PSAboot: An R Package for Bootstrapping Propensity Score Analysis

Jason M. Bryer[1,*]

1. Excelsior College
*Contact author: jason@bryer.org

Propensity score analysis [5] is an approach to estimate causal effects in observational studies where randomization is not used. However, even when balance has been achieved between treatment and control units on observed covariates, there is a risk of bias due to unobserved covariates. Rosenbaum[4] has suggested testing one hypothesis multiple times using multiple matching and/or stratification methods to avoid exaggerated bias due to method selection. The **PSAboot** extends Rosenbaum's suggestion by implementing bootstrapping [1] for PSA. Like typical bootstrapping methods, the **PSAboot** package will draw *n* random samples and provide a pooled effect size estimate and confidence intervals. However, for each bootstrap sample multiple PSA methods will be used. The package includes implementations for matching (using the **Matching**[6] and **MatchIt**[2] packages), stratification with logistic regression, and stratification using classification trees (using the **rpart**[7] and **party**[3] packages). The package emphasizes the use of visualizations for evaluating balance as well as summarizing results. Additionally, examples on extending the package for other PSA methods will be discussed.

References

[1] Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

[2] Ho, D. E., K. Imai, G. King, and E. A. Stuart (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software 42*(8), 1–28.

[3] Hothorn, T., K. Hornik, and A. Zeileis (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics 15*(3), 651–674.

[4] Rosenbaum, P. R. (2012). Testing one hypothesis twice in observational studies. *Biometrika 99*, 763–774.

[5] Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

[6] Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software 42*(7), 1–52.

[7] Therneau, T., B. Atkinson, and B. Ripley (2013). *rpart: Recursive Partitioning*. R package version 4.1-3.