Amazon.com: Jim Grace, Lafayette, LA's review of Causal Inference in Statistics: A Primer

Verified Purchase(What's this?)

This review is from: Causal Inference in Statistics: A Primer (Kindle Edition)

The book by Judea Pearl and collaborators Madelyn Glymour and Nicholas Jewell, Causal Inference in Statistics: A Primer, provides a concise introduction to a topic of fundamental importance for the enterprise of drawing scientific inferences from data. The book, which weighs in at a trim 125 pages, is written as a supplement to traditional training in statistics and I believe it fills that role admirably. I was very excited when my copy arrived because I am one of those folks who thinks that most statistics texts provide only the technical specs for quantitative science, not the driver's manual that is needed by researchers who collect and interpret data. After reading it, I think the book is going to be a big hit with both scientists and practicing statisticians. I believe this book will also prove to be useful support for those who teach statistics and data analysis, because the current omission of causal principles in most curricula is an intolerable oversight we must correct.

The book starts off by challenging the reader with the intriguing proposition that data, by themselves, lack sufficient information to permit proper causal analysis. What is required for sensible evaluations of data are causal hypotheses. Clever, simple examples are used to show that if we make the wrong scientific assumptions about how a system works, we can derive very incorrect conclusions from our data. This illustration sets the stage for the rest of the book by leaving us wondering, "What additional set of rules would we need in order to draw causal inferences from data?"

Chapter 1 does a rather brilliant job of providing the minimum essential set of background information for the task at hand. Basic concepts such as conditional probability and conditional independence are defined, along with essential quantities and relationships, to set the stage for later computations. After a few pages, the book then departs from conventional treatments by presenting the elements of graph theory as an equally-important set of background information. Graphs, specifically probabilistic causal networks, represent one of the key pieces that has been missing from the field of statistics, but that is absolutely essential for representing and evaluating causal hypotheses for analysis. The reader simply needs to get to page 24 to begin to encounter the unique information in this highly readable treatment. As Chapter 1 continues, probability theory and graph theory are married though the combination of the "Structural Causal Model", which specifies the variables and connecting functions, and the "Graphical Causal Model", which summaries the causal logic of network relations.

In Chapter 2, in only a few pages, the book presents the core "rules" that establish much of the logic for causal analysis within a graph-theoretic framework. Again, the book is truly outstanding in its capacity to distill fundamental ideas to their basics and clearly illustrate with examples. Following this treatment, Chapter 3 then begins to move the reader into a thorough consideration of the interventionist perspective. In essence, causal modeling asks questions about the outcomes of interventions – "What would happen to Y if we were to change X?" Of course sometimes we have information from manipulative experiments, but the greater challenge is to address this question using observational data and causal rules. In this chapter, the rules of engagement are presented. We encounter new mathematical concepts, like the "do" operator and formulae for adjusting for covariates and calculating causal effects. We also encounter rules like the backdoor and front-door criteria. A central feature of causal networks, mediation, is presented and described. Chapter 3 ends by transitioning from general rules that apply to models of all forms to illustrations obtained through reference to linear Gaussian systems. This final set of examples connects the graph-theoretic perspective with more traditional formulations and examples of structural equations. Here more direct comparisons between, for example, regression coefficients and structural coefficients are made. An elegant and crystal clear introduction to instrumental variables ends the chapter and in the process, links the new material presented in this book with yet another historical body of causal modeling literature. It is impressive to see this accomplished in such a compact fashion.

Chapter 4 turns to a topic that will be unfamiliar to many as a formal subject – counterfactuals. Simply put, counterfactuals are questions about, “What would have happened to individual i if they had not been exposed to treatment X=1 (if they had not received the drug treatment)?” This seemingly innocent question, as the chapter goes on to reveal, unlocks much additional power derived from the causal modeling system presented in the book. To begin with, counterfactuals lead us necessarily from the population to the individual level, since these are questions about what would have happened to an individual if a different choice or event had happened in the past. Considering the individual level, we begin to realize that all along we have had unique information about individuals that has been ignored via summarization. With counterfactuals, ignoring is no longer appropriate. At the outset, the reader will assume perhaps that the counterfactual question is an impossible question to answer, even with randomized experiments. If individual X(1) is included in the treatment group that received a placebo, how are we to know what might have happened if they had actually received the drug? Surprisingly, a general solution to this problem is offered using the logic of the Structural Causal Model and the fundamental law of counterfactuals. Following a series of illustrations developed for a variety of situations, the chapter ends with a summary of essential information in the form of a computational toolkit for causal analysis. Clearly, this book goes beyond an exposition of ideas to provide the reader with a functional knowledge of causal analysis principles.

Throughout, this lucid and concise book explains concepts through the presentation of multiple, simple examples – a strategy that works exceptionally well, making this the most accessible presentation of this material I have read. The reader will be well rewarded for buying and reading this book and I recommend it with enthusiasm for both practicing scientists and students of statistics.