Paper SAS332-2017

Propensity Score Methods for Causal Inference with the PSMATCH Procedure

Yang Yuan, Yiu-Fai Yung, and Maura Stokes, SAS Institute Inc.

Abstract

In a randomized study, subjects are randomly assigned to either a treated group or a control group. Random assignment ensures that the distribution of the covariates is the same in both groups and that the treatment effect can be estimated by directly comparing the outcomes for the subjects in the two groups. In contrast, subjects in an observational study are not randomly assigned. In order to establish causal interpretations of the treatment effects in observational studies, special statistical approaches that adjust for the covariate confounding are required to obtain unbiased estimation of causal treatment effects.

One strategy for correctly estimating the treatment effect is based on the propensity score, which is the conditional probability of the treatment assignment given the observed covariates. Prior to the analysis, you use propensity scores to adjust the data by weighting observations, stratifying subjects that have similar propensity scores, or matching treated subjects to control subjects.

This paper reviews propensity score methods for causal inference and introduces the PSMATCH procedure, which is new in SAS/STAT® 14.2. The procedure provides methods of weighting, stratification, and matching. Matching methods include greedy matching, matching with replacement, and optimal matching. The procedure assesses covariate balance by comparing distributions between the adjusted treated and control groups.

Introduction

In a randomized study, such as a randomized clinical trial, the subjects are randomly assigned to a treated (exposure) group or a control (non-exposure) group. Random assignment ensures that the distribution of the covariates is the same in both groups and that the treatment effect can be estimated by directly comparing the outcomes for the subjects in the two groups.

In contrast, the subjects in an observational study, such as a retrospective cohort study or a nonrandomized clinical trial, are not randomly assigned to the treated and control groups. Confounding can occur if one or more covariates are related to both the treatment assignment and the outcome. Consequently, there can be systematic differences between the treated subjects and the control subjects prior to the application of the treatment. Such differences are sources of confounding, which obscure the actual treatment effect.

Under the potential outcomes framework, in an observational study whose goal is to estimate the effect of a treatment, each individual typically has two potential outcomes:

- $Y(1)$, the outcome that would be observed if the individual receives the treatment
- $Y(0)$, the outcome that would be observed if the individual does not receive the treatment under identical circumstances to those under which the subject would have received the treatment

However, only one outcome can be observed.

The treatment effect is defined as $Y(1) - Y(0)$, and the average treatment effect is defined as

$$\text{ATE} = E(Y(1) - Y(0))$$

The average treatment effect for the treated (individuals who actually receive treatment) is defined as

$$\text{ATT} = E(Y(1) - Y(0) \mid T = 1)$$

where T denotes the treatment assignment.

In a randomized trial, the potential outcomes $(Y(0), Y(1))$ and the treatment assignment T are independent:

$$(Y(0), Y(1)) \perp\!\!\!\perp T$$

Thus, the average treatment effect (ATE) is identical to the average treatment effect for the treated (ATT), which can be expressed as follows and can be estimated from the observed data:

$$E(Y(1) \mid T = 1) - E(Y(0) \mid T = 0)$$

In an observational study, the potential outcomes $(Y(0), Y(1))$ and the treatment assignment T might not be independent. In this case, the ATE and ATT are not the same, and outcomes cannot be compared directly to estimate the treatment effect.

In the presence of confounding, statistical approaches that remove the effects of confounding are required in order to estimate the effect of treatment. One such approach is regression adjustment, which estimates the treatment effect after adjusting for differences in the baseline covariates. However, this approach has practical limitations, such as the difficulty in assessing the degree of overlap between the distribution of baseline covariates for the treated and control groups, as discussed by Austin (2011). Propensity score analysis is an alternative approach that circumvents many of these limitations.

The propensity score was defined by Rosenbaum and Rubin (1983, p. 47) as the probability of assignment to treatment conditional on a set of observed baseline covariates. Propensity score analysis minimizes the effects of confounding and provides some of the advantages of a randomized study. The theoretical foundation for propensity score methods is based on the causal effect model introduced by Rubin (1974).

Propensity Score Analysis

Rosenbaum and Rubin (1983) defined treatment assignment to be strongly ignorable when two conditions are met. The first condition (unconfoundedness) states that the potential outcomes $(Y(0), Y(1))$ and the treatment assignment T are conditionally independent given the observed baseline variables X:

$$(Y(0), Y(1)) \perp\!\!\!\perp T \mid X = x$$

This condition is called the "no unmeasured confounders" assumption because it assumes that all the confounding variables that affect both the outcome and the treatment assignment have been measured in X. The second condition (probabilistic assignment) states that there is a positive probability that a subject is assigned to the treated group or the control group:

$$0 < \Pr(T = 1 \mid X = x) < 1$$

In a randomized study, the potential outcomes within treated and control groups are unrelated to treatment assignment because individuals are randomly assigned to the groups. Consequently, the treatment assignment given the variables **X** is strongly ignorable.

When the treatment assignment in an observational study is assumed to be strongly ignorable, Rosenbaum and Rubin (1983, p. 43) showed that unbiased estimates of average treatment effects can be obtained by conditioning on the propensity score $e(x)$, which is the probability of the treatment assignment conditional on the set of confounding variables X:

$$e(x) = \Pr(T = 1 \mid X = x)$$

At any value of the propensity score $e(x)$, the difference between the means of the treated and control groups is an unbiased estimate of the average treatment effect at $e(x)$. Consequently, propensity score matching and propensity score stratification also produce unbiased estimates of treatment effects (Rosenbaum and Rubin 1983, p. 44).

Furthermore, the propensity score is a balancing score. At each value of the propensity score, the distributions of the variables X are the same in the treated and control groups (Rosenbaum and Rubin 1983, p. 44; Stuart 2010, p. 6). Thus, the treatment assignment T and observed variables X are conditionally independent given the propensity score Rosenbaum (2010, p. 72):

$$x \perp\!\!\!\perp T \mid e(x)$$

Propensity score analysis attempts to replicate the properties of a randomized trial with respect to the observed variables X. The following three methods are commonly used in propensity score analysis:

- weighting, which creates weights that are appropriate for estimating the ATE and ATT
- stratification, which creates strata based on propensity scores
- matching, which matches treated units with control units

You can use propensity score methods to create an output data set that contains a sample that has been adjusted (by matching, stratification, or weighting) so that the distributions of the variables are balanced between the treated and control groups. Thus, the two groups differ only randomly in their observed or measured variables, as in a randomized study. You can then use the output data set in an outcome analysis to estimate the effect of the treatment.

A propensity score analysis usually involves the following steps (Guo and Fraser 2015, p. 131):

1. You specify a set of confounding variables that might be related to both the treatment assignment and the outcome.
2. You use this set of variables to fit a logistic regression model and compute propensity scores. The response is the probability of assignment to the treated group.
3. You choose a propensity method (weighting, stratification, or matching) to compute observation weights (weighting), to construct strata of observations (stratification), or to select matched observations (matching).
4. You assess the balance of variables by comparing the distributions (before and after the propensity score weighting, stratification, or matching) between the treated and control groups.
5. To improve the balance, you can repeat the process with a different set of variables for the logistic regression model, a different region of support, a different set of matching criteria, or a different matching method.
6. When you are satisfied with the variable balance of the propensity score analysis, you save the output data set for subsequent outcome analysis.

Note that the outcome variable is not used in this process, and the variable selection is not related to the observed outcomes (Rubin 2001; Stuart 2010, p. 5). Any variables that might have been affected by the treatment should not be included in the process (Rosenbaum and Rubin 1984; Stuart 2010, p. 5).

After adequate variable balance has been achieved and assuming that no other confounding variables are associated with both the treatment assignment and the outcome, you can add the response variable to the output data set and perform an appropriate outcome analysis that is the same as the analysis you would perform with data from a randomized study. For example, if you used propensity score matching, a simple univariate test or analysis might be sufficient to estimate treatment effect.

The following section describes the features of the PSMATCH procedure, followed by descriptions of propensity score weighting, propensity score stratification, and propensity score matching, and then by examples that illustrate propensity score matching, variable balance assessment, and propensity score weighting.

Features of the PSMATCH Procedure

You can use the PSMATCH procedure to create propensity scores for observations from treated and control groups by fitting a binary logistic regression model. Alternatively, you can input propensity scores that have already been created by using a different model or even a different approach such as a tree-based method. For example, you can input propensity scores that have been computed by the LOGISTIC procedure using a binary probit model or by the HPSPLIT procedure using a classification tree.

The PSMATCH procedure provides the following methods for using the propensity scores to allow for valid estimation of treatment effect in a subsequent outcome analysis:

- Inverse probability of treatment weighting and ATT weighting (weighting by odds): The procedure computes weights from the propensity scores. These weights can then be incorporated into a subsequent analysis that estimates the effect of treatment.
- Stratification: The procedure creates strata of observations that have similar propensity scores. In a subsequent analysis, the treatment effect can be estimated within each stratum, and the estimates can then be combined across strata.
- Matching: The procedure matches each treated unit with one or more control units whose propensity score is similar. In a subsequent analysis, the treatment effect can be estimated by comparing outcomes between treated and control subjects in the matched sample. If the outcome values for a study have not been collected prior to matching, only the matched units are needed for follow-up. Thus, the cost of the trial is reduced (Stuart 2010, p. 2).

The PSMATCH procedure provides various ways to assess how well the distributions of variables are balanced between the treated and control groups after matching, weighting, or stratification. These variables include the propensity score, the logit of the propensity score, variables used in the logistic regression model, and other variables in the data set. The assessments include the following:

- differences in the distributions of the variables between the treated and control groups
- standardized differences in the variables between the treated and control groups
- percentage reductions of absolute differences

When you use stratification, the differences are also computed within each stratum.

The PSMATCH procedure also provides various plots for assessing balance. These plots include the following:

- cloud plots, which are scatter plots in which the points are jittered to prevent overplotting
- box plots for continuous variables
- bar charts for classification variables
- a standardized differences plot that summarizes differences between the treated and control groups

When you use stratification, the plots are also produced by stratum.

After adequate variable balance has been achieved, and assuming that no other confounding variables are associated with both the treatment assignment and the outcome, the PSMATCH procedure creates an output data set according to the specified method. The output data set then serves as input for an appropriate outcome analysis procedure, and many other SAS/STAT procedures can be used for this purpose. The following three sections describe the three propensity score methods and corresponding output data sets.

Propensity Score Weighting

Propensity score weighting creates weights that are appropriate for estimating the ATE and ATT. The PSMATCH procedure provides the following methods for weighting observations:

- inverse probability of treatment weighting, which is used to estimate the ATE
- ATT weighting (also called weighting by odds), which is used to estimate the ATT

You can save the input observations with their weights in an output data set. For the subsequent outcome analysis, you can use the same analysis on this output data set (augmented with responses) as you would have used on the original data set to estimate the treatment effect.

Propensity Score Stratification

Propensity score stratification divides the observations into strata that have similar propensity scores, with the objective of balancing the observed variables between treated and control units within each stratum. The treatment effect can then be estimated by combining stratum-specific estimates of treatment effect. Rosenbaum and Rubin (1984, p. 521) show that an adjusted estimate of this type that is based on five strata can remove approximately 90% of the bias in the crude or unadjusted estimate.

You can save the strata identification in the output data set. For the subsequent outcome analysis, you can estimate the treatment effect within each stratum (augmented with responses) and combine these estimates across strata to estimate the treatment effect (Stuart 2010, pp. 13–14).

Propensity Score Matching

Propensity score matching matches observations in the control group to observations in the treated group such that the effect of the treatment can be estimated from the resulting matched sample. The PSMATCH procedure provides the following matching strategies:

- greedy nearest neighbor matching, which sequentially and without replacement selects the control unit whose propensity score is closest to that of the given treated unit
- optimal matching, which selects all matches simultaneously and without replacement to minimize the total absolute difference in propensity score across all matches. This approach includes the following optimal matching methods:
  - fixed ratio matching, which matches a fixed number of control units to each treated unit
  - variable ratio matching, which matches one or more control units to each treated unit
  - full matching, which matches each treated unit to one or more control units and matches each control unit to one or more treated units
- matching with replacement, which selects with replacement the control unit whose propensity score is closest to that of each treated unit

Matching can be based on the difference in the logit of the propensity score (LPS) in addition to the difference in the propensity score (PS). Furthermore, matching can be based on Mahalanobis distance that is computed from a set of continuous covariates (possibly including LPS and LS).

You can save only the matching observations in the output data set. For the subsequent outcome analysis, you can use the same analysis on the matched data set (augmented with responses) as you would have used on the original data set to estimate the treatment effect (Ho et al. 2007, p. 233).

Example 1: Optimal Matching

At the completion of a school year, a school administrator asks whether taking a music class causes an improvement in the grade point averages (GPAs) of students. The reasoning behind this question is that learning to read and perform music might improve general reading ability, concentration, and memory.

The data set **School** (which is available at the end of the school year) contains the following variables about students:

- **StudentID**, the student identification number
- **Music**, the music class indicator
- **Gender**, the gender of the student
- **Absence**, the percentage of absences
- **Gpa**, the grade point average

In this data set, only three covariates (**Music**, **Gender**, and **Absence**) are used to illustrate the PSMATCH procedure, but in practice a propensity score analysis often involves many more covariates. Figure 1 lists the first eight observations.

Figure 1 School Data Set

First 8 Observations of the Input School Data Set

Obs	StudentID	Music	Gender	Absence	Gpa
1	18	No	Female	3.71200	3.13894
2	61	No	Male	2.07552	3.31835
3	95	No	Female	2.53865	3.31256
4	41	No	Male	3.00637	3.13982
5	19	Yes	Female	0.08081	4.35331
6	51	No	Female	1.20229	3.87599
7	110	No	Male	2.20710	3.21211
8	87	No	Female	2.30150	3.27578

The outcome data (GPAs) for the students happen to be available at the time of the propensity score analysis, but the recommended practice is not to use the outcome values in the propensity score analysis (Stuart 2010, p. 2). Instead, the response variable is added to the output data set that the PSMATCH procedure creates, and that output data set is subsequently used in an outcome analysis.

First, the following TTEST procedure is used to perform a simple analysis for the effect of the music class:

```
proc ttest data=School;
   class Music;
   var Gpa;
run;
```

The *t* test for difference table in Figure 2 shows that the effect of the music class on GPA is significant.

Figure 2 *t* Test for Difference

The TTEST Procedure

Variable: Gpa

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	198	-3.44	0.0007
Satterthwaite	Unequal	148.13	-3.86	0.0002

Although the test shows a significant effect of the music class, the effect might be related to the student's gender or absence record. The following regression analysis is performed to control for the gender and absence effects:

```
proc glm data=School;
   class music(ref='No') gender;
   model gpa= music gender absence / solution;
run;
```

The parameter estimates table in Figure 3 shows that the effect of the music class on GPA has a *p*-value of 0.1075, which is not as significant as in the previous analysis.

Figure 3 Music Class Effect Estimate

The GLM Procedure

Dependent Variable: Gpa

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		3.902969836	B	0.03911525	99.78	<.0001
Music	Yes	0.066797980	B	0.04131117	1.62	0.1075
Music	No	0.000000000	B	.	.	.
Gender	Female	0.059642566	B	0.03698803	1.61	0.1085
Gender	Male	0.000000000	B	.	.	.
Absence		-0.152941168		0.01483172	-10.31	<.0001

However, the regression adjustment requires a sufficient covariate overlap between the distribution of covariates for the group of students who take the music class and the group of students who do not take the class. You can perform a propensity score analysis that uses a greedy nearest neighbor matching method to ensure a sufficient covariate overlap between the distribution of covariates for the two groups of students.

The following statements invoke the PSMATCH procedure and request optimal matching to match students in the treated group (those who took music) with students in the control group (those who did not take music):

```
ods graphics on;
proc psmatch data=School region=cs;
   class Music Gender;
   psmodel Music(Treated='Yes')= Gender Absence;
   match method=optimal(k=1) exact=Gender stat=lps caliper=0.25;
   assess ps var=(Gender Absence) / plots=all weight=none;
   output out(obs=match)=OutEx1 matchid=_MatchID;
run;
```

The **School** data set contains the student information. The REGION= option specifies an interval region of propensity scores such that only observations that have propensity scores in the region are used in the analysis. Because the MATCH statement is also specified, the REGION=CS option requests that only observations that have propensity scores in the common support region be used for matching. The lower endpoint of the common support region is the larger of the minimum propensity score for the treated group and the minimum propensity score for the control group; the upper endpoint is the smaller of the maximum propensity score for the treated group and the maximum propensity score for the control group. By default, the region is extended by 0.25 times a pooled estimate of the common standard deviation of the logit of the propensity score.

The CLASS statement specifies the classification variables. The PSMODEL statement specifies the logistic regression model that creates the propensity score for each student, which is the probability that the student enrolled in the music class. The **Music** variable is the binary treatment indicator variable, and TREATED='Yes' identifies Yes as the treated group. The **Gender** and **Absence** variables are included in the model because they are believed to be related to enrolling in the music class.

The MATCH statement requests matching and specifies the criteria for matching. The STAT=LPS option (which is the default) requests that the logit of the propensity score be used in computing differences between pairs of observations. The METHOD=OPTIMAL(K=1) option (which is the default) requests optimal matching of one control unit to each unit in the treated group in order to minimize the total within-pair difference, The EXACT=GENDER option forces the treated unit and its matched control unit to have the same value of the **Gender** variable.

The CALIPER=0.25 option specifies the caliper requirement for matching. This option specifies that in order for a match to be made, the difference in the logits of the propensity scores for pairs of individuals from the two groups must be less than or equal to 0.25 times the pooled estimate of the common standard deviation of the logits of the propensity scores.

Although this example illustrates only an optimal one-to-one matching, other matching methods (such as a greedy nearest neighbor matching) can also be specified in the procedure with a similar syntax.

The "Data Information" table in Figure 4 displays information about the input and output data sets, the numbers of observations in the treated and control groups, the lower and upper limits for the propensity score support region, and the numbers of observations in the treated and control groups that fall within the support region. Because REGION=CS is specified, not all 140 observations in the control group fall within the support region.

Figure 4 Data Information

The PSMATCH Procedure

Data Information	
Data Set	WORK.SCHOOL
Output Data Set	WORK.OUTEX1
Treatment Variable	Music
Treatment Group	Yes
All Obs (Treatment)	60
All Obs (Control)	140
Support Region	Extended Common Support
Lower PS Support	0.079975
Upper PS Support	0.529621
Support Region Obs (Treatment)	60
Support Region Obs (Control)	132

The "Propensity Score Information" table in Figure 5 displays summary statistics by treatment group for all observations, for the support region observations, and for the matched observations.

Figure 5 Propensity Score Information

Propensity Score Information

	Treated (Music = Yes)					Control (Music = No)				
Observations	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
All	60	0.347143	0.096184	0.092831	0.490191	140	0.279796	0.124997	0.026465	0.488875
Region	60	0.347143	0.096184	0.092831	0.490191	132	0.294024	0.113997	0.083346	0.488875
Matched	60	0.347143	0.096184	0.092831	0.490191	60	0.348211	0.098808	0.092963	0.488875

The "Matching Information" table in Figure 6 displays the matching criteria, the number of matched sets, the numbers of matched observations in the treated and control groups, and the total absolute difference in the logit of the propensity score for all matches.

Figure 6 Matching Information

Matching Information	
Difference Statistic	Logit of Propensity Score
Method	Optimal Fixed Ratio Matching
Control/Treated Ratio	1
Caliper (Logit PS)	0.163128
Matched Sets	60
Matched Obs (Treated)	60
Matched Obs (Control)	60
Total Absolute Difference	2.113146

The ASSESS statement produces tables and plots that summarize differences in the specified variables between treated and control groups for all observations, for the support region observations, and for the matched observations. You can use these results to assess how well matching achieves a balance in the distributions of these variables. The balance assessment is illustrated in Example 2.

If you are not satisfied with the variable balance, you can use a different propensity score analysis (such as choosing a different set of variables to fit the propensity score model) until you are satisfied. If you are satisfied with the variable balance, you can output matched observations so that they can be used for subsequent outcome analysis.

The OUT(OBS=MATCH)=OutEx1 option in the OUTPUT statement creates an output data set, **OutEx1**, that contains the matched observations. Figure 7 displays the eight observations in **OutEx1** that have lowest propensity scores.

Figure 7 Output Data Set with Matching Numbers

Obs	StudentID	Music	Gender	Absence	Gpa	_PS_	_MATCHWGT_	_MatchID
1	33	Yes	Female	3.50313	3.86316	0.09283	1	1
2	82	No	Female	3.50036	3.20905	0.09296	1	1
3	67	Yes	Female	2.71352	3.80732	0.13790	1	2
4	95	No	Female	2.53865	3.31256	0.15009	1	2
5	47	No	Female	2.49866	3.39185	0.15300	1	3
6	4	Yes	Female	2.49425	3.78497	0.15333	1	3
7	37	No	Male	2.93955	3.89897	0.15549	1	4
8	152	Yes	Male	2.88102	3.73670	0.15988	1	4

By default, the output data set includes the variable **_PS_** (which provides the propensity score) and the variable **_MATCHWGT_** (which provides matched observation weights). The MATCHID=_MatchID option creates a variable named **_MatchID** that identifies the matched sets of observations.

If you assume that no other confounding variables are associated with both the GPA and the music class indicator **Music**, you can add the GPAs for the students to the data set **OutEx1** and perform an outcome analysis of GPA on this data set to estimate the effect of the music class.

For example, because the outcome variable **Gpa** is also available, you can use the following statements to estimate the effect of the music class:

```
proc ttest data=OutEx1;
   class Music;
   var Gpa;
run;
```

The *t* test for difference table in Figure 8 shows that the effect of the music class is not significant.

Figure 8 *t* Test for Difference

The TTEST Procedure

Variable: Gpa

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	118	-0.31	0.7556
Satterthwaite	Unequal	117.86	-0.31	0.7556

The effect of the music class on GPA is not significant according to the propensity score analysis, which contradicts the earlier conclusion that the effect is significant.

You can perform the following regression analysis to further control for the gender and absence effects:

```
proc glm data=OutEx1;
   class music(ref='No') gender;
   model gpa= music gender absence / solution;
run;
```

The "Parameter Estimates" table in Figure 9 shows that the effect of the music class also has a large *p*-value of 0.7246.

Figure 9 Music Class Effect Estimate

The GLM Procedure

Dependent Variable: Gpa

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		3.940060720	B	0.04584263	85.95	<.0001
Music	Yes	0.015668870	B	0.04437397	0.35	0.7246
Music	No	0.000000000	B	.	.	.
Gender	Female	0.034153752	B	0.04447264	0.77	0.4441
Gender	Male	0.000000000	B	.	.	.
Absence		-0.121616920		0.02692121	-4.52	<.0001

Example 2: Variable Balance Assessment

This example builds on Example 1 to illustrate variable balance assessment in a propensity score analysis. For your convenience, the following statements are copied from Example 1:

```
proc psmatch data=School region=cs;
   class Music Gender;
   psmodel Music(Treated='Yes')= Gender Absence;
   match method=optimal(k=1) exact=Gender stat=lps caliper=0.25;
   assess ps var=(Gender Absence) / plots=all weight=none;
   output out(obs=match)=OutEx1 matchid=_MatchID;
run;
```

The ASSESS statement produces tables and plots that summarize differences in the specified variables between treated and control groups for all observations, for the support region observations, and for the matched observations. You can use these results to assess how well matching achieves a balance in the distributions of these variables. As requested by the PS and VAR= options, the variables are the propensity score and the covariates **Gender** and **Absence**. The WEIGHT=NONE option suppresses the display of differences for weighted matched observations. For a matching of one control unit to each treated unit, the weights are all 1 for matched treated and control units, and the results are identical for the weighted matched observations and the matched observations.

The "Standardized Variable Differences" table in Figure 10 displays standardized differences between the treated and control groups, which are computed for all observations, support region observations, and matched observations. For the binary classification variable (**Gender**), the computed difference is the difference in the proportion of the first ordered level (Female) between the treated and control groups.

Figure 10 Standardized Differences

The PSMATCH Procedure

Standardized Variable Differences (Treated - Control)

	Mean Difference				Standardized Mean Difference			Percent Reduction	
Variable	All Obs	Region Obs	Matched Obs	Divisor	All Obs	Region Obs	Matched Obs	Region Obs	Matched Obs
PS	0.067347	0.053119	-0.001068	0.111525	0.603876	0.476301	-0.009579	21.13	98.41
Absence	-0.697568	-0.485721	0.006177	1.136945	-0.613546	-0.427216	0.005433	30.37	99.11
Gender	-0.045238	-0.034848	0	0.496344	-0.091143	-0.070210	0	22.97	100.00

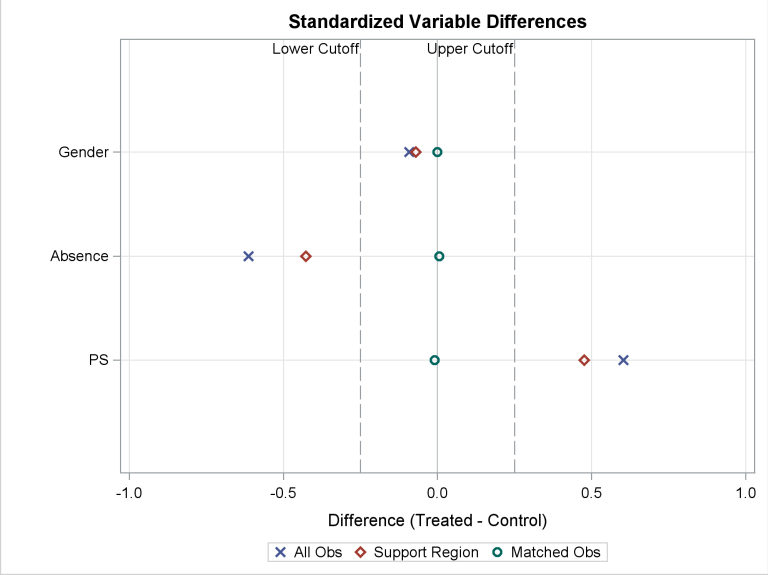
Standardized Variable Differences (Treated - Control)

	Variance Ratio		
Variable	All Obs	Region Obs	Matched Obs
PS	0.5921	0.7119	0.9476
Absence	0.3550	0.5560	0.9618
Gender	1.0208	1.0144	1.0000

The standardized mean differences are significantly reduced in the matched observations, and the largest of these differences is 0.0096 in absolute value, which is less than the recommended upper limit of 0.25 (Rubin 2001, p. 174; Stuart 2010, p. 11). The variance ratios between the two groups are between 0.9476 and 1 for all variables in the matched observations, which is within the recommended range of 0.5 to 2. Because both EXACT=GENDER and METHOD=OPTIMAL are specified in the MATCH statement, the standardized difference for **Gender** is 0 in the matched observations.

By default, the PSMATCH procedure displays a standardized variable differences plot for the variables that are specified in the ASSESS statement, as shown in Figure 11.

Figure 11 Standardized Differences Plot

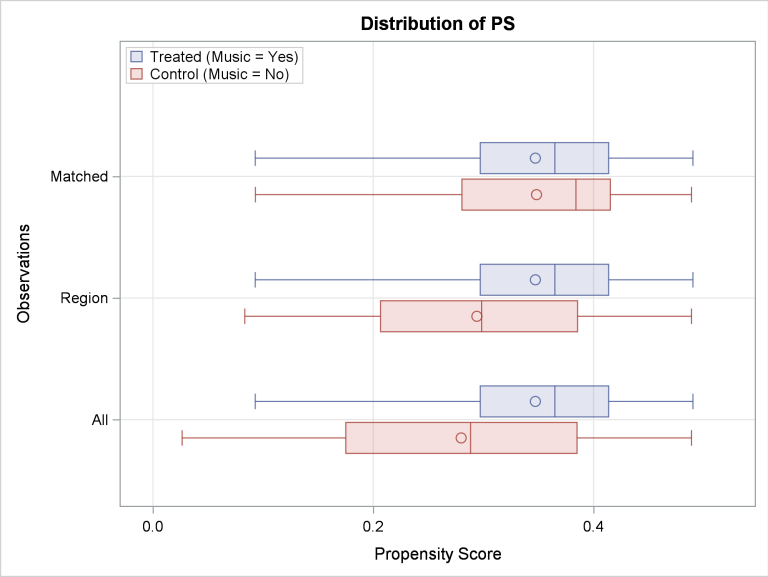


The standardized variable differences plot displays the standardized differences that are shown in the “Standardized Variable Differences” table in Figure 10. All differences for the matched observations are within the recommended limits of –0.25 and 0.25, which are indicated by reference lines.

When you specify PLOTS=ALL, the PSMATCH procedure uses ODS Graphics to create all applicable plots for variables in the ASSESS statement for assessment. For this paper, only a propensity score (PS) box plot, a PS cloud plot, and a bar chart of **Gender** are displayed.

Figure 12 displays PS box plots that compare the propensity score distributions for units in the treated and control groups, based on all observations, observations in the support region, and matched observations.

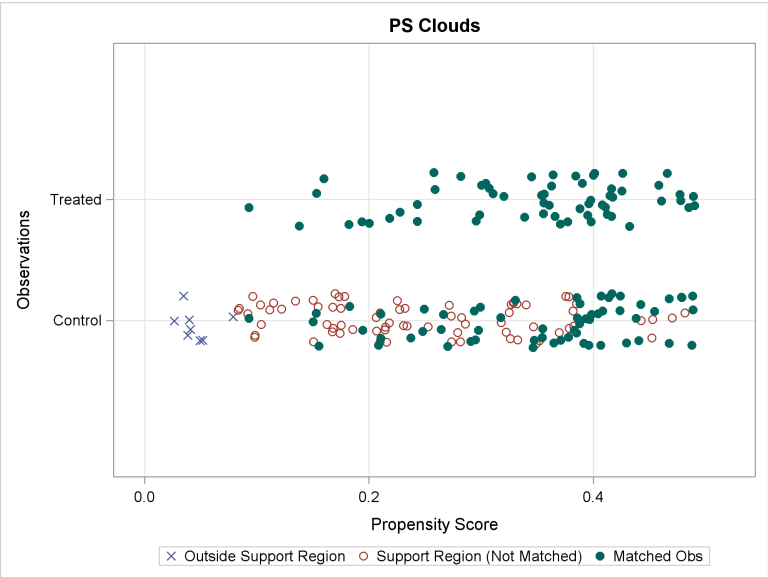
Figure 12 PS Box Plots



The PS box plots show side-by-side box-and-whiskers plots of propensity scores for units in the treated and control groups. Each box-and-whiskers plot displays the mean (as a circle), quartiles (as vertical lines), and minimum and maximum observations for a group. The two distributions are well-balanced for matched observations.

Figure 13 displays a PS cloud plot that compares the values of the propensity score for observations in the treated and control groups, based on all observations, observations in the support region, and matched observations.

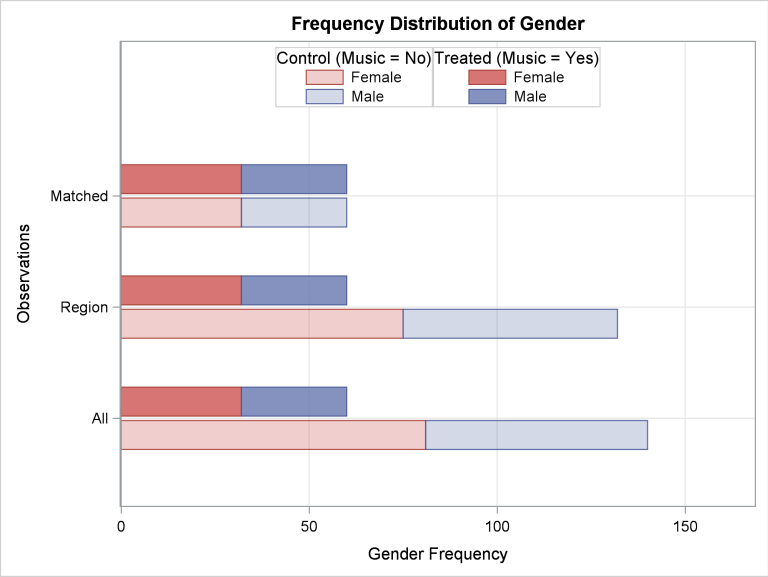
Figure 13 PS Cloud Plot



The points are jittered in the vertical direction to avoid overlap. In the control group, you can see that some observations are not in the support region and some observations in the support region are not matched.

Figure 14 displays a **Gender** bar chart that compares the distributions of **Gender** for units in the treated and control groups, based on all observations, observations in the support region, and matched observations.

Figure 14 Gender Bar Chart



Note that with the one-to-one matching, the two distributions are identical for matched observations.

When you are satisfied with the variable balance of the propensity score analysis, you can output matched observations so that they can be used for subsequent outcome analysis.

Example 3: Propensity Score Weighting

This example creates observation weights for patients in the trial of a propensity score analysis.

A pharmaceutical company is conducting a nonrandomized clinical trial to demonstrate the efficacy of a new treatment (Drug_X) by comparing it to an existing treatment (Drug_A). Patients in the trial can choose the treatment that they prefer; otherwise, physicians assign each patient to a treatment. The possibility of treatment selection bias is a concern because it can lead to systematic differences in the distributions of the baseline variables in the two groups, resulting in a biased estimate of treatment effect.

The data set **Drugs** contains the following baseline variable measurements for individuals from both treated and control groups:

- **PatientID**, the patient identification number
- **Drug**, the treatment group indicator
- **Gender**, the gender
- **Age**, the age
- **Bmi**, the body mass index (a measure of body fat based on height and weight)

In this data set, only a few variables are used to illustrate the PSMATCH procedure. Typically, more variables are used in a propensity score analysis. Figure 15 lists the first eight observations.

Figure 15 Drugs Data Set

First 8 Observations of the Input Drugs Data Set

Obs	PatientID	Drug	Gender	Age	Bmi
1	284	Drug_X	Male	29	22.02
2	201	Drug_A	Male	45	26.68
3	147	Drug_A	Male	42	21.84
4	307	Drug_X	Male	38	22.71
5	433	Drug_A	Male	31	22.76
6	435	Drug_A	Male	43	26.86
7	159	Drug_A	Female	45	25.47
8	368	Drug_A	Female	49	24.28

Note that the **Drugs** data set does not contain a response variable, because the response variable is not available yet. Instead, the response variable is added to the output data set for an subsequent outcome analysis.

The following statements invoke the PSMATCH procedure and create observation weights that are appropriate for estimating the ATT:

```
proc psmatch data=drugs region=allobs(psmin=0.05 psmax=0.95);
   class Drug Gender;
   psmodel Drug(Treated='Drug_X')= Gender Age Bmi;
   assess lps var=(Gender Age Bmi)
          / weight=attwgt;
   output out(obs=region)=OutEx3;
run;
```

The **Drugs** data set contains the patient information. The CLASS statement specifies the classification variables. The PSMODEL statement specifies the logistic regression model that creates the propensity score for each observation, which is the probability that the patient receives Drug_X. The **Drug** variable is the binary treatment indicator variable and TREATED='Drug_X' identifies Drug_X as the treated group. The **Gender**, **Age**, and **Bmi** variables are included in the model because they are believed to be related to the assignment.

The REGION=ALLOBS(PSMIN=0.05 PSMAX=0,95) option selects only observations whose propensity scores are between 0.05 and 0.95 are in the support region, and the OUT(OBS=REGION)= option in the OUTPUT statement includes only these observations in the output data set.

The “Data Information” table in Figure 16 displays information about the input and output data sets, the numbers of observations in the treated and control groups, the lower and upper limits for the propensity score support region, and the numbers of observations in the treated and control groups that fall within the support region. Because REGION=ALLOBS(PSMIN=0.05 PSMAX=0.95) is specified, not all 373 observations in the control group fall within the support region.

Figure 16 Data Information

The PSMATCH Procedure

Data Information	
Data Set	WORK.DRUGS
Output Data Set	WORK.OUTEX3
Treatment Variable	Drug
Treatment Group	Drug_X
All Obs (Treatment)	113
All Obs (Control)	373
Support Region	PS Bounded Obs
Lower PS Support	0.05
Upper PS Support	0.685757
Support Region Obs (Treatment)	113
Support Region Obs (Control)	353

The "Propensity Score Information" table in Figure 17 displays summary statistics by treatment group for all observations and for the support region observations.

Figure 17 Propensity Score Information

Propensity Score Information

	Treated (Drug = Drug_X)					Control (Drug = Drug_A)				
Observations	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
All	113	0.310773	0.132467	0.060231	0.641148	373	0.208801	0.131969	0.020157	0.685757
Region	113	0.310773	0.132467	0.060231	0.641148	353	0.218410	0.129132	0.050185	0.685757

The ASSESS statement produces the tables and plots that summarize differences in the specified variables between treated and control groups for all observations and for the support region observations. As requested by the LPS and VAR= options, the variables listed in the table are the logit of propensity score and the variables **Gender**, **Age**, and **Bmi**. The WEIGHT=ATTWGT option also summarizes differences in the specified VAR= variables between the treated and control groups for the weighted observations.

The "Standardized Variable Differences" table, as shown in Figure 18, displays standardized differences between the treated and control groups for all observations, the support region observations, and the weighted support region observations.

Figure 18 Standardized Differences

The PSMATCH Procedure

Standardized Variable Differences (Treated - Control)

	Mean Difference				Standardized Mean Difference			Percent Reduction	
Variable	**All Obs**	**Region Obs**	**ATT Weighted Region Obs**	**Divisor**	**All Obs**	**Region Obs**	**ATT Weighted Region Obs**	**Region Obs**	**ATT Weighted Region Obs**
LPS	0.639971	0.543499		0.767448	0.833894	0.708190		15.07	
Age	-4.095091	-3.506130	0.536029	6.079104	-0.673634	-0.576751	0.088176	14.38	86.91
Bmi	0.739296	0.618147	-0.128157	1.923178	0.384414	0.321420	-0.066638	16.39	82.66
Gender	-0.024817	-0.016797	-0.008696	0.496925	-0.049941	-0.033801	-0.017500	32.32	64.96

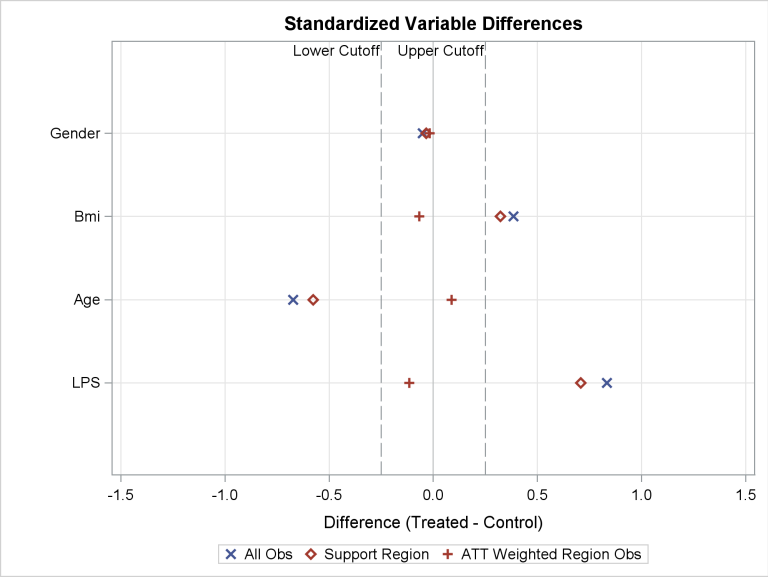
Standardized Variable Differences (Treated - Control)

	Variance Ratio		
Variable	**All Obs**	**Region Obs**	**ATT Weighted Region Obs**
LPS	0.6517	0.8066	
Age	0.7076	0.7897	0.8180
Bmi	0.8854	0.9183	0.9256
Gender	0.9892	0.9921	0.9956

The standardized mean differences are significantly reduced in the weighted support region observations; the largest difference in absolute value is 0.0882, which is still less than the recommended upper limit of 0.25 (Rubin 2001, p. 174; Stuart 2010, p. 11). The variance ratios between the two groups are within the recommended range of 0.5 to 2.

The PSMATCH procedure displays a standardized variable differences plot for the variables that are specified in the ASSESS statement, as shown in Figure 19.

Figure 19 Standardized Differences Plot



The standardized variable differences plot displays the standardized differences that are shown in the "Standardized Variable Differences" table in Figure 18. As expected, all differences for the matched observations are within the recommended limits of –0.25 and 0.25, which are indicated by reference lines.

If you are not satisfied with the variable balance, you can use a different propensity score analysis (such as choosing a method of matching) until you are satisfied. If you are satisfied with the variable balance, you can output all observations (including added observation weights) so that they can be used for subsequent weighted outcome analysis.

The OUT(OBS=REGION)=OutEx3 option in the OUTPUT statement creates an output data set, **OutEx3**, that contains all observations in the support region. Figure 20 displays the first eight observations in **OutEx3**.

Figure 20 Output Data Set with PS Weights

Obs	PatientID	Drug	Gender	Age	Bmi	_PS_	_ATTWGT_
1	284	Drug_X	Male	29	22.02	0.36444	1.00000
2	201	Drug_A	Male	45	26.68	0.22296	0.28694
3	147	Drug_A	Male	42	21.84	0.11323	0.12768
4	307	Drug_X	Male	38	22.71	0.19733	1.00000
5	433	Drug_A	Male	31	22.76	0.35311	0.54586
6	435	Drug_A	Male	43	26.86	0.27263	0.37482
7	159	Drug_A	Female	45	25.47	0.14911	0.17523
8	368	Drug_A	Female	49	24.28	0.07780	0.08437

By default, the output data set includes the variable **_PS_** (which provides the propensity score) and the variable **_ATTWGT_** (which provides observation weights to estimate ATT). The weight for each treated unit is 1, and the weight for each control unit is computed as p / (1 – p), where p is the propensity score.

If you assume that no other confounding variables are associated with both the response variable and the treatment group indicator **Drug**, then after the responses for the trial are observed for the individuals in the support region and added to the data set **OutEx3**, you can use the same outcome analysis with weights on this output data set as you would have used on the original data set **Drugs** (augmented with responses) to estimate the treatment effect.

Summary

In an observational study, propensity score analysis attempts to replicate the properties of a randomized trial to estimate the treatment effect. To this end, the PSMATCH procedure provides methods of weighting, stratification, and matching in propensity score analysis. The matching methods include greedy nearest neighbor matching, matching with replacement, and optimal matching.

The procedure also assesses covariate balance by comparing distributions between the adjusted treated and control groups. When you are satisfied with the variable balance, you can save the output data set that contains a sample of observations that has been adjusted by weighting, stratification, or matching. If you can assume that all important confounding variables have been included in the analysis (weighting, stratification, or matching) and the covariate balance results are satisfactory, you can perform the outcome analysis on the output data set (with the response variable added).

REFERENCES

Austin, P. C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46:399–424.

Guo, S., and Fraser, M. W. (2015). *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Ho, D., Imai, K., King, G., and Stuart, E. A. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15:199–236.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York: Springer-Verlag.

Rosenbaum, P. R., and Rubin, D. B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41–55.

Rosenbaum, P. R., and Rubin, D. B. (1984). "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:516–524.

Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688–701.

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.

Stuart, E. A. (2010). "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25:1–21.

Acknowledgment

The authors are grateful to Bob Rodriguez and Anne Baxter of the Advanced Analytics Division at SAS for their valuable assistance in the preparation of this paper.

Contact Information

Your comments and questions are valued and encouraged. Contact the author:

Yang Yuan
SAS Institute Inc.
111 Rockville Pike, Suite 1000
Rockville, MD 20850
240-618-1055
Yang.Yuan@sas.com