make the conclusion, and then explicitly lists the questionable assumptions made by the researchers. The model's dependence on these (unjustified) assumptions makes a strong case against the validity of the findings.

Half of the book is composed of references, solutions to exercises, and reprints of papers discussed in the text. This turns out to be a wonderful source of exploration for the material covered by the book, and I found myself flipping to the solutions frequently, particularly on the discussion questions. This creates the illusion of making the reader an expert at the kind of reasoning that takes years to develop.

No book can do everything, and this one is no exception. While the book states the meanings of terms it introduces and formulas it uses, it does not belabor the point. Students are expected to have a strong background coming into this book; topics such as regression and standard deviation are reviewed in a way that encourages a second look, but a student who hasn't mastered them before will struggle. There are few examples that speak to the "how to" rather than the "why" behind the ideas, which keeps the book focused but may lose students who need repetitive reinforcement. If it is used as a text, many students would need excellent lectures and computational examples to accompany each of the new ideas.

In some sense, the book is as much about the limits of statistics as it is a how-to book for learning statistics; understanding how it works and when it works confers power and also diminishes it. Unfortunately, David Freedman died in 2008, but his concise reading of this power balance cannot be denied. The more we mathematicians understand statistics, the more we can contribute to the public conversation about an assortment of professionally relevant issues, such as math education.

An appropriate joke comes to mind. One college student says to another, "I took this statistics class, and I learned that correlation does not imply causation." The other student responds, "Seems like the class was pretty influential." The first replies, "Well, maybe. . . ."

REFERENCES

1. F. B. Hu and F. Grodstein, Postmenopausal hormone therapy and the risk of cardiovascular disease: The epidemiologic evidence, *American Journal of Cardiology* **90** (2002) 26F–29F. doi:10.1016/S0002-9149(01)02219-6
2. J. E. Manson, J. Hsia, K. C. Johnson, et al., Estrogen plus progestin and the risk of coronary heart disease, *New England Journal of Medicine* **349** (2003) 523–534. doi:10.1056/NEJMoa030808

George Mason University, Fairfax, VA 22039
rgoldin@math.gmu.edu