**Introduction**

There has been a common issue of creating randomized controlled trials due to the nature of said experiments [1]. This has led to the development of many different causal learning methods in order to determine causes and effects from previous conducted randomized controlled trials. The Intelligent Randomized Controlled Trials (iRCT) method is a supervised learning method that was developed in order to apply new methods of determining cause and effect to randomized controlled trials. iRCT is a python package that is partly based on the findings of matching estimators by Abadie [2], the use of propensity scoring using logistical regression, and using those propensity scores in order to do matching to determine an average treatment effect [3].

**Propensity Scoring**

Propensity scoring is defined as “a conditional probability of being exposed given a set of covariates” [4]. For iRCT the use of propensity scoring is based around the idea that for most datasets there will likely be more than a single covariate outside of the treatment and outcomes, so using a propensity scoring method would allow all covariates to become a single value that could then be used to identify matches.

In iRCT the current propensity scoring method is based around the sklearn package and its linear regression function and pipelining feature, as well as pandas for general database manipulation. The propensity score is also eventually converted into another value known as the “propensityScoreLogit”, which is a function of the propensity score and is what is usually used to match upon and also what iRCT uses for its matching function [5].

**Previous Versions and Developments**

The initial iRCT method was developed was based upon Abadie’s work [2] and was developed with two key assumptions in mind. The first being that the dataset this would be applied to would be relatively small. The second being that there would be only 4 distinct columns, an index column, a covariate column, a treatment column, and an outcome column. Once this original method was tested on simple datasets with these assumptions in place it was clear that a method would need to be developed in order to allow for multiple covariate columns. This initial method can still be found in the iRCT package and is named “firstAttempt\_calculateRelationVal”.

This allowed for the introduction of propensity scoring into the iRCT methodology. Due to propensity scorings ability to combine multiple covariates into a single covariate and the research of using propensity scoring for determining causal effects [6]. This new method would be applied to the main dataset used for all comparisons and tests from here on, COVID3\_4Nodes3. This method was quickly deemed necessary to be adjusted due to the extreme length of calculating just one average treatment effect between a single pair of treatment and outcome. This second method can still be found in the iRCT package and is named “secondAttempt\_calculateRelationVal”.

Finally, in order to deal with the extremely slow process time from the second attempt, it was determined that the problem was the way the matching was done due to needing to iterate over every single row. However, a simple fix was implemented by using the apply function from the pandas package. Everything was already being stored in a pandas data frame for ease of data manipulation and thus, the implementation of the new matching method allowed for a significant speed up in processing time. This is still the current method iRCT is using and can be found in the iRCT package under the name “calculateRelationVal”.

**Comparison against other causal methods**

**Sources**

1. [**https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5539637/**](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5539637/)
2. [**https://journals.sagepub.com/doi/10.1177/1536867X0400400307**](https://journals.sagepub.com/doi/10.1177/1536867X0400400307)
3. [**https://florianwilhelm.info/2017/04/causal\_inference\_propensity\_score/**](https://florianwilhelm.info/2017/04/causal_inference_propensity_score/)
4. [**https://www.publichealth.columbia.edu/research/population-health-methods/propensity-score-analysis**](https://www.publichealth.columbia.edu/research/population-health-methods/propensity-score-analysis)
5. [**https://www.youtube.com/watch?v=gaUgW7NWai8&t=981s&ab\_channel=PyData**](https://www.youtube.com/watch?v=gaUgW7NWai8&t=981s&ab_channel=PyData)
6. **Rosenbaum and Rubin (1983).**