

SDS 384: Causal Inference Methodology
Homework 2

February 18, 2020

Professor Zigler

Spencer Woody

Work with...

My own code is available at github.com/spencerwoody/sds384causal

Exercise 1

Separately for 2002 and 2014, conduct an unadjusted “crude” analysis comparing the average NOx levels for treated and untreated units. Evaluate whether the observed covariates are balanced in this unadjusted analysis.

Exercise 2

In this exercise you will use a variety of propensity score methods to estimate the causal effect of having an SnCR in a given year on NOx emissions in that year, under the assumption that the covariates listed in Table 1 are sufficient to adjust for confounding (i.e., that having an SnCR installed is conditionally unconfounded with respect to NOx emissions). For all parts of this exercise:

- *Use logistic regression with all of the variables in Table 1 (besides Tx and Outcome) included as covariates to estimate the propensity score.*
 - *Be sure to check covariate balance for each analysis*
 - *Conduct each analysis separately for 2002 and 2014, and comment (in ~3 sentences) on the differences between the analyses in the two years Be sure to check covariate balance for each analysis*
 - *I strongly suggest you read up on the following R packages to conduct these analyses: MatchIt, survey, ipw, twang*
- (a) *When you arrive at a propensity score model, plot the histograms of the estimated propensity scores in treated and untreated units.*
- (b) *Conduct a 1-1 nearest neighbor propensity score matching procedure without replacement*
- (c) *Conduct a 1-1 nearest neighbor propensity score matching procedure without replacement and a caliper set to 0.1 standard deviations of the estimated propensity score distribution.*
- (d) *Conduct an analysis that subclassifies units based on the estimated propensity score*
- (e) *Conduct an IPW analysis using weights $\frac{W_i}{\hat{e}(X_i)} + \frac{1-W_i}{1-\hat{e}(X_i)}$ and be sure to include a visual summary (e.g., histogram) of the estimated weights.*
- (f) *Conduct an IPW analysis using stabilized weights and be sure to include a visual summary (e.g., histogram) of the estimated weights.*

Exercise 3

Describe in ~5 sentences why the answers you obtained with the different propensity score methods in Exercise (1) were different from one another.

Exercise 4

Repeat Exercise (1e), but use a more advanced prediction model (your choice) to estimate the propensity score. Describe (~3 sentences) any differences.

References

- Jennifer L Peel, Mitchell Klein, W Dana Flanders, James A Mulholland, Paige E Tolbert, HEI Health Review Committee, et al. *Impact of improved air quality during the 1996 Summer Olympic Games in Atlanta on multiple cardiovascular and respiratory outcomes*. Health Effects Institute Boston, MA, 2010.
- Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 04 1983. ISSN 0006-3444. doi: 10.1093/biomet/70.1.41. URL <https://doi.org/10.1093/biomet/70.1.41>.
- Donald B. Rubin. Bayesian inference for causal effects: The role of randomization. *Ann. Statist.*, 6(1): 34–58, 01 1978. doi: 10.1214/aos/1176344064. URL <https://doi.org/10.1214/aos/1176344064>.