

Using SAS® to Employ Propensity Score Matching in an Institutional Research Office to Create Matched Groups for Outcomes Analyses

Bobbie E .Frye, Central Piedmont Community College; James E. Bartlett, North Carolina State University

ABSTRACT

It is common to encounter student unit record data in the community college and to analyze the impact of educational interventions using two groups of students, those exposed to the intervention and those not exposed. Yet results are limited in that the students are not typically randomly selected into experimental and control groups. Non-random selection implies that the two groups of students may be very different on key factors that affect the results of analyses through self-selection bias and other differences. Propensity matching is a technique designed to simulate an experimental design, controlling for selection bias and creating almost equivalent experimental and comparison groups on key indicators. Propensity score matching using key characteristics such as diagnostic/placement test scores, Pell status, age, gender and race/ethnicity will be used to select the experimental and comparison groups. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010).

INTRODUCTION

Researchers acknowledge the limitations of comparing student performance, progression, or retention in a non-scientific study where participants are not randomly assigned or equivalent in terms of motivation, intentions, background, or skill level (Titus, 2007). While random selection is the “gold standard” of experimental designs (St. Pierre, 2006), random selection is often impractical, perceived unethical and resisted in educational settings. Propensity score matching (PSM) techniques are alternatively used to measure the counterfactual; that is, what would have happened to a similar group not receiving the treatment through choice or self-selection (Titus, 2007). An assumption of randomized selections and experimental designs is that biases are randomly distributed across categories in both the experimental and control groups. Propensity score matching is a technique designed to simulate an experimental design, controlling for selection bias, and creating almost equivalent experimental and control groups on key indicators.

APPROACHES AVAILABLE

There are several types of approaches available for institutional researchers when creating matched groups for analyses. New student groups are recommended as a best practice so that student groups are at the same place in their studies:

- Compare a new student cohort group of students exposed to intervention or program (treatment) to a group not exposed (control).
- Compare a new student cohort group of students exposed to intervention (treatment) to historical past group not exposed (control).
- Compare multiple cohort treatment and control groups, with different entry points, and track each derived set for the same amount of time.
- Compare cohorts of students taking the same course(s) with treatment students experiencing different delivery methods, interventions, etc. than control group.
- Compare students in multiple institutions, which requires establishing hierarchy to multilevel data (i.e., student level and institutional level).

SETTING UP THE DATASET

The student record dataset design includes independent variables on all cases and, if possible, derived outcomes from students in the full sample. A research identification number should be established and retained in the dataset. The dependent variable is coded 1 for cases in the treatment group and 0 for cases in the control group. It is up to the analyst to prepare the dataset for analysis and to take into account the context with which the study is conducted. Independent variables that are categorical are designated as categorical during the logistical regression procedure, or dummy value pre-coding is an option. For convenience this study used dummy coding of categorical variables to facilitate the analyses. **Table 1** shows the coding scheme for the study (finalprop) dataset used in the paper located at:

`LIBNAME bje6 'g:\finalprop';`

Table 1

Coding Scheme for Variables in the Analysis Dataset – bje6 .finalprop

Covariates	Variable Type	Variable Name	Coding
ResearchID		ResearchID	
DevMathAttempter	Categorical	Count1	1, Yes 0 No
Black	Dummy variable	Race1	1, Yes 0 No
Other	Dummy variable	Race2	1, Yes 0 No
White (reference)	Dummy Variable	Race3	1, Yes 0 No
Female	Dummy variable	Gen1	1, Yes 0 Male
Non-US Citizen	Dummy variable	Citizen	1, Yes 0 No
Age < 20	Dummy Variable	Age1	1, Yes 0 No
Age greater than 20 and less than 24	Dummy Variable	Age2	1, Yes 0 No
Age greater than 24 and less than 29	Dummy Variable	Age3	1, Yes 0 No
Age greater than 30 (reference)	Dummy Variable	Age4	1, Yes 0 No
Pell Recipient	Dummy variable	PellAmt1	1, Yes, 0 No
Enrolled Full Time	Dummy variable	Status	1, Yes, 0 Part-time
Associates College Transfer Program	Dummy variable	Prog1	1, Yes, 0 No
Associates Career and Technical Education Program (CTE)	Dummy variable	Prog2	1, Yes, 0 No
Certificates & Diplomas	Dummy variable	Prog3	1, Yes 0 No
Non-Declared (reference)	Dummy variable	Prog4	1, Yes, 0 No
Placed into level 4-devmath	Dummy variable	Mat1	1, Yes, 0 No
Placed into level 3-devmath	Dummy variable	Mat2	1, Yes, 0 No
Placed into level 2-devmath	Dummy variable	Mat3	1, Yes, 0 No
Placed into level 1-devmath	Dummy variable	Mat4	1, Yes, 0 No

(reference)

Outcomes

# Transfer Courses Attempted	Continuous	Transfer
# Technical CTE Courses Attempted	Continuous	TechCareer
# Developmental Courses Attempted	Continuous	Developmental
Attempted ACA Course	Continuous	ACA
Attempted College English Course	Continuous	ENG
Attempted College Algebra Course	Continuous	ALG
Total Courses Attempted	Continuous	Courseattempt
Total Courses Completed	Continuous	Coursecomp
Total Courses Completed C or better	Continuous	Success
Completed Credential	Continuous	Completer
Transferred Out	Continuous	Transfer

SAMPLE SIZE

Logistic regression requires adequate sample sizes based on the number of covariates selected. Hosmer & Lomeshow recommend sample sizes greater than 400 and 10 observations per estimated parameter (as cited in Hair et al., 2006). Larger sample sizes are likely to show statistical significant results when running statistical t-tests analyses and it is recommended to compute effect sizes for the power of treatment effects such as Cohen's d. The control group should be 2.5 to 3 times larger than the treatment group in order to find comparable propensity score matches for the treatment group examined (Caliendo & Kopeinig, 2005; Imbens, 2000).

THE ESSENTIAL STEPS

Since PSM is a multivariate statistical technique, there are multiple steps and decisions involved in the analysis: data pre-screening, covariate identification, propensity score estimation, matching of propensity scores, determination of matching success, and presentation of results. General outlines of the components of the essential steps are presented in the following sections. In the study, student record data were used to implement propensity score matching with the full new student sample. The greedy match macro employed here was provided in the Proceedings of the SUGI 26, by Lori S, Parsons. This application of the technique at Central Piedmont Community College is described.

PRE-SCREENING

Pre-screening data involves running frequencies and checking for missing values on all study variables. A rule of thumb is that a small number of missing values are not an issue. List-wise deletion is the preferred method when there are less than 5% of missing values (Hair et al., 2006). Variables with more than 5% of missing values should be further analyzed for adjustments.

Multi-collinearity is addressed by the researcher prior to the execution of the propensity score estimation. When trying to determine the importance of individual independent variables, multi-collinearity tends to

distort the prediction equation because some of the independent variables are highly correlated to each other. Collinearity statistics are available through options in the PROC REG technique or PROC CORR in order to address multi-collinearity among the independent variables, the researcher should exclude the variable of least importance to the study and retain the most important variable.

Appendix1 contains the code for the pre-screening analyses including running PROC FREQ.

SELECTION OF PRE-TREATMENT COVARIATES

The researcher begins the study by identifying, coding, and deriving independent variables that can bias comparison studies. Although the selection of pre-treatment variables considers those variables that explain the variance in-group membership, there is some debate in the literature concerning the appropriate use of pre-treatment covariates (Titus, 2007). In general, covariate selection should be based on relevant theory and prior research findings. Since logistic regression techniques are used to predict group membership, PSM is used to examine a group exposed to a treatment compared to a similar group not exposed to the treatment.

PROPENSITY SCORE ESTIMATION

SAS/STAT® allows users to perform multivariate logistic regression with the PROC LOGISTIC procedure. PROC LOGISTIC options allow users to calculate and save the predicted probability of the dependent variable, the propensity score, for each observation in the data set.

The initial use of stepwise logistic regression permits the researcher to identify significant covariates by using multiple quantitative independent variables to predict the probability of group membership (dependent variable). PSM allows the researcher to simplify the analysis by creating a one-number composite of all the covariates and then using the propensity score to match students. Propensity scores represent the “conditional probability of a person being in one condition rather than another given a set of observed covariates used to predict a person’s condition” (Rosenbaum & Rubin, 1984, p. 4).

A stepwise logistic regression was conducted to determine which independent variables were predictors of the dependent variable defined as participating in the developmental math intervention versus not participating in the developmental math intervention. Regression results indicated that the overall model of 7 predictors (mat1, gen1, pellamt1 age1, status, prog1, prog2) were statistically reliable in distinguishing between taking and not taking developmental math intervention among students. (-2 Log Likelihood=2364.44), chi-squared =206.405, $p < .001$, R Squared=.141. The model correctly classified 70.0% of the cases and explained 15% of the variance in the dependent variable. Regression coefficients are presented in **Table 2**.

Table 2

Results of Stepwise Logistic Regression- Covariates in the Model

Parameter	DF	Estimate	Standard	Wald	Pr > ChiSq	Exp(Est)
Intercept	1	-0.4846	0.2023	5.7384	0.0166	0.616
mat1	1	0.8195	0.1081	57.4441	<.0001	2.269
gen1	1	-0.2448	0.1085	5.0913	0.024	0.783
pellamt1	1	-0.4303	0.1203	12.8003	0.0003	0.65
age1	1	-0.6567	0.1107	35.2124	<.0001	0.519
status	1	0.8341	0.1302	41.0366	<.0001	2.303

prog1	1	-0.9428	0.1701	30.7381	<.0001	0.39
prog2	1	-0.8169	0.1762	21.492	<.0001	0.442

Appendix 2 contains the code for PROC LOGISTIC procedure used to determine which covariates were predictors of the dependent variable- participation in the intervention.

PROPENSITY SCORE MATCHING

Matching refers to a variety of functions that capture students who are similar to each other and create a subset of data that, on average, is balanced in terms of relevant variables. The matching functions allow the researcher to limit the number of matches and to control the standard deviation or distance from the propensity scores of the student in the comparison group and the corresponding study group. There are several matching algorithms available but with large sample sizes, such as are often found in student unit record data, the outcomes among techniques are typically similar (Rojewski et al., 2010). However, it is important to compare results across specifications.

The most common matching algorithm is nearest neighbor matching. Within nearest neighbor matching, a few options are available to researchers, specifically, matching with replacement and without replacement. *With replacement* means a case is considered more than once in the matching procedure. Matching *without replacement* means, after matching, the case is removed from further consideration in matching. Both types of nearest neighbor matching will affect the variance explained by the model and the bias on key indicators. With replacement is preferred when there are many cases in the treated group with high propensity scores but only a few matching cases in the comparison group (Caliendo & Kopeinig, 2008). The matching algorithm used in this paper is a greedy, nearest neighbor matching function. Once a match is made the treatment case is not reconsidered. The treatment cases are ordered and sequentially matched to the nearest un-matched control. If more than one un-matched control matches to a treatment case, the control match is selected at random (Parsons, SUGI 26)

APPENDIX 3. contains the code for the SAS Greedy 5 to 1 Digit Match Macro. Greedy 5 to 1 match means that the cases were first matched to controls on 5 digits of the propensity score. For those that did not match, a subsequent match was run to match remaining cases on 4 digits of the propensity score. The matching process continued attempts to match remaining cases until reaching a 1 digit match as the final attempt (Parsons, SUGI 26). Although other matched processes are available, the most common one used at CPCC is the 5 to 1 Greedy Matched Macro.

DETERMINATION OF MATCHING SUCCESS

After matching, the quality of the match should be assessed and measured statistically using t-tests or chi-square as appropriate (Gemici, Rojweksi & Lee, 2012; Oakes & Johnson, 2006). Alternatively, covariate imbalance can be assessed using a standardized measure of difference or effect size. The standardized effect size is the difference between the sample means in the treated and control (full or matched) samples as a measure of the square root of the average of the sample variances in the treated and control groups (Rosenbaum & Rubin, 1985). A standardized difference in means less than 0.25 has been suggested as a threshold, where differences in means are standardized by the standard deviation in the initial active treatment group (Rubin, 2001, Stuart, 2010). The effect size can be calculated using an online calculator using means and standard deviations. When presenting data from propensity score matching analysis, results are reported both pre- and post- match. The method, i.e., PROC LOGISTIC, used to determine propensity score estimation should be identified, as should the method used to determine matching success. A match is deemed successful if after matching there are little or no differences between the groups on the initial covariates (Gelman & Hill, 2007).

In **Table 3** T-tests were conducted to determine which variables were significantly different among the two groups of the dependent variable defined as placing into developmental math and participating or not

participating in a developmental math intervention. Prior to the propensity match, the groups were statistically different on all but three variables: gen1 race1 prog2.

Table 3

Sample SAS Output-T-Tests Before Matching

T-Tests					
Variable	Method	Variances	DF	t Value	Pr > t
mat1	Pooled	Equal	1984	7.77***	<.0001
mat1	Satterthwaite	Unequal	1004	7.69***	<.0001
mat2	Pooled	Equal	1984	-2.99**	0.0028
mat2	Satterthwaite	Unequal	1209	-3.22**	0.0013
mat3	Pooled	Equal	1984	-3.04**	0.0024
mat3	Satterthwaite	Unequal	1133	-3.18**	0.0015
gen1	Pooled	Equal	1984	-1.83	0.0681
gen1	Satterthwaite	Unequal	1020	-1.82	0.0689
pellamt1	Pooled	Equal	1984	-2.68**	0.0094
pellamt1	Satterthwaite	Unequal	1077	-2.66**	0.008
age1	Pooled	Equal	1984	-6.92***	<.0001
age1	Satterthwaite	Unequal	953	-6.68***	<.0001
age2	Pooled	Equal	1984	3.04**	0.0024
age2	Satterthwaite	Unequal	927	2.89**	0.0039
age3	Pooled	Equal	1984	3.45***	0.0006
age3	Satterthwaite	Unequal	822	3.06**	0.0023
race1	Pooled	Equal	1984	-0.11	0.911
race1	Satterthwaite	Unequal	1031	-0.11	0.9108
race2	Pooled	Equal	1984	2.7**	0.0071
race2	Satterthwaite	Unequal	1002	2.67**	0.0078
status	Pooled	Equal	1984	8.05***	<.0001
status	Satterthwaite	Unequal	1297	8.93***	<.0001
citizen	Pooled	Equal	1984	2.09*	0.0364
citizen	Satterthwaite	Unequal	879	1.93	0.0536
prog1	Pooled	Equal	1984	-4.42***	<.0001
prog1	Satterthwaite	Unequal	1025	-4.42***	<.0001
prog2	Pooled	Equal	1984	0.720	0.4723
prog2	Satterthwaite	Unequal	1017	0.720	0.4742
prog3	Pooled	Equal	1984	2.54*	0.0111
prog3	Satterthwaite	Unequal	838	2.28*	0.0229

Note: ***p<.001, ** p <.01, *p <.05

After propensity matching, no covariates were statistically different between the two groups. The match was deemed successful in creating two equivalent groups as shown in **Table 4**.

Table 4

Sample SAS Output-T-Tests After Matching

T-Tests					
Variable	Method	Variances	DF	t Value	Pr > t
mat1	Pooled	Equal	1010	0	1
mat1	Satterthwaite	Unequal	1010	0	1
mat2	Pooled	Equal	1010	0.1	0.9212
mat2	Satterthwaite	Unequal	1010	0.1	0.9212
mat3	Pooled	Equal	1010	-0.31	0.756
mat3	Satterthwaite	Unequal	1010	-0.31	0.756
gen1	Pooled	Equal	1010	0	1
gen1	Satterthwaite	Unequal	1010	0	1
pellamt1	Pooled	Equal	1010	-0.27	0.7867
pellamt1	Satterthwaite	Unequal	1010	-0.27	0.7867
age1	Pooled	Equal	1010	-0.26	0.7985
age1	Satterthwaite	Unequal	1010	-0.26	0.7985
age2	Pooled	Equal	1010	0.31	0.7551
age2	Satterthwaite	Unequal	1010	0.31	0.7551
age3	Pooled	Equal	1010	0.23	0.816
age3	Satterthwaite	Unequal	1009	0.23	0.816
race1	Pooled	Equal	1010	0.1	0.9229
race1	Satterthwaite	Unequal	1010	0.1	0.9229
race2	Pooled	Equal	1010	1.02	0.3078
race2	Satterthwaite	Unequal	1010	1.02	0.3078
status	Pooled	Equal	1010	0.08	0.9353
status	Satterthwaite	Unequal	1010	0.08	0.9353
citizen	Pooled	Equal	1010	0.79	0.4321
citizen	Satterthwaite	Unequal	1002	0.79	0.4321
prog1	Pooled	Equal	1010	0.31	0.7532
prog1	Satterthwaite	Unequal	1010	0.31	0.7532
prog2	Pooled	Equal	1010	0.06	0.9487
prog2	Satterthwaite	Unequal	1010	0.06	0.9487
prog3	Pooled	Equal	1010	-0.53	0.5949
prog3	Satterthwaite	Unequal	1006	-0.53	0.5949

Note: ***p<.001, ** p <.01, *p <.05

Before and after matching propensity scores (saved in the datasets) should be assessed to ensure that the distributions are similar across the two groups and that outliers are not present in the propensity scores that could affect the analysis. Box plot examinations and propensity score distributions are useful for determining whether outliers should be addressed. In some cases, outliers can be eliminated using a minimum maximum score range of common support. However, if a small number of outliers are detected or the outliers are believed to represent a portion of the population, the researcher may decide not to eliminate the outliers and continue with the analysis. One option is to run both analyses and determine the differences for both analyses (Rojewski et al., 2010). The example shown below did not eliminate any outliers. The treatment cases and controls contain a different distribution of the propensity scores before and after the match. However, after matching the median values and propensity score ranges are balanced (**Tables 5-8**).

Table 5

Before Matching- Propensity Score Distributions for Control Group

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	0.7629593
99%	0.7629593
95%	0.625337
90%	0.5864927
75% Q3	0.4797995
50% Median	0.3558641
25% Q1	0.222688
10%	0.1744478
5%	0.1236647
1%	0.0887554
0% Min	0.0595659

Table 6

Before Matching- Propensity Score Distributions for Treatment Group

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	0.7629593
99%	0.625337
95%	0.4953372
90%	0.4199425
75% Q3	0.3372975
50% Median	0.222688
25% Q1	0.1551205
10%	0.110647
5%	0.0887554
1%	0.0595659
0% Min	0.0595659

Table 7*After Matching- Propensity Score Distributions for Control Group*

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	0.7629593
99%	0.625337
95%	0.5864927
90%	0.5268047
75% Q3	0.4199425
50% Median	0.3372975
25% Q1	0.222688
10%	0.1570424
5%	0.1236647
1%	0.0887554
0% Min	0.0595659

Table 8*After Matching- Propensity Score Distributions for Treatment Group*

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	0.7629593
99%	0.625337
95%	0.5864927
90%	0.5268047
75% Q3	0.4199425
50% Median	0.3372975
25% Q1	0.222688
10%	0.1570424
5%	0.1236647
1%	0.0887554
0% Min	0.0595659

The SAS System

The UNIVARIATE Procedure

Variable: prob (Estimated Probability)

Schematic Plots

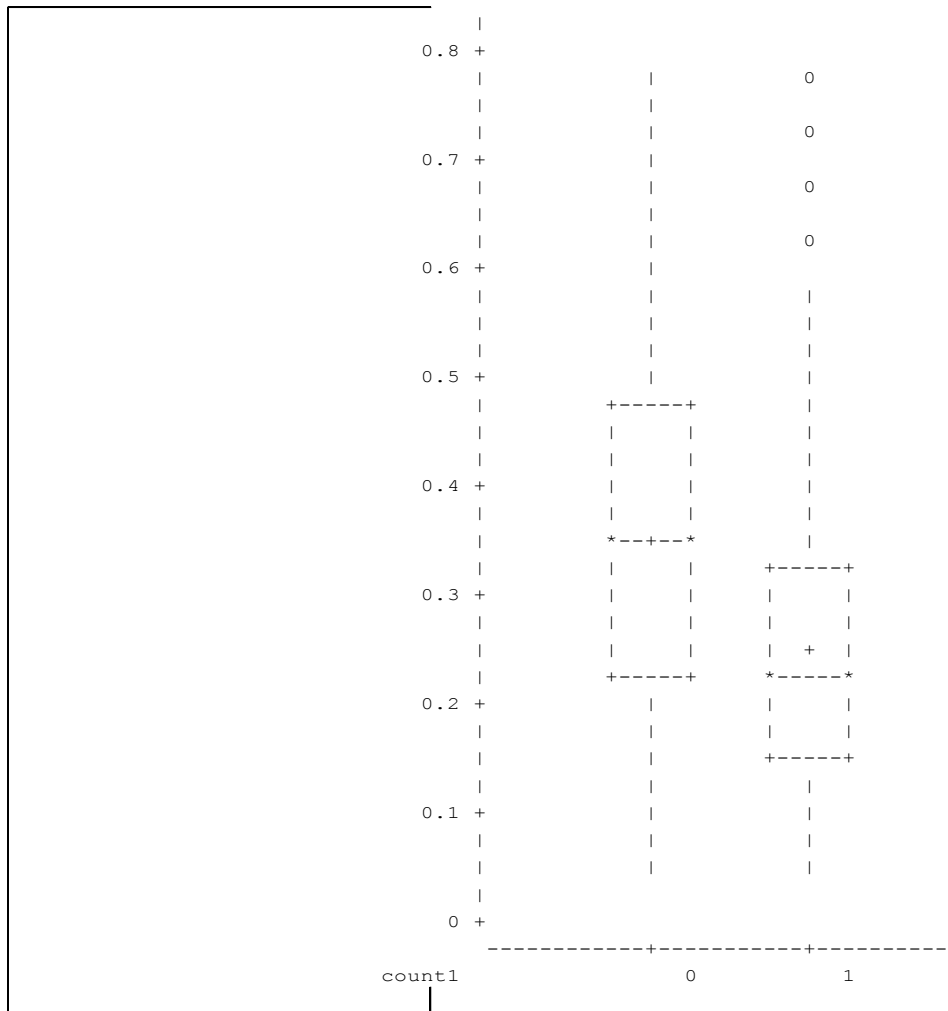


Figure 1 Boxplots of estimated probabilities output before applying GREEDYMATCH algorithm

The SAS System

The UNIVARIATE Procedure

Variable: prob (Estimated Probability)

Schematic Plots

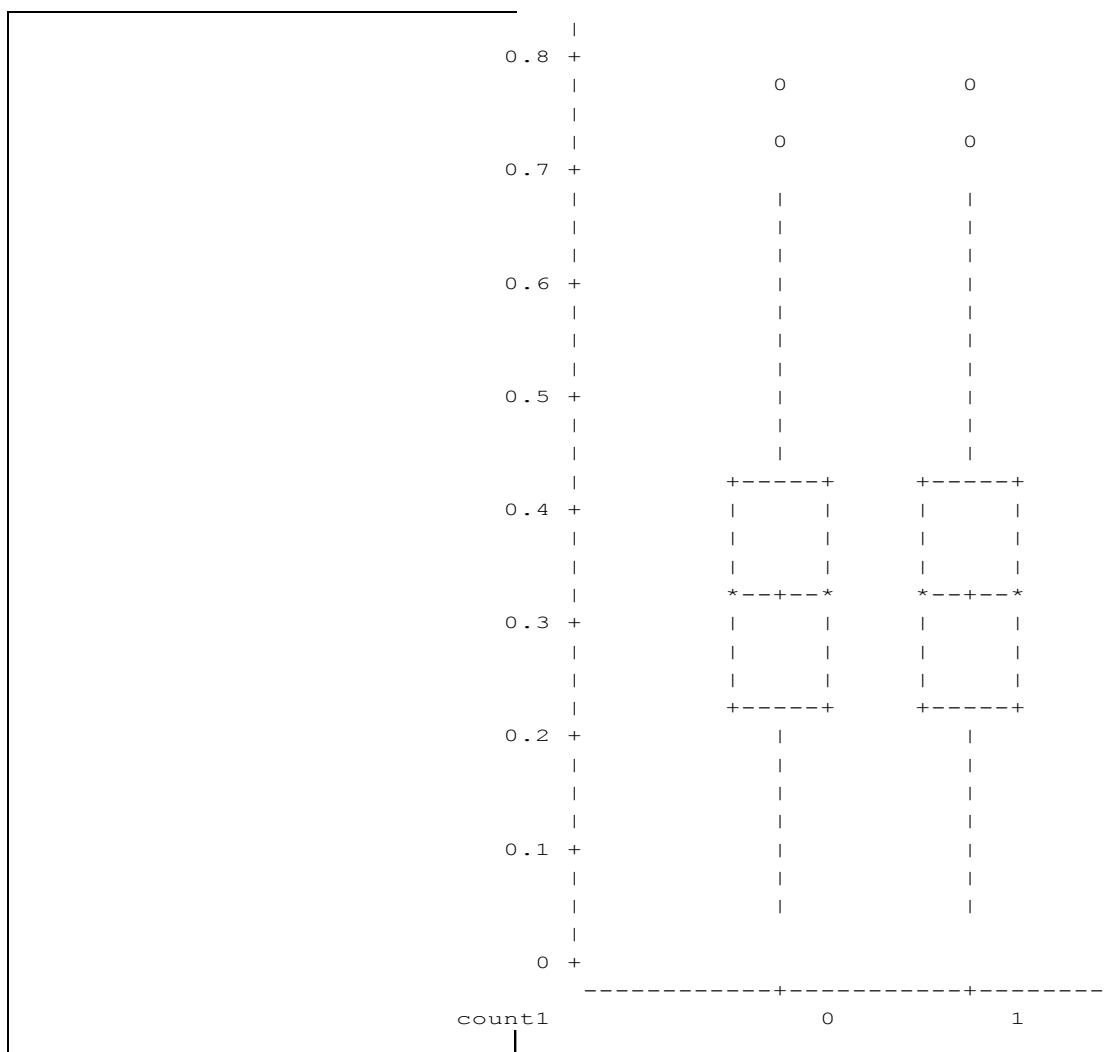


Figure 2 Boxplots of estimated probabilities output after applying GREEDYMATCH algorithm

Appendix 4 contains the code employed to assess matching success. Particular attention is given to the differences in covariates and the propensity score distributions before and after matching.

PRESENTING THE RESULTS

Outcomes of interest are reported after the matching procedure and measure the treatment effect or difference in outcomes between the two groups. When propensity scores matching is used to create a matched group for analyses, multiple regression can subsequently be employed to measure program impact with a continuous dependent variable (such as credits completed) and the treatment variable as an independent variable (Schuler, 2015). In theory, any differences in outcomes between the two groups are associated with the treatment or intervention since PSM has eliminated the variation between the two groups (Gelman & Hill, 2007; Gemici et al., 2012).

Employing t-tests for analyses, the students in the intervention treatment group attempted an average of 6.39 transferrable courses compared to 4.43 courses for the control group and successfully completed an average of 7.37 courses compared to 4.91 for the control group. T-tests indicated that the treatment group students did significantly better on every outcome assessed except attempted number of techcareer courses (**Table 9**).

Table 9

Sample SAS Output-Four- Year Outcomes of Treatment and Control Groups After Matching

T-Tests					
Variable	Method	Variances	DF	t Value	Pr > t
transfer	Pooled	Equal	1010	-4.72***	<.0001
transfer	Satterthwaite	Unequal	1010	-4.72***	<.0001
techcareer	Pooled	Equal	1010	-0.28	0.7794
techcareer	Satterthwaite	Unequal	1001	-0.28	0.7794
developmental	Pooled	Equal	1010	-18.87***	<.0001
developmental	Satterthwaite	Unequal	883	-18.87***	<.0001
success	Pooled	Equal	1010	-5.57***	<.0001
success	Satterthwaite	Unequal	1000	-5.57***	<.0001
coursecomp	Pooled	Equal	1010	-7.17***	<.0001
coursecomp	Satterthwaite	Unequal	993	-7.17***	<.0001
courseattempt	Pooled	Equal	1010	-9.08***	<.0001
courseattempt	Satterthwaite	Unequal	1001	-9.08***	<.0001
aca	Pooled	Equal	1010	-6.97***	<.0001
aca	Satterthwaite	Unequal	977	-6.97***	<.0001
eng	Pooled	Equal	1010	-7.02***	<.0001
eng	Satterthwaite	Unequal	976	-7.02***	<.0001
alg	Pooled	Equal	1010	-2.11*	0.0349
alg	Satterthwaite	Unequal	985	-2.11*	0.0349

Note: ***p<.001, ** p <.01, *p <.05

APPENDIX 5 contains the code used to evaluate differences in outcome means between the intervention and control group after matching. Comparisons of student outcomes using propensity matching has been used to yield less biased results than are derived using simple comparisons (Rojewski et al., 2010).

CONCLUSION

As we have shown, we used a variety of SAS building blocks or procedures to employ propensity score matching. Evaluating frequencies, descriptive statistics, boxplots, probability distributions and statistical differences before and after matching are important throughout the analyses. Primarily, start with a dataset of treatment and control students and create a purposeful sample using PSM. The matching algorithm used in this paper is a greedy, nearest neighbor matching function. Once a match is made the treatment case is not reconsidered. (Parsons, SUGI 26). Comparisons of student outcomes using propensity matching has been shown to yield less biased results than are derived using simple t-test comparisons (Rojewski, Lee, & Gemici, 2010). Therefore, propensity score matching is an extremely useful tool for evaluating interventions and programs in educational environments.

REFERENCES

- Agresti, A., & Finlay, B. (2009). *Statistical methods for the social science* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.
- Caliendo, M., & Kopeinig, S. (2005, May). *Some practical guidance for the implementation of propensity score matching* [Discussion Paper No. 1588]. Bonn, Germany: University of Bonn, Institute for the Study of Labor (IZA). National Bureau of Economic Research Cambridge, Mass., USA.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gemici, S., Rojewski, J. W., & Lee, I. H. (2012). Use of propensity score matching for training research with observational data. *International Journal of Training Research*, 10(3), 219-232
- Green, S. B., & Salkind, N. J. (2001). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Imbens, G. (2000). The role of the propensity score in estimating does-response functions. *Biometrika*, 87, 706-710.
- Mertler, C. A., & Vannatta, R. A. (2010). *Advanced and multivariate statistical methods: Practical applications and interpretations* (4th ed). Glendale, CA: Pycrzak Publishing.
- Oakes, J.M., & Johnson, P.J. (2006). Propensity score matching for social epidemiology. In J.M. Oakes & J.M. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 364-386). San Francisco, CA: Jossey-Bass.
- Parsons S. Lori. "Reducing Bias in a Propensity Score Matched -Pair Sample Using Greedy Matching Techniques, Paper 214-26." Proceedings of the SUGI 26 , SAS Institute Inc. Available at <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>
- Rojewski, J.W., Lee, I.H., & Gemici, S. (2010). Using propensity score matching to determine the efficacy of secondary career academies in raising educational aspirations. *Career and Technical Education Research*, 35(1), 3-27.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.
- Rosenbaum, P.R., & Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387), 516-524.

- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.
- Schuler, M. (2015). Propensity score analysis: fundamentals and developments. In W. Pan & H. Bai (Eds.), *Overview in implementing propensity score analyses in statistical software* (pp. 20-48). New York, NY: Guilford.
- St. Pierre, E.A. (2006). Scientifically-based research in education: Epistemology and ethics. *Adult Education Quarterly*, 56(4), 239-266. doi: 10.1177/0741713606289025.
- Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21.
- Titus, M. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: an application to the private returns to a master's degree. *Research in Higher Education*, 48(4), 487-521.

ACKNOWLEDGEMENTS

Parsons, Lori S. GMATCH Algorithm SAS Code Reference -SAS paper 214-26 Reducing Bias in a Propensity Score Matched -Pair Sample Using Greedy Matching Techniques

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social science* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall. – SAS Code Reference

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Bobbie E. Frye
Central Piedmont Community College
704-330-6459
bobbie.frye@cpcc.edu

James E. Bartlett
North Carolina State University
919-208-1697
James_bartlett@ncsu.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX 1.

```
/*Run PROC FREQUENCY PROCEDURE on all variables in the dataset
bje6 .finalpropen checking for missing values*/

PROC SORT DATA=bje6.finalpropen; by count1;
PROC FREQ DATA=bje6.finalpropen; by count1;
    TABLES mat1 mat2 mat3 gen1 pellamt1 age1 age2 age3 race1 race2
    status citizen prog1 prog2 prog3;
RUN;
```

APPENDIX 2.

```
/*Perform a PROC LOGISTIC PROCEDURE
on the dataset bje6 .Finalpropen and save
the propensity score data set
bje6 .propen and the propensity score (prob).
Categorical dependent variable Name=count1
Statistic Name=prob
*/

LIBNAME bje6 'g:\finalprop';

PROC LOGISTIC DATA= bje6.finalprop;
Class <List Categorical Variables if applicable>;
MODEL count1(event='1')=
    mat1 mat2 mat3 gen1 pellamt1 age1 age2 age3 race1 race2 status
    citizen prog1 prog2 prog3 <List Categorical Variables if
    applicable> / expb selection=stepwise risklimits lackfit rsquare;
OUTPUT OUT=bje6.propen prob=prob;
RUN;
```

APPENDIX 3.

```
/*Evaluate the propensity scores using
frequencies to determine minimum and maximum values.
Ensure there is considerable overlap between the propensity scores in
in both groups.*/

PROC SORT DATA=bje6.propen; by count1;
PROC FREQ DATA=bje6.propen; by count1;
TABLES prob;
RUN;

/*Evaluate boxplots of the estimated probabilities to determine
minimum and maximum values and to observe probability distributions in
each group */
```

```

PROC CHART DATA=bje6.propen; vbar prob; by count1;
PROC UNIVARIATE PLOT DATA=bje6.propen; var prob; by count1;
  RUN;

/*Evaluate the difference in means, between groups on each covariate
by running a PROC TTEST PROCEDURE with the treatment variable
as the class variable.*/

PROC TTEST DATA=bje6.propen; class count1;
  var mat1 mat2 mat3  gen1 pellamt1 age1 age2 age3 race1 race2
  status  citizen prog1 prog2 prog3;
  RUN;

/*Evaluate the difference in associations on each categorical
covariate between groups by running chi-square with the treatment
variable as the weighted variable.*/

PROC FREQ DATA=bje6.propen; weight count1;
TABLES (categorical variables) /chisq expected measures;
  RUN;

```

APPENDIX 4.

```

/*Call statement for greedy match macro*/;

%greedmtch(bje6,propen,count1,matches);

/*Greedy 5>1 digit matching macro*/
%macro greedmtch
(lib, /*Library name*/
dataset, /*Data set of all students*/
depend, /*Dependent variable that indicates treatment case or control,
code 1 for cases,0 for controls*/
matches/*Output file of matched pairs*/
);
%macro sortcc;
proc sort data=tcases
out=&lib..Scase;
by prob;
  RUN;
proc sort data=tctrl
out=&lib..Scontrol;
by prob randnum;
  RUN;
%mend sortcc;

%macro initcc(digits);

data tcases (drop=cprob) tctrl (drop=aprob);
set &LIB..&dataset.;

```

```

if &depend.= 0 and prob ne . then do;
cprob=round(prob,&digits.);
Cmatch=0;
length randnum 8;
randnum=ranuni(1234567);
label randnum='Uniform Randomization Score';
output tctrl;
end;
else if &depend.= 1 and prob ne . then do;
Cmatch=0;
aprob=round(prob,&digits.);
output tcases;
end;
RUN;

%sortcc;
%mend initcc;

%macro match (matched,digits);
data &lib..&matched.
(drop=Cmatch randnum aprob cprob start oldi curctrl matched);
set &lib..Scase ;
curob + 1;
matchto=curob;

if curob=1 then do;
start=1;
oldi=1;
end;

do i= start to n;
set &lib..Scontrol point= i nobs= n;

if i gt n then goto startovr;
if _error_=1 then abort;

curctrl = i;
if aprob = cprob then do;
Cmatch=1;
output &lib..&matched.;
matched=curctrl;
goto found;
end;
else if cprob gt aprob then
goto nextcase;

startovr: if i gt n then goto nextcase;
end;

nextcase:

```

```

if Cmatch=0 then start=oldi;

found:
if Cmatch=1 then do;
oldi=matched+1;
start=matched +1;
set &lib..scase point=curob;
output &lib..&matched.;
end;
retain oldi start;
if _error_=1 then _error_=0;
RUN;

proc sort data=&lib..Scase out=sumcase;
by researchid;
RUN;
proc sort data=&lib..Scontrol
out=sumcontrol;
by researchid;
RUN;
proc sort data=&lib..&matched. out=smatched (keep=researchid matchto);
by researchid;
RUN;
data tcases (drop=matchto);
merge sumcase (in=a) smatched;
by researchid;
if a and matchto= .;
Cmatch=0;
aprob=round(prob,&digits.);
RUN;
data tctrl (drop=matchto);
merge sumcontrol(in=a) smatched;
by researchid;
if a and matchto= .;
Cmatch=0;
cprob=round(prob,&digits.);
RUN;
%sortcc
%mend match;

%initcc(.00001);
%match(Match5,.0001);
%match(match4,.001);
%match(match3,.01);
%match(match2,.1);
%match(match1,.1);
Data &lib..&matches.;
set &lib..match5 (in=a)
&lib..match4 (in=b) &lib..match3 (in=c) &lib..match2 (in=d)
&lib..match1 (in=e);
if b then matchto=matchto + 100000;
if c then matchto=matchto + 1000000;

```

```

if d then matchto=matchto + 1000000000;
if e then matchto=matchto + 100000000000;
RUN;

proc sort data=&lib..&matches. out=&lib..S&matches.;
by &depend.;
RUN;

%mend greedmtch;

```

APPENDIX 5.

```

/*Evaluate the propensity scores using
frequencies to determine minimum and maximum values.
Ensure there is considerable overlap between the propensity scores in
in both groups after matching.*/

PROC SORT DATA=bje6.smatches; by count1;
PROC FREQ DATA=bje6.smatches; by count1;
TABLES prob;
RUN;

/*Evaluate boxplots of the estimated probabilities by employing the
PROC CHART and PROC UNIVARIATE procedures to determine balance of the
estimated probabilities for matched groups and to observe probability
distributions in each group after matching*/

PROC CHART DATA=bje6.smatches; vbar prob; by count1;
PROC UNIVARIATE PLOT DATA=bje6.smatches; var prob; by count1;
RUN;

/*Evaluate the difference in means on each covariate between groups by
running a t-tests with the treatment variable
as the class variable.*/

PROC TTEST DATA=bje6.smatches; class count1;
var mat1 mat2 mat3 gen1 pellamt1 age1 age2 age3 race1 race2
status citizen prog1 prog2 prog3;
RUN;

/*Check outcomes between groups, on the matched dataset, by employing
the PROC TTEST PROCEDURE and listing the class variable as the
treatment or group variable. */

PROC TTEST DATA=bje6.smatches; class count1;
var transfer techcareer developmental success coursecomp courseattempt
aca eng alg completer transfer;
RUN;

```