Matching for Adjustment and Causal Inference
Escuela de Invierno en Métodos y Análisis de Datos UCU-DCSP

Jake Bowers
jwbowers@illinois.edu
Online: http://jakebowers.org/

Invierno 2020

Overview

This class is an introduction to statistical adjustment using matched stratification in the style pioneered by Rubin and Rosenbaum and currently in rapid development across the social science and statistical disciplines. An important motivation for matching is to approximate an experimental design. And, since such a motivation arises from a desire to make transparent and defensible statements about causal relations, we will introduce the counterfactual conception of causal inference and the potential outcome formalization of these ideas. We will also spend some time on statistical inference (hypothesis testing, confidence interval creation) after the creation of a matched design. Finally, we will grapple with some of the questions that are current research topics in this area: When and how one can claim to have adjusted "enough"? How can we engage with concerns about unobserved confounds even if we have adjusted for what we observe?

Since methods of matching are rapidly developing in the methodology literature, we will here focus on the simplest and oldest form: post-stratification. The general concepts and work-flow should be transportable to more sophisticated methods of matched adjustment.

Goals and Expectations

This course aims to help you think about statistical adjustment using stratification and matching as compared to statistical adjustment using the linear model directly (adjustment by "residualization").

The course ought to give you opportunities to practice producing matched designs for your data and to ask questions that puzzle you as you do this work.

The point of the course is to position you to do the future learning that is at the core of your work as an academic analyzing data.

This course does not delve deeply into the theories of causal inference, statistical inference, or algorithms at the heart of these methods of statistical adjustment. Rather, through practice using tools, I hope that your curiosity is awakened and you begin to read more broadly and understand more deeply on your own.

Expectations I assume some previous engagement with high school mathematics, probability and statistical computing in the R statistical computing environment. If you have not used R, you are welcome to take the class, but I encourage you to get a little experience with R before the first class session. Feel free to email me to ask for advice about how to practice with R before the class begins.

Participation We will be doing hands-on work. I plan to lecture very little and instead will hope to pose problems of statistical theory, research design, and data for you to solve at your computers. I anticipate that you'll work in small groups, asking me and/or the group questions as you proceed. I will break away to draw on the board or demonstrate on my own computer now and then to clarify points or help you around particularly difficult tasks.

Computing We will be using R in class so those of you with laptops available should bring them. Of course, I will not tolerate the use of computers for anything other than class related work during active class time. Please install R (`http://www.r-project.org`) on your computers before the first class session. You may prefer to use R in the context of the Rstudio IDE (`http://www.rstudio.com/`).

Computing is an essential part of modern statistical data analysis — both for turning data into information and for conveying that information persuasively (and thus transparently and reliably) to the scholarly community. In this course we will pay attention to computing, with special emphasis on understanding what is going on behind the scenes. You will be writing your own routines for a few simple and common procedures.

Books

We will use the Rosenbaum book as our primary source. The other books are useful for further study.

Required: Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer (pdf free to download from some university ip addresses or via university library springerlink subscriptions: `http://www.springerlink.com/content/978-1-4419-1212-1/contents/`)

Recommended: Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press (particularly chapters 9,10 and 23 see `http://www.stat.columbia.edu/~gelman/arm/`).

(See `http://www.wjh.harvard.edu/~cwinship/cfa.html` for some links and background reading)

(see `http://www-stat.wharton.upenn.edu/~rosenbap/index.html` for lots of papers and presentations).

Schedule

Note: This schedule is preliminary and subject to change. We will spend roughly 4 hours together for three of the days and 3 hours on one of the days. I anticipate mixing group discussions of your questions from the readings with in-class work using your own laptops or those provided by the school to get practice creating, analyzing, and assessing matched designs.

1— 1—July 14—Experiments, Potential Outcomes, and Treatment Effects

Questions and Reading: What is the point of experiments? What are the key characteristics of experiments? Why are experiments special?

?

?, Chap 1

How can we bolster the interpretability of our comparisons if we do not have an experiment?

Rosenbaum, 2010, Chap 1

What do we mean by "causal inference"? Can we convince ourselves that experiments have special advantages like "unbiased estimation of averages of potential outcomes"?

?, Chap 2

Extra Reading: Gelman and Hill, 2007, Chap 9.0 – 9.3 (On potential outcomes and causal inference)

?, Chap 2

? (on the Counterfactual/Manipulationist conception of causality)

? (for a survey of other major conceptions of causality from the perspective of applied social science)

2— 2—July 15—Adjustment by Simple and Complex, Algorithmic Stratification

Given the problems that arise from the use of the linear model, how can we use research design to approximate the randomized experiment? How would we assess whether a matched design adjusted enough or not? What does it mean to adjust enough? What actual steps do we take in order to flexibly express our ideas about what it means for "like to be compared with like."

Questions and Reading: Why not use the linear model for adjustment? How do we know when we have adjusted enough to make a strong case for clear, interpretable comparisons?

Rosenbaum, 2010, Chap 6

Gelman and Hill, 2007, Chap 9.5–9.6

What is the basic intuition behind post-stratified designs and modern optimal versions of post-stratification that we call "matched designs"? What is "balance"? How do we use R to create and evaluate matched designs?

Rosenbaum, 2010, Chap 7–9, 13 (Especially Chap 7, Chap 8.6, Chap 9)

Extra: Rosenbaum, 2010, Chap 3

?

? for an example walk-through of a matched analysis including a discussion of missing data on covariates.

? [esp. their discussion of model sensitivity, for example their Fig 2]

Three different ideas about balance testing: (1) **?**; (2) **?**[1]; (3) Hansen and Bowers (2008) **?** or for a less mathematical version of the same argument (**?**, §3).

[1] `http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf`

3— 3—July 16—Statistical Inference for Matched/Post-stratified Designs

Questions and Reading: Given a matched design, how can we produce tests of substantively meaningful hypotheses about the unobserved comparisons of potential outcomes that we call "causal effects?"

Rosenbaum, 2010, Chap 2

How can we produce confidence intervals for an estimate of an average treatment effect?

?, Chap 6.1 and Appendix 6.1

?, Chap 17

Extra: **?**, Chap 4 on general requirements for statistical inference (i.e. what does it mean to do statistical inference at all, what are we inferring to?)

? provides some useful proofs supporting the idea that linear regression with "robust" HC2 standard errors provides a useful large-sample way to do statistical inference about average treatment effects.

? teach us about statistical inference for average treatment effects when matching after experimental outcomes have been collected.

???? Suggesting that even the large sample statistical inference from using linear regression in randomized experiments is biased. Also arguing that the Huber-White standard errors are not a good idea.

? and **?** showing how covariance adjustment is compatible with Fisher's randomization inference (and thus can be unproblematic after matching).

?? Suggesting that in large samples these biases worried about by Freedman ought not to worry us.

?, Chap 6–8 Suggesting, similarly to Green and Schochet, that regression is fine for statistical inference in experiments (and further suggesting the use of the Huber-White robust standard errors).

? suggesting that the bootstrap is not a good approach with matched designs.

For advanced reading on the latest in statistical theory for statistical inference for "matching estimators" (which include but are not restricted to post-stratified studies) see:

? for theory using randomization-inference.

? for a large-sample, Normal theory approach.

4— 4—July 17— Sensitivity Analysis: An observational study is not a randomized experiment.

Questions and Reading: We addressed concerns about variables that we do observe using matching. Although randomization addresses concerns about all background covariates (observed and unobserved), any non-randomized study may be criticized on the basis that it does not adjust for variables that were not observed.

Since an observational study (no matter how well matched) cannot adjust for unobserved confounders, how can we address concerns about such unobserved variables?

?

?

Extra: **?**, Chap 4

Rosenbaum, 2010, Chap 3, 14

?

5— *—Other Topics

If we move quickly, or if the class has some organized preferences, we could change the syllabus to discussion some of these ideas.

Non-bipartite matching – Advances in Multivariate Matching: Beyond Binary Treatment Rosenbaum, 2010, Chap 11; Lu et al. (2011); **?**

Longitudinal Matching – Advances in Multivariate Matching: Matching with Longitudinal Data Rosenbaum, 2010, Chap 12

6— Stuff that was painfully left out but which is important

Here are just a few extra citations to launch self-study of aspects of matching which we did not cover in our class.

The class elected to focus on matching for longitudinal problems for the last class. We thus are unable to cover other approaches to matching that have been developed by political methodologists such as Genetic Matching (**??**) or Coarsened Exact Matching (**??**) or Balance Optimization Subset Selection (**?**) or more fine-tuned versions of the optimal post-stratification that we consider in this class (for example, (**?**)). Nor did we have time to engage with many other applied and theoretical topics in causal inference for observational studies such as the work establishing causal interpretation of the propensity score (cited in the Rosenbaum textbook), or the alternative approaches to causal inference based on weighting by functions of the propensity score such as those arising from work by Jamie Robins (**?**), let alone alternative conceptualizations of causal relations such at those developed by Judea Pearl (**?**) or the work on estimation by Heckman or bounding causal inferences by Manski.

7— References

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Hansen, B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618.

Hansen, B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23:219.

Lu, B., Greevy, R., Xu, X., and Beck, C. (2011). Optimal nonbipartite matching and its statistical applications. *The American Statistician*, 65(1):21–30.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer.