Regression tree analysis implementation - ADIA Studies 1&3

This document outlines the code implementing the regression tree analysis for ADIA studies 1 and 3.

'0.preprocess.r': Create separate training and estimation sample

- Use stratified random sample (stratified by demographics characteristics and any ACE)

'1.rtree.r': Implement regression tree using training sample

- Incorporate sampling weights
- Use all cases, even if missing covariate info (surrogate)
- 3 variants
  1. Classical (Breiman, 1998)

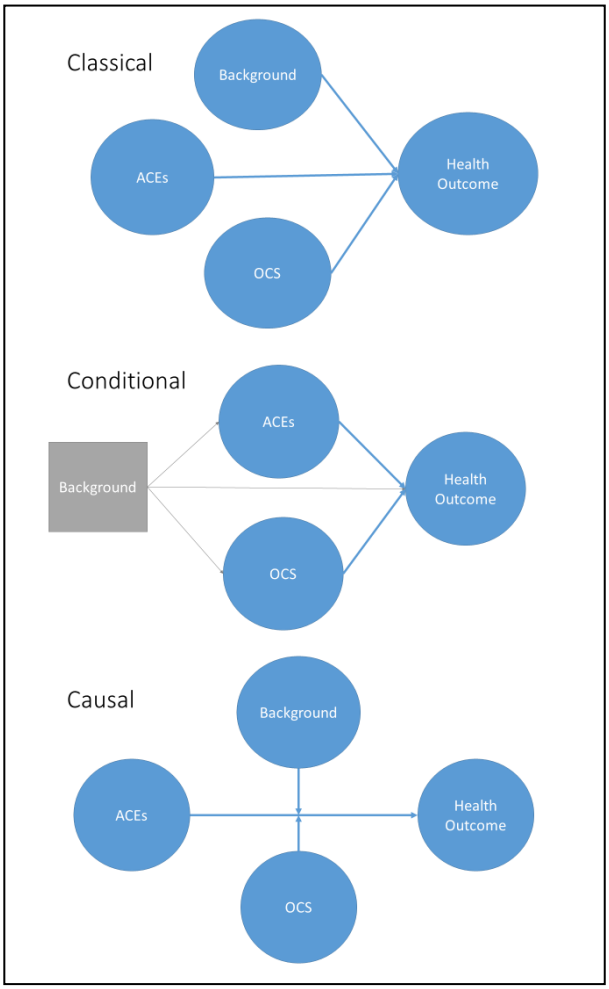
     All predictors (either exposure or background characteristics) are treated symmetrically

  2. Conditional (Stanfill et al., 2019)

     The importance of exposure to ACEs and OCS is examined after adjusting for background characteristics. Can only handle a limited number of strata.

  3. Causal (Athey & Imbens, 2015)

     Focuses on how OCS and background characteristics modify the causal relationship between conventional ACEs and outcomes. Start by estimating casual effect of conventional ACEs adjusting for observed confounding.



'2.inference.r': Obtain inference using the estimation sample

- Use svyglm to incorporate sampling design (Lumley, 2004)

References

Athey, S., & Imbens, G. (2015). *Recursive Partitioning for Heterogeneous Causal Effects*. https://doi.org/10.48550/ARXIV.1504.01132

Breiman, L. (Ed.). (1998). *Classification and regression trees* (Repr). Chapman & Hall [u.a.].

Lumley, T. (2004). Analysis of Complex Survey Samples. *Journal of Statistical Software*, *9*(8). https://doi.org/10.18637/jss.v009.i08

Stanfill, B., Reehl, S., Bramer, L., Nakayasu, E. S., Rich, S. S., Metz, T. O., Rewers, M., Webb-Robertson, B.-J., & TEDDY Study Group. (2019). Extending Classification Algorithms to Case-Control Studies. *Biomedical Engineering and Computational Biology*, *10*, 117959721985895. https://doi.org/10.1177/1179597219858954