Heterogeneous Treatment Effects Workshop
SREE 2019

Select References

Please note: This list of references is by no means comprehensive as a literature review of these areas of methodology. We selected papers that we found to be generally clear and provide the main substance of what we discussed in the workshop. These papers' reference lists are also generally comprehensive and quite useful.

General Overview

Schochet, P. Z., Puma, M., & Deke, J. (2014). Understanding Variation in Treatment Effects in Education Impact Evaluations: An Overview of Quantitative Methods. NCEE 2014-4017. *National Center for Education Evaluation and Regional Assistance*.

- Overview of heterogeneity, with a focus on the issues that arise in the domain of education.

Weiss, M., H. S. Bloom, and T. Brock (2014). A Conceptual Framework for Studying the Sources of Variation in Program Effects. MDRC.

- Nice discussion of where one might find heterogeneity in a RCT.

Moderation analysis and thinking about variation

VanderWeele, T. J., & Knol, M. J. (2011). Interpretation of Subgroup Analyses in Randomized Trials: Heterogeneity Versus Secondary Interventions. *Annals of Internal Medicine*, *154*(10), 680–683. http://doi.org/10.7326/0003-4819-154-10-201105170-00008

- This paper discusses how to think about subgroup effects in moderation analysis.

Knol, M. J., & VanderWeele, T. J. (2012). Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology*, *41*(2), 514–520. http://doi.org/10.1093/ije/dyr218

- Similar to the above, but looks at relative risk and other aspects of binary outcomes.

Ding, P., Feller, A., & Miratrix, L. (2018). Decomposing treatment effect variation. *Journal of the American Statistical Association*, 1–14.

- Proves that linear regression with interaction "works" in that we get good summaries of the systematic variation (variation explained by covariates) in an RCT.

Ding, P., Feller, A., & Miratrix, L. (2015). Randomization Inference for Treatment Effect Variation, 1–18. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/rssb.12124/full

- This talks about detecting idiosyncratic variation, and has the Head Start example in it.

Feller, A., & Gelman, A. (2015). Hierarchical models for causal effects. *Emerging Trends in the Social and Behavioral Sciences: An interdisciplinary, searchable, and linkable resource.*

Multisite Trials

Raudenbush, S. W., & Bloom, H. S. (2015). Learning About and From a Distribution of Program Impacts Using Multisite Trials, *36*(4), 475–499. http://doi.org/10.1177/1098214015600515

- Overview of important themes in this domain.

Bloom, H. S., Raudenbush, S. W., Weiss, M. J., & Porter, K. (2017). Using multisite experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient. Journal of Research on Educational Effectiveness, 10(4), 817–842.

- This discusses the FIRC model

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How Much Do the Effects of Education and Training Programs Vary Across Sites? Evidence From Past Multisite Randomized Trials. *Journal of Research on Educational Effectiveness*, 1–35. http://doi.org/10.1080/19345747.2017.1300719

- This paper is one of a set of papers published in JREE dealing with cross-site variation. (The other papers, listed above and below, are also quite excellent.) Also see appendices for more on testing for cross-site variation with the Q-statistic.

Bloom, H. S., & Spybrook, J. (2017). Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects, *10*(4), 877–902. http://doi.org/10.1080/19345747.2016.1271069

- This paper focuses on power issues in the context of cross-site heterogeneity (both for ATE and for detecting the cross-site heterogeneity).

Machine Learning

Hill, J. L. (2010). Bayesian Nonparametric Modeling for Causal Inference. Journal of Computational and Graphical Statistics 20(1), 217–240.

- One of the current easy to use machine learning tools being adapted for treatment variation is BART. This showcases BART for this area.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, *116*(10), 4156-4165.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353-7360.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228-1242.

Nie, X., & Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. *arXiv:1712.04912*.

Hahn, P. R., Murray, J. S., & Carvalho, C. (2017). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. *arXiv:1706.09523*.

Classic References and Foundational Papers

Bryk, A. S. and S. W. Raudenbush (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin* 104(3), 396.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008). Nonparametric tests for treatment effect heterogeneity. *The Review of Economics and Statistics* 90(3), 389–405.

- This talks about using more flexible modeling by extending the linear model. It is a precursor to our idiosyncratic testing work.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64(4), 487–535.