Matching Technical Appendix

This document provides a technical overview of how the RCE Coach matching tool functions. For a brief introduction to matching as a research design, you can refer to our Matching Overview.

Matching attempts to group students (or the individuals you are trying to affect) with similar characteristics. The ideal match would be two students identical on every observed and unobserved characteristic. In practice, it is very unlikely that two students will be identical on all observed characteristics, so you must determine which observable characteristics you will match on based on how likely those characteristics are to affect the outcome of interest.

Matching on only one variable is conceptually simple. Suppose that you have a set of students using U-Read, an educational technology, and another set not using it. You are trying to create a matched data set using only a pre-test score. A simple algorithm would look like this:

1. Define a distance measure. For example, the absolute distance between pre-test scores of 240 and 265 is: |240-265|= 25
2. Take a student using U-Read and calculate the distance between his pre-test and the pre-test of each of the students not using U-Read.
3. Match this student to the student with the minimum distance.
4. Repeat Steps 2 through 4 for all students using U-Read.

It is more complicated to define distance when you are trying to match students using more than one characteristic. When you use multiple matching variables, you have to summarize this information into a single distance number. For example, suppose that you want to include an indicator for English as a second language and another for economic disadvantage. In this case, we can create a measure that summarizes all this information and match based on the summary. We can combine all of this information, even though the scales differ, by calculating the probability that a student with a given set of observable characteristics uses the educational technology. The goal is that after matching, students in the intervention group will be matched to students from the comparison group who look the same on observable key characteristics.

HOW DOES THE TOOL WORK?

The RCE Coach uses R, a free software package for statistical computing. The Coach will first perform several checks on the data to confirm that there are no data issues that would cause an error with the matching code. Then the Coach will create a matched data set using the R package MatchIt. The rest of this document describes the specific data checks and randomization algorithm.

Step 1: Check for data issues

The Coach will perform the following checks to verify that the Coach has the necessary inputs to randomly assign the students:

1. Data is NULL, not readable, or has 0 observations.
2. No treatment variable specified.
3. No matching variables specified.
4. One or more matching variables not found in the data.
5. Treatment variable has values other than 0, 1, and NA.
6. One or more matching variables are not numeric.

If any test fails, the Coach will print a message to help the user identify the problem. If the Coach detects no problems, it will proceed to Step 2.

Step 2: Create a matched data set

To create a matched data set, the RCE Coach uses one-to-one nearest neighbor matching (Rubin 1973). The Coach will select a comparison individual with the smallest distance for each individual using the educational technology.

After the Coach has selected matches for each individual using the educational technology, it will assess whether the two groups—the intervention group and the matched comparison group—are similar on these key characteristics. The Coach uses the baseline equivalence standard from the What Works Clearinghouse to make this assessment. We consider the matched group to be a valid comparison group if the baseline differences on all matching variables are less than 0.25 standard deviations.

If the Coach fails to generate a valid comparison group, it will define a maximum distance, known as a caliper, and repeat the matching algorithm. If the distance between the individual using the technology and the nearest individual not using it is larger than this caliper, the Coach will drop this observation. If the Coach cannot generate a valid comparison group with this caliper, it will use a smaller one. The Coach will start by using no caliper, and then move from 1.00 standard deviation of the distance measure to 0.25 standard deviations in 0.25 intervals. Using a caliper can mean that some users of the educational technology will not receive a match. If this is the case, you should be cautious when interpreting the results because you are measuring the effect only for the individuals who have a good match. This algorithm can be summarized as follows:

1. The Coach attempts to match treatment and comparison observations using nearest neighbor matching.

2. If matching in Step 1 successfully produced balanced samples, the matched groups are returned as a downloadable file and relevant sample sizes and graphic displays of balance are shown on the screen.
3. If matching in Step 1 was not successful, the Coach attempts to match using caliper matching, with an initial caliper value of 1.00 and decreasing by 0.25 until either matching produces balanced samples or the caliper reaches 0.
   a. If matching was successful, a downloadable file is available as in Step 2 and the same sample sizes and graphics are shown on screen.
   b. If matching was not successful for any caliper value, balance graphics are shown but there is no file to download.

The code for the RCE Coach is open source under the General Public License Version 3 and will be available soon on our github repository.

Which variables should I include for matching?

You should always include some baseline, or pre-intervention, measurement that is related to your outcome of interest. In the case of U-Read, you would include pre-test scores as one of your matching variables. However, there are cases in which you might want to use additional characteristics for matching to improve precision.

In our example, if the number of students using and not using U-Read is very different (one group is substantially larger than the other), the literature recommends including:

- Variables that affect the likelihood of using U-Read
- Variables that affect the outcome you want to study (for example, achievement), controlling for use of the educational technology
  (Rubin and Thomas 1996; Heckman et al. 1998; Ho et al. 2007)

If you do not have substantially more students not using U-Read than the number of students using U-Read, you **should not** include all available characteristics (Ho et. al. 2007). The intuition is that as you increase the number of characteristics used for matching it might become too difficult to find a good match. For example, think about searching for a new apartment: if you include too many characteristics (location, bedrooms, laundry, pets, size, and so on), your search could end up too narrow and you won’t find any results.

Should matching variables also be included as covariates in the analysis?

YES! Matching is seldom perfect—some imbalance often remains on at least some variables. Regression adjustment for the matching variables can mitigate any remaining imbalance.

REFERENCES

Heckman, J., H. Ichimura, J. Smith, and P. Todd. "Characterizing Selection Bias Using Experimental Data." No. w6699. Cambridge, MA: National Bureau of Economic Research, 1998.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference." Journal of Statistical Software, [vol. 42, no. 8], 2007, [add page span]. Available at http://gking.harvard.edu/matchit.

Rubin, D.B., and N. Thomas. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." Biometrics, 1996, pp. 249–264. 7