



Institutional Analytics – Research & Analytics

A Rapid Approach to Program Assessment Through Matching Algorithms

March 9, 2023

Outline

1. Why Perform Matching?
2. Overview of How to Perform Matching
3. Demo of Distance-Based Matching Program Assessment Tool



SECTION 1

Why Perform Matching?

Motivation

- Our team (Gina in particular) is consistently asked to perform program assessments
- Work mostly with observational datasets
 - No randomly assigned control group
 - Difficult to ensure that the treatment group “looks” like the group that we’re comparing results against
- Most of those assessments require some type of matching



What is Matching?

- Method that attempts to create two sets of similar “looking” observations, usually to compare results
- Helps to ensure that units receiving an intervention have the same background and experiences and that the impact of the intervention isn’t conflated by other variables
- Can be thought of as a data pre-processing step that creates a pseudo-experimental dataset
- Matched data make statistical analyses more robust and descriptive/visual analyses more trustworthy



SECTION 2

Overview of How to Perform Matching

Steps in Matching Process

1. Identify and Gather Matching Variables
2. Compute Scores or Distances between Potential Treatment/Control Pairs
3. Match Treatment Units to Control Units Based on Distance/Score
4. Evaluate Matching Balance
5. Analysis After Matching



1. Identify and Gather Matching Variables

1. Choose variables that relate to being selected into the treatment and that impact the measured outcome.
2. Be intentional. Don't throw everything into the matching.
3. What not to include
 - Actual outcome variables
 - Variables measured after the intervention



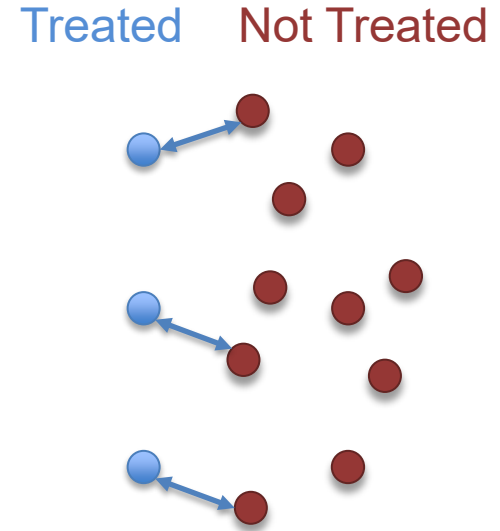
2. Compute Scores or Distances

1. Scores (e.g., propensity scores)

- Likelihood of being selected into the treatment group calculated from the matching variables
- Single balancing score modeled from the covariates using logistic regression, tree-based models, etc..
- Minimize the distance between the propensity scores

2. Distances (e.g., Euclidean distance)

- Distance between each matching variable is calculated
- Minimize the distance between the values of the covariates, for example by using the k-nearest neighbor algorithm.



2. Compute Scores or Distances (*continued*)

Modified Gower Distance Measure

1) Compute the distance between matching variables for each potential pair

- *Rescale variables between 0 and 1*
- *Compute the absolute value of the difference between the paired individuals*

2) Compute the overall distance between the paired individuals

- *Define weights for the variables based on perceived/modeled importance in the selection criteria*
- *Compute the weighted average of the distances across all variables*



3. Match Treatment and Control Units

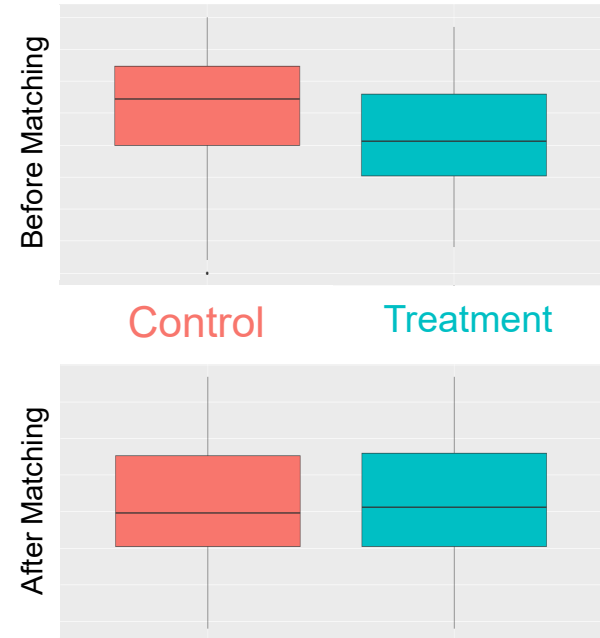
- **Type of matching**
 - **Nearest Neighbor (Greedy)**
 - Sequentially matches each treatment unit to the closest control unit without replacement
 - **Pros:** Quick and intuitive
 - **Cons:** Operates sequentially, so not all units will have an optimal match
 - **Optimal**
 - Considers every possible combinations of treatment/control pairs and chooses the set of matches that minimizes a global distance
 - **Pros:** Guaranteed to find optimal match
 - **Cons:** Computationally intensive
 - **Many others**

When in doubt try optimal unless it's too computationally intensive.



4. Evaluate Matching Balance

- After matching, ensure that treatment and control populations “look” the same on average
 - Visual Inspections
 - Bar charts for categorical variables
 - Histograms or box plots for numeric variables
 - Perform basic statistical testing
 - Chi-square tests for categorical variables
 - T-test or comparable non-parametric test (e.g., Wilcoxon Rank-Sum) for numeric variables



5. Analysis After Matching

1. Statistical Analysis to Test for Significant Treatment Effect
2. Tableau report to explore outcomes for Treatment versus Control
3. Give matched dataset to clients to explore the impact of their program



SECTION 5

Demo of Distance-Based Program Assessment Tool

Implementing Matching With Software

- Statistical software have packages and functions that make it easy to implement matching
- For the Statistical Software R...
 - MatchIt package and function:
<https://www.rdocumentation.org/packages/MatchIt/versions/4.5.0>
 - Allows specification of distance matrices, matching algorithm (nearest/optimal), ratio of controls to treatments, etc.



Thank you! Questions?

Contact Us:

Research and Analytics – iara@iu.edu

[Link to Matching Application](https://github.com/iu-ia-research-analytics/distance-based-matching-program-assessment-tool)

<https://github.com/iu-ia-research-analytics/distance-based-matching-program-assessment-tool>

