PSYCHOMETRIKA
2013
DOI: 10.1007/s11336-013-9389-2

# BOOK REVIEW

ROSENBAUM, P.R. (2010). *Design of Observational Studies*. NY: Springer. 385 pages. US$129.00. ISBN: 978-1441912121

Design of Observational Studies (DOS) is an introductory textbook on analyzing observational data built on permutation-based inference, matching, and sensitivity analyses. The book is both a pragmatic how-to manual and a delightfully written manifesto. It is quite refreshing to read a textbook that has any personality whatsoever, not to mention one with a philosophy quite different from what one typically encounters. The book contains four parts. The first, in large part a distillation of Rosenbaum's prior book "Observational Studies," outlines an overall approach for tackling observational studies and gives background material on analyzing randomized trials. The second and third parts give further detail and techniques to accommodate various circumstances. The final part ties the book together and provides an overall philosophical framework for analyzing data.

Rosenbaum's light writing style combined with a somewhat dry humor makes for a pleasant read. Stylistic features such as the heavy repetition of a phrase, e.g., "if the null were true" (pp. 29–37), gives rhythm to the prose, but can sometimes make it a tad slow. DOS is not a particularly dense book; as Rosenbaum says in Chapter 1, he is talking about statistics, not necessarily writing about them. Due to "talking," Rosenbaum goes to some lengths to avoid technical detail so we can focus on the overall ideas behind the decisions a statistician has to make when analyzing data. But this is no puff piece; Rosenbaum does not shy away from difficult, deep ideas. It is also a useful reference, containing many worked examples, a decently thought out index, sections on further readings at the end of each chapter, and several appendices and glossaries, including an analysis FAQ, at the back. DOS extensively uses fascinating data sets, but often also has toy examples running in parallel to aid comprehension.

To analyze an observational study, Rosenbaum argues, we need to look for a source of randomness; hence "an observational study may begin with an opportunity, an arbitrary capricious chaotic disruption of the ordered world of everyday" (p. 114). And in this text, matching is the tool for exploiting this randomness because it can, if appropriately used, pull a de facto randomized experiment, in particular a matched-pairs experiment, from observational data. Matching comes in part from the goal of transparency. "In matching, people remain intact as people. In contrast, in model-based adjustments, the people themselves disappear or recede into the background to be replaced by features of the model" (p. 322). Such philosophical underpinnings bring this book alive, and remind us of what we should be focusing on when conducting statistical analyses.

The presented method for analyzing a matched pairs experiment, once one is in hand, is the permutation test. This test is magnificently introduced in Chapter 2, which gives a tour of several treatment impact estimators applied to the same worker-training data set. Here Rosenbaum presents a completely worked out table for each analysis approach using a miniature data set of only five pairs of workers, and also concludes each test with the results for the full data set. This allows the reader to both see exactly what is happening and also see its implications in practice. By the end of the chapter, the concepts behind permutation methods are well explicated, and furthermore readers are left with a sense of the many possible test statistics and definitions of treatment effect one might use, depending on circumstance. It is quite refreshing to see an explicit discussion of the great diversity of estimators that exist.



Published online: 28 November 2013

Such discussion of what is possible is a hallmark of this book. This is very well illustrated by Chapter 4, where Rosenbaum underlines the idea of testing for competing theories with a very nice series of examples and case studies including Galileo. Rosenbaum also here introduces the concept of replication, distinguishing it from repetition. Chapter 5 carries this further, with talk of multiple control groups, looking at multiple outcomes, placebo tests, and other approaches for exploiting features of the natural world in order to determine causality. It is quite nice to have a library of things to look for and techniques to try from a very high level. These chapters underscore the importance of thinking about the context one is studying, and contain a lot of good advice on how to carry out that thinking.

Part II (Chapters 7 through 13) goes into matching in more detail. Following a detailed overview of matching, including propensity score matching and exact matching, there are several short chapters that provide a wide array of matching approaches with discussion on how these various techniques might be more or less appropriate in different contexts. As before, Part II is chock full excellent examples and interesting problems. Chapter 13 is devoted to pragmatic concerns of how to do matching in R, a useful supplement for the applied researcher. One thing that does seem odd in its absence is any discussion of matching with replacement, and other matching approaches that are more similar to weighting adjustment. Given that some very specialized forms of matching, such as non-bipartite matching, are discussed with examples, the absence is conspicuous.

But an observational study, even after matching, is no experiment, and this has to be addressed. A major tenant of this book is that the main concern in an observational study is bias, represented as a departure from the assumed randomization of an ideal experiment, and therefore our inference should focus on this issue. The recommended method for handling this departure is to conduct a sensitivity analysis, which boils down to (in this book) bounding the odds of a given unit being the treated one for each matched pair.

This approach is introduced in Part I and is extensively expanded in Part III. Part III is the section that would possibly feel the most alien to many readers. A focus on sensitivity over uncertainty has direct consequences on what sorts of estimators are more or less efficient when analyzing data. From a sensitivity standpoint, even with unlimited data one estimator might fail while another might not, given a minor model misspecification, even if the former is classically more powerful under correct model specification. Therefore, Rosenbaum argues, we should select a statistical analysis plan by considering estimators' relative resiliences to misspecification rather than their precisions. This perspective shift opens up a whole new area for statistics, with different results on, e.g., how the strength of an instrumental variable, the homogeneity of the units, or using dose level in the analysis might impact the power of an analytic approach.

The book closes with Part IV, an argument for the role of planning as being the foundation of any convincing analysis. A plan gives structure, prevents fishing, and guides transparency. A plan is a rich form of argument that provides a framework for thinking of an entire arc of a sequence of tests. As Rosenbaum puts it, many pitfalls can be avoided with a moments thought and "a moment's thought before the fact is readily available; it just takes a moment's thought" (p. 328). Talk of overall plans raises several ideas not touched by many introductory texts such as detailed advice on how to handle multiple testing issues rather than silence, and a method for explicitly testing for equivalence rather than a stern lecture on how one cannot affirm the null. Of particular interest to me was a case study using medical records where a qualitative investigation into the success of a matching attempt led to a much better overall analysis. In this are real hints as to how qualitative and quantitative research could really inform each other in a principled and highly productive way, a holy grail we all possibly should be seeking.

DOS includes several curious side paths (sometimes in chapter appendices, sometimes in small subsections easily recognized as ignorable), which leaves something for everyone, and reminds us that there are many possible roads to take when analyzing data. The book provides

very solid methodology for a fairly narrow (but flexible) type of analysis centered on matched pairs data, and also provides excellent advice for thinking about analyzing *any* data. People very new to statistics could read this book quite readily, with the caveat that mathematical notation, language, and prose is used far more than in most classic introductory statistics books. This book seems most useful for statisticians and applied statisticians who are looking to learn about a new area that they might not have been exposed to before.

Design of Observational Studies deftly illustrates how cleverness and deep thinking about the particular problem at hand can result in some very nice and clean approaches to analyzing data. On the flip side, it ignores other statistical methodology pretty much completely (e.g., regression, not even to critique it), or treats them oddly (e.g., instrumental variables with a strong dose-type model and no local treatment effect view in Chapter 5) and in my mind does not really dive into some of the shortcomings of the statistical techniques proposed. For example, if one is interested in confidence intervals and estimation more than hypothesis testing, permutation-based methods get a little tricky as they implicitly rely on a constant treatment effect (unless something really fancy is happening). As a fan of permutation-based methods myself, I was hoping to see some discussion of how to take these techniques further to address this important area of concern. Perhaps that will be Rosenbaum's next book.

HARVARD UNIVERSITY *Luke W. Miratrix*