Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 19, Number 18, November 2014 ISSN 1531-7714

A Step-by-Step Guide to Propensity Score Matching in R

Justus J. Randolph, Kristina Falbe, Austin Kureethara Manuel, Joseph L. Balloun
Mercer University

Propensity score matching is a statistical technique in which a treatment case is matched with one or more control cases based on each case's propensity score. This matching can help strengthen causal arguments in quasi-experimental and observational studies by reducing selection bias. In this article we concentrate on how to conduct propensity score matching using an example from the field of education. Our goal is to provide information that will bring propensity score matching within the reach of research and evaluation practitioners.

Propensity score matching is a statistical technique in which a treatment case is matched with one or more control cases based on each case's propensity score. This matching can help strengthen causal arguments in quasi-experimental and observational studies by reducing selection bias. Because there have been many thorough explanations and rationales for propensity score matching published elsewhere (Adelson, 2013; Holland, 1986; Rubin, 2005; Rudner & Peyton, 2006; Shadish, Cook, & Campbell, 2002; Stone & Tang, 2013), in this article we will concentrate on how to conduct propensity score matching using an example from the field of education. Specifically, in this document we provide a step-by-step example of conducting propensity score matching in **R** using the MatchIt package with nearest-neighbor 1-to-1 matching. While there is other software than **R** for conducting propensity score matching, we have chosen **R** because it is open-source software and is widely used by data scientists across many different fields. Our goal in this article is to provide information that will bring propensity score matching within the reach of research and evaluation practitioners.

Information on the Dataset Used Here

Data from an observational study by Falbe (2014) are used here to illustrate propensity score matching. In that study, Falbe used publicly available school-level data from several states to investigate whether an intervention (i.e., being designated a Schools to Watch © (*stw*) school) was a predictor of success in reading and mathematics achievement, when controlling for school size (*tot*), percentage of minority students (*min*), and percentage of students receiving free and reduced lunch (*dis*). For our example here, we use Falbe's school-level data only from the state of New York. In the New York data set, there were 25 *stw* schools and 560 non-*stw* schools. As matching variables, Falbe chose school size, percentage of minority students, and percentage of students receiving free and reduced lunch. Her rationale for choosing those matching variables was that previous research had shown that they tended to covary with academic achievement. By matching on those variables, her goal was to reduce selection bias between "treated" (i.e., *stw*) and "control" (i.e., non-*stw*) schools. Note that although Falbe's study was correlational and not experimental, we use the

terms *treated* and *control* here because those are the terms reported in the output of the MatchIt package.

The Steps in Conducting Propensity Score Matching in R

Step 1. Install R.

R is a free statistical package that can be downloaded from the URL in the **R** Core Team (2014) reference in the References section of this article. Specific installation instructions are provided after downloading and opening the software. **R** is available for Windows, Mac OS X, and Linux operating systems. The directions presented in this article are based on **R** version 3.0.3.

Step 2. Install and load the MatchIt package.

MatchIt is an **R** package that easily enables **R** users to conduct propensity score matching; specific information on the MatchIt package can be found from Ho, Kosoke, King, and Stuart (2007a, 2007b, 2011, 2013). To use the MatchIt package, you must first install and load it. You only need to install MatchIt the first time you use it; however, you will need to load it each time you re-open **R** software.

To install a package, open **R** and select Packages menu from the top of the screen. Then choose Install Package(s). . . A pop-up window, CRAN Mirror, will be displayed. Choose the site you prefer from the list and click OK. Another window labeled "Packages" will be displayed. Scroll down the list in the window to select MatchIt and click OK. The package will get downloaded immediately.

Next, to load the MatchIt package, choose Load Package… from the dropdown menu under the Packages menu in R. A window "Select one" with the list of available packages will be displayed. Choose MatchIt from the list and hit OK. The MatchIt package should now be loaded. You should only have to install the package one time. However, you will need to load the MatchIt package in **R** every time you wish to run it.

Step 3. Prepare and load the data.

To perform propensity score matching, you will need a data set that consists of cases in rows and variables in columns. You will need a grouping variable and one or more matching variables. The grouping variable is the variable that specifies which group a case belongs to (e.g., treatment or control). The matching variables are the ones that you want to attempt to equalize the groups on. For example, in the Falbe (2014) data set, the stw variable is the grouping variable. It specifies whether a particular case (a school) has been designated as a Schools to Watch © school (1) or not (0). The other variables tot (school size), min (percentage of minority students in the school), and dis (percentage of students receiving free and reduced lunch) are the matching variables. The dataset we used here can be downloaded from Randolph (2014a). In your own datasets, make sure that there are no missing data or **R** may not be able to perform the analysis.

Although there are functions to import Excel, SPSS, or other data formats into R, we have found it is most convenient to save it as a .csv file before trying to load the data into R. When you save the file in Excel, you will have the option to save it as a .csv file. As you save the file, note its location.

Next, you will need to replace the file location between the parentheses in the first line of **R** code in Figure 1 with your own file location. Note that in the first line, forward slashes rather than backslashes are used to specify the file location. The example below is a file location in Windows, where the data are located in a file called `newyork.csv` in folder called `r` on the C drive. The first line of code reads the data from your computer and renames it `mydata`. The second line makes those data available in the current **R** session. The third line of the code below prints the variable names and the first ten cases in the data set. We do this just to check the data set and understand what each column represents.

```
mydata <- read.csv ("C:/r/newyork.csv")
attach(mydata)
mydata[1:10,]
```

Figure 1. Code example for inputting a data set.

Figure 2 shows the results of running the code in Figure 1. It shows the first ten cases in the data set and what variables are included in the columns. It shows that the columns from left to right are the case number (a unique id number for each school), school (the name of each school), tot (the total number of students in the school), min (the percentage of minority students in the school), dis (the number of students receiving free or

reduced lunch), and stw (whether a school is designated as (1) Schools to Watch © or (0) not).

	school	tot	min	dis	stw
1	SKANEATELES MIDDLE SCHOOL	380	0.03	0.00	0
2	MARCUS WHITMAN MIDDLE SCHOOL	276	0.04	0.00	0
3	BLIND BROOK-RYE MIDDLE SCHOOL	376	0.09	0.00	0
4	BRONXVILLE MIDDLE SCHOOL	404	0.11	0.00	0
5	BRIARCLIFF MIDDLE SCHOOL	374	0.12	0.00	0
6	RYE MIDDLE SCHOOL	754	0.17	0.00	0
7	EASTCHESTER MIDDLE SCHOOL	704	0.26	0.00	0
8	SCARSDALE MIDDLE SCHOOL	1172	0.27	0.00	0
9	EDGEMONT JUNIOR-SENIOR HIGH SCHOOL	920	0.42	0.00	0
10	SEVEN BRIDGES MIDDLE SCHOOL	619	0.17	0.01	0

Figure 2. Results of running the code in Figure 1. The first ten cases are displayed.

Step 4. Perform matching and evaluate the results.

The next step is to perform the matching and evaluate the results. The first line in the code shown in Figure 3 performs the matching where the grouping variable is stw and the variables being matched on are tot, min, and dis. You will need to replace these variable names in the code with the variable names of your own data set. The method command in Figure 3 specifies that the nearest neighbor method will be used. The ratio command indicates one-to-one matching—every treatment case will be matched with one control case. You can increase the number of control cases matched to each treatment case by increasing this number; usually this number is between 1 and 5.

```
m.out = matchit(stw ~ tot + min + dis,
      data = mydata, method = "nearest",
      ratio = 1)

summary(m.out)

plot(m.out, type = "jitter")

plot(m.out, type = "hist")
```

Figure 3. Code to perform propensity score matching and get results.

There are many matching methods that can be used; a short description of them can be found in the list below. We encourage MatchIt users to try out the different matching methods to see which method works best for a particular data set. In this case, we tried all of the matching methods currently available in MatchIt and chose the nearest neighbor method because it resulted in the lowest mean differences between groups. Some of the other methods call for the installation of additional packages.

- Exact Matching – This technique matches each treated unit with a control unit that has exactly the same values on each covariate. When there are many covariates and/or covariates that can take a large range of values, exact matching may not be possible (method = "exact").
- Subclassification – This technique breaks the data set into subclasses such that the distributions of the covariates are similar in each subclass (method = "subclass").
- Nearest Neighbor – This technique matches a treated unit to a control unit(s) that is closest in terms of a distance measure such as a logit (method = "nearest").
- Optimal Matching – This technique focuses on minimizing the average absolute distance across all matched pairs (method = "optimal"). This method of matching requires the optmatch package.
- Genetic Matching – This technique uses a computationally intensive genetic search algorithm to match treatment and control units (method = "genetic"). It requires the Matching package.
- Coarsened Exact Matching – Finally, this technique matches on a covariate while maintaining the balance of other covariates. It is claimed to work "well for multicategory treatments, determining blocks in experimental designs, and evaluating extreme counterfactuals" (Ho, Kosuke, King, & Stuart, 2011, p.12) (method = "cem").

See the documentation on MatchIt for more details on the matching methods mentioned above. (Ho, Kosuke, King, & Stewart, 2007a, 2007b, 2011, 2013).

The results of matching are saved in a variable called m.out. The second line of code in Figure 3 prints a summary of the matching results (see Figure 4). The third and fourth lines produce jitter plots and histograms.

The results show that matching worked very well for this data set. In the summary of balance for all data section of Figure 4, before matching the mean number of students (tot) in the treated schools (i.e., stw

```
Summary of balance for all data:
         Means Treated Means Control SD Control Mean Diff  eQQ Med eQQ Mean   eQQ Max
distance        0.0943        0.0405     0.0503    0.0537    0.0559    0.0599    0.1875
tot           832.6400      568.8998   333.6746  263.7402 300.0000 310.9600 1124.0000
min             0.1664        0.2767     0.3011   -0.1103    0.0200    0.1276    0.6300
dis             0.1840        0.4079     0.2500   -0.2239    0.2500    0.2276    0.4900

Summary of balance for matched data:
         Means Treated Means Control SD Control Mean Diff eQQ Med eQQ Mean  eQQ Max
distance        0.0943        0.0942     0.0513    0.0001  0.0004   0.0005   0.0024
tot           832.6400      830.6400   315.6859    2.0000 99.0000 115.8400 247.0000
min             0.1664        0.1772     0.1330   -0.0108  0.0200   0.0260   0.1500
dis             0.1840        0.1808     0.1361    0.0032  0.0100   0.0256   0.0900

Percent Balance Improvement:
         Mean Diff. eQQ Med eQQ Mean eQQ Max
distance    99.8180 99.3674  99.1609 98.7414
tot         99.2417 67.0000  62.7476 78.0249
min         90.2061  0.0000  79.6238 76.1905
dis         98.5707 96.0000  88.7522 81.6327

Sample sizes:
          Control Treated
All           559      25
Matched        25      25
Unmatched     534       0
Discarded       0       0
```

Figure 4. Results showing the effectiveness of the propensity scores matching.

schools) was 263.74 students less than in the control (i.e., non-stw schools) schools. The treated schools had 11% less minority students (min) and 22% percent less students of poverty (dis) than control schools. After matching, however, those differences reduced dramatically as shown in the summary of balance for matched data section of Figure 4. The mean difference in number of students between treated and control schools reduced to 2; it was 263 before matching. The percent difference in minority students between treated and control schools reduced to 1%; it was 11% before matching. Finally, the mean difference between treated and control schools in terms of percent of impoverished students reduced to 3/10ths of a percent; it was 22% before matching. In short, the treated and control schools after matching are very similar now in terms of number of students, percentage of minority students, and percentage of students receiving free and reduced lunch. Before matching, the treated schools were on average larger, had less minority students, and had less impoverished students than control schools. The rightmost columns in these summary data show the median, mean, and maximum quartile-differences between the treated and control data; smaller QQ values indicates better matching. Note that the QQ values are all smaller after matching than before matching. The third and fourth lines of the code in Figure 2 creates jitter plots and histograms to visualize the quality of the matching.

Figure 5 is a jitter plot where each circle represents a case's propensity score. The absence of cases in the uppermost stratification indicates that there were no unmatched treatment units. The middle stratifications show the close match between the treatment units and the matched control units. The final stratification shows the unmatched control units, which will not be

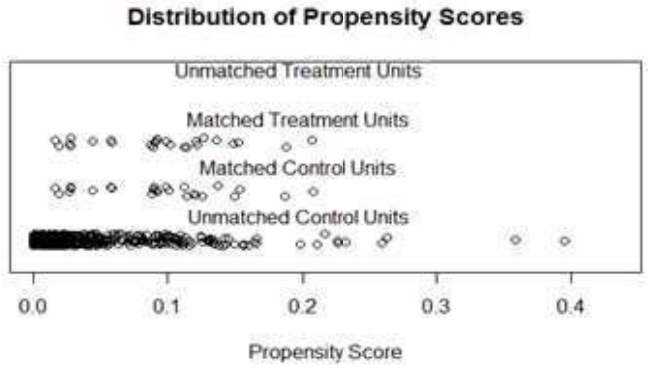

Figure 5. Distribution of propensity scores.

used in any follow-up analyses. Figure 6 shows the histograms before and after matching. The histograms before matching on the left differ to a great degree. The histograms after matching on the right are very similar however. In sum, both the numerical and visual data show that the matching was successful.

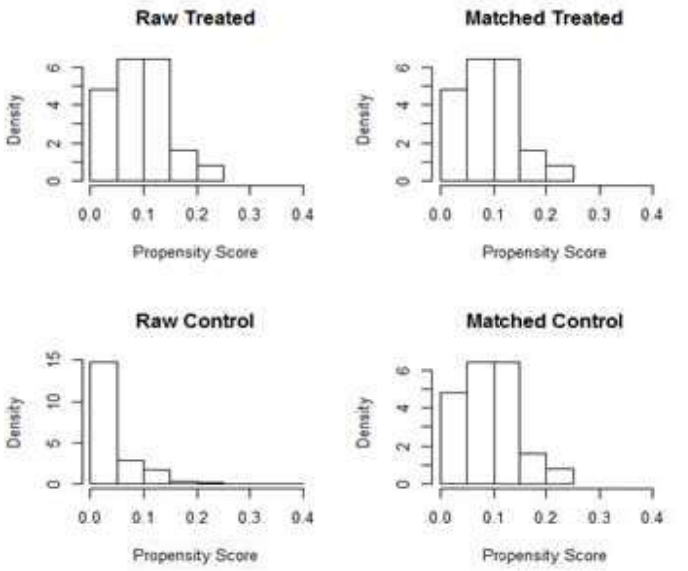

Figure 6. Histograms of propensity scores before and after matching.

Step 5. Export a data file to do follow-up analyses

Once the matching has been completed, you will want to create a data set that only has the matched cases to do follow up statistical analyses now that the data are matched. In Figure 7, the first line of code creates an **R** data set that only has the matched control and treatment cases (i.e., it deleted the 500+ control cases that were unmatched.) The second line of data converts the matched data set back into a .csv file that can either be further analyzed in **R** or exported into other statistical software; in this case, the output data set was saved as a file called `newyork_nearest1` in a folder called `r` on the C drive. (The matched data set for the New York data can be downloaded from Randolph (2014b)).

```
m.data1 <- match.data(m.out)
write.csv(m.data1, file =
     "C:/r/newyork_nearest100.csv")
```

Figure 7. Code to output a matched data set.

In terms of the follow-up analysis, Falbe (2014) was interested in whether stw schools performed better than non-stw schools in terms of mathematics achievement. To do that analysis, she added mathematics achievement scores to the matched data set and predicted mathematics scores from the following variables: whether a school was designated as an stw school or not, the school size, the percentage of minority students in a school, and the percentage of students who received free or reduced lunch. It turns out that matching was important in this case. Without matching, stw schools had statistically significantly better achievement than non-stw schools. However, with matching, Fable found no statistically significant difference between stw and non-stw schools in terms of academic achievement. Without propensity score matching, Falbe would have come to a different conclusion about the efficacy of the intervention, most likely as a result of selection bias.

It is clear that propensity score matching is a useful tool for reducing selection bias and strengthening causal conclusions. We hope that this step-by-step guide will enable a wide variety of researchers and evaluators to add propensity score matching to their repertoire of data analysis techniques.

References

Adelson, J. L. (2013). Educational research with real-world data: Reducing selection bias with propensity score analysis. Practical Assessment Research & Evaluation, 18(15). Retrieved from http://pareonline.net/getvn.asp?v=18&n=15

Falbe, K. (2014). The relationship between Schools to Watch © designation and academic achievement: A study of Colorado, New York, Ohio, and Virginia (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3581272)

Ho, D., Kosuke, I., King, G., & Stuart, E. (2007a). MatchIt: Nonparametric preprocessing for parametric causal inference. Political Analysis, 15(3), 199-236.

Ho, D., Kosuke, I., King, G., & Stuart, E. (2007b). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Journal of Statistical Software. Retrieved from http://gking.harvard.edu/matchit/

Ho, D., Kosuke, I. King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference [software documentation]. Retrieved from http://gking.harvard.edu/matchit

Ho, D., Kosuke, I., King, G., & Stuart, E. (2013). MatchIt: Nonparametric preprocessing for parametric causal inference [software]. Retrived from http://gking.harvard.edu/matchit

Holland, P. W. (1986). Statistics and causal inference. Journal of the American Statistical Association, 81(396), 945-960.

Randolph, J. J. (2014a). New York educational data set example before matching. Retrieved from http://justusrandolph.net/psm/newyork.csv

Randolph, J. J. (2014b). New York educational data set example after matching. Retrieved from http://justusrandolph.net/psm/newyork_nearest100.csv

R Core Team (2014). **R:** A language and environment for statistical computing. (3.0.3) [Computer software]. Vienna, Austria: Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Rubin D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(496), 322-331.

Rudner, L. M., & Peyton, J. (2006). Consider propensity scores to compare treatments. Practical Assessment Research & Evaluation, 11(9). Retrieved from: http://pareonline.net/getvn.asp?v=11&n=9

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Stone, C. A. & Tang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. Practical Assessment, Research & Evaluation, 18(13). Retrieved from: http://pareonline.net/getvn.asp?v=18&n=13

Note:

A previous version of this paper was delivered at the 2014 annual meeting of Mercer University Atlanta Research Conference, Atlanta, GA.

Citation:

Randolph, Justus J., Falbe, Kristina, Manuel, Austin Kureethara, & Balloun, Joseph L. (2014). A Step-by-Step Guide to Propensity Score Matching in **R**. *Practical Assessment, Research & Evaluation*, 19(18). Available online: http://pareonline.net/getvn.asp?v=19&n=18

Corresponding Author:

Justus J. Randolph
Tift College of Education
Mercer University
3001 Mercer University Dr.
Atlanta, GA 30341
randolph_jj [at] mercer.edu