Application of Propensity Score Matching in Observational Studies Using SAS

Yinghui (Delian) Duan, M.Sc., Ph.D candidate

Department of Community Medicine and Health Care,
University of Connecticut Health Center

Connecticut Institute for Clinical and Translational Science (CICATS)

Email: yduan@uchc.edu
RCTs & Observational Studies
Randomized Control Trials (RCTs)

› Treatment assignment is randomized
› Pre-treatment characteristics are balanced, no confounding effects
› Difference in post-treatment outcomes can be attributed to treatment effects
› “Gold standard” to estimate the effects of treatment, interventions, and exposures
Observational Studies

› Non-experimental
› Treatment assignment is not determined by design
› Usually the “treated” and “untreated” are systematically different in some characteristics that can affect outcome of interest (i.e. confounders)
› Difficult to conclude causal effects due to confounders
Propensity Score Method

A useful tool to control confounding effects in observational studies
Propensity Score (PS)

- Defined by Rosenbaum & Rubin in 1983: the probability of treatment assignment conditional on observed baseline covariates

\[ PS_i = Pr (\text{Treatment}_i = 1 \mid X_i) \]

- A useful tool to remove confounding effects and enhance causal inference in observational studies
Estimating PS

- PS is most often estimated by a logistic regression model.
- Can also be estimated using other methods, e.g., bagging or boosting, recursive partitioning or tree-based methods, random forests, and neural networks.
- No significant advantages reported compared to logistic regression model.
Estimating PS in SAS
Example dataset: 6-month Mortality after Percutaneous Coronary Intervention (PCI)

- Study sample: patients who received PCI
- Treatments: usual care alone vs. usual care + a blood thinner
- Baseline confounders: age, gender, height, coronary stent placement, acute myocardial infarction within 7 days, and diabetes
- Outcome: 6-month mortality (0 or 1)
Example dataset: 6-month Mortality after Percutaneous Coronary Intervention (PCI)

› Table: Sample description

<table>
<thead>
<tr>
<th></th>
<th>Usual care alone (N = 2,830)</th>
<th>Usual care + blood thinner (N = 2,332)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>64.6 ± 4.2</td>
<td>62.0 ± 3.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female</td>
<td>938 33.1</td>
<td>760 32.6</td>
<td>0.673</td>
</tr>
<tr>
<td>Height (mean ± SD)</td>
<td>172.4 ± 10.2</td>
<td>171.6 ± 9.5</td>
<td>0.002</td>
</tr>
<tr>
<td>Stent</td>
<td>1,794 63.4</td>
<td>1,611 69.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>659 23.3</td>
<td>438 18.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Acute MI</td>
<td>193 6.8</td>
<td>356 15.3</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
PROC LOGISTIC DATA = new DESC;
CLASS trtm female diabetic stent acutemi;
MODEL trtm = age height female diabetic stent acutemi;
OUTPUT OUT=new_ps PROB = prob;
RUN;

PROB = PREDICTED = PRED = P
PROC LOGISTIC DATA = new DESC;
CLASS trtm female diabetic stent acutemi;
MODEL trtm = age height female diabetic stent acutemi;
OUTPUT OUT=new_ps PROB = prob;
RUN;

Snapshot of output dataset “new_ps”

<table>
<thead>
<tr>
<th>Obs</th>
<th>mort6mo</th>
<th>trtm</th>
<th>stent</th>
<th>height</th>
<th>female</th>
<th>diabetic</th>
<th>acutemi</th>
<th>age</th>
<th><em>LEVEL</em></th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>159</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>66</td>
<td>1</td>
<td>0.37115</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>156</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>66</td>
<td>1</td>
<td>0.40767</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>159</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>67</td>
<td>1</td>
<td>0.35918</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>157</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>68</td>
<td>1</td>
<td>0.32239</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>156</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>63</td>
<td>1</td>
<td>0.51497</td>
</tr>
</tbody>
</table>
Remove Confounding Effects using PS
Two Important Assumptions

› The assignment of treatment is independent of potential outcomes conditional on the observed baseline covariates
› Every subject has a nonzero probability to receive either treatment
Four Methods

- PS matching – most widely used
- Stratification using PS
Stratification using PS

<table>
<thead>
<tr>
<th>PS</th>
<th>Treated</th>
<th>Untreated</th>
<th>Strata</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.4</td>
<td>U</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>T</td>
<td>U</td>
<td>2</td>
</tr>
<tr>
<td>0.6</td>
<td>T</td>
<td>U</td>
<td>3</td>
</tr>
<tr>
<td>0.7</td>
<td>T</td>
<td>U</td>
<td>4</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Trimming
Four Methods

- **PS matching** – most widely used
- Stratification using PS
- Weighting adjustment, e.g., Inverse probability of treatment weighting (IPTW) using PS
- Covariate adjustment using PS – not recommended
Propensity Score Matching

To form matched sets of treated and untreated subjects who share a similar value of PS.
Common Support

Frequency

Untreated

Region of Common Support

Treated

Propensity Score
Four Methods – Common Support

› **PS matching** – ✔

› Stratification – only when used together with trimming

› IPTW – not explicitly examine common support

› Covariate adjustment – not explicitly examine common support
PS Matching

Some decisions to be made:

1:1 or N:1 matching

- N:1 can improve efficiency, reduce variance, but increase bias

With or without replacement

- With-replacement may yield less bias, but higher variance

Which algorithm?
PSM Algorithms: Nearest-Neighbor

› Each treated will get a match, even if it isn’t a very good one
› Will create problem when a treated subject just doesn’t have any controls with similar PS
› If there are multiple untreated subjects with the same PS value as the treated subject, randomly select one
PSM Algorithms: Match within Caliper

› Caliper: limit matches to be within some range of PS values
  › 0.2 of the standard deviation of the logit of the PS (Austin, 2011)
  › 0.25 or 0.5 of the PS standard deviation
PSM Algorithms: **Greedy vs. Optimal**

<table>
<thead>
<tr>
<th>Treated</th>
<th>Untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>PS</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>112</td>
<td>0.43</td>
</tr>
<tr>
<td>113</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

› Overall absolute distance = 0.01 + 0.03 = 0.04
PSM Algorithms: Greedy vs. Optimal

<table>
<thead>
<tr>
<th>Treated</th>
<th>Untreated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>PS</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>0.43</td>
</tr>
<tr>
<td>103</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

› Overall absolute distance = 0.02 + 0.01 = 0.03
PSM Algorithms: Greedy vs. Optimal

- Often does not make huge difference
- Generate the same results if matching with replacement
PSM Example

A macro performing N:1 match on propensity score
Performing a 1:N Case-Control Match on Propensity Score

Lori S. Parsons, Ovation Research Group, Seattle, Washington

› N:1 match
› Matching iterations are from 8-digit to 1-digit
  › E.g., in the 3rd iteration, 6-digit matching,
    PS = 0.12345698 is matched with PS = 0.12345605
All macro variables are required except and **SiteN**

- **Lib** has to be specified even if it’s “work” (otherwise error will occur)
- If **SiteN** is specified, then subjects will be matched within each site
These statements can be modified or removed to change matching precision.
Run Matching for the Example Dataset

NOTE: There were 0 observations read from the data set WORK.MATCH8.
NOTE: There were 0 observations read from the data set WORK.MATCH7.
NOTE: There were 20 observations read from the data set WORK.MATCH6.
NOTE: There were 152 observations read from the data set WORK.MATCH5.
NOTE: There were 1094 observations read from the data set WORK.MATCH4.
NOTE: There were 1828 observations read from the data set WORK.MATCH3.
NOTE: There were 502 observations read from the data set WORK.MATCH2.
NOTE: There were 42 observations read from the data set WORK.MATCH1.
NOTE: The data set WORK.MATCHED1 has 3638 observations and 12 variables.
NOTE: DATA statement used (Total process time):
    real time    0.01 seconds
    cpu time    0.00 seconds
Examine balance after PS Matching

<table>
<thead>
<tr>
<th></th>
<th>Usual care alone (N = 1,819)</th>
<th>Usual care + blood thinner (N = 1,819)</th>
<th>Standardized Mean Difference p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>62.7 ± 3.6</td>
<td>62.8 ± 3.5</td>
<td>0.818</td>
</tr>
<tr>
<td>Female</td>
<td>599 32.9</td>
<td>615 33.8</td>
<td>0.574</td>
</tr>
<tr>
<td>Height (mean ± SD)</td>
<td>171.9 ± 10.2</td>
<td>171.8 ± 9.5</td>
<td>0.7511</td>
</tr>
<tr>
<td>Stent</td>
<td>1,203 66.1</td>
<td>1,214 66.7</td>
<td>0.699</td>
</tr>
<tr>
<td>Diabetes</td>
<td>373 20.5</td>
<td>371 20.4</td>
<td>0.935</td>
</tr>
<tr>
<td>Acute MI</td>
<td>174 9.6</td>
<td>182 10.0</td>
<td>0.655</td>
</tr>
</tbody>
</table>

- *P-value* can be misleading, especially in large sample and with many confounders
- Standardized mean difference < 10
Standardized Mean Difference

› For continuous variables:

\[ d = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \times 100 \]

› For categorical variables:

\[ d = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)}{2}}} \times 100 \]

› ± Sign does not matter
Another Example

Matching using specified caliper = 0.2 of SD of logit of PS
Calculate Logit of PS

```plaintext
PROC LOGISTIC DATA = eg.new DESC;
   CLASS trtm female diabetic stent acutemi;
   MODEL trtm = age height female diabetic stent acutemi;
   OUTPUT OUT=new_ps PROB = prob XBETA = logit;
RUN;
```

Calculate SD of Logit of PS

![The MEANS Procedure]

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis Variable : logit Value of the Linear Predictor</td>
<td>5162</td>
<td>-0.2159201</td>
<td>0.7918129</td>
<td>-3.4224572</td>
<td>2.9084397</td>
</tr>
</tbody>
</table>

$0.2 \times SD = 0.158$
An Introduction to Implementing Propensity Score Matching With SAS®
Kathy Hardis Fraeman, United BioSource Corporation, Bethesda, MD

- pat_dsn  = The name of the SAS data set with the treated patients
- pat_idvar = The name of the patient ID variable in PAT_DSN, can be character or numeric
- pat_psvar = The name of the propensity score probability variable in PAT_DSN
- cntl_dsn  = The name of the SAS data set with the untreated patients
- cntl_idvar = The name of the patient ID variable in CNTL_DSN, can be character or numeric
- cntl_psvar = The name of the propensity score probability variable in CNTL_DSN
- match_dsn  = The name of the output SAS data set with the patient IDs for the matched pairs
- match_ratio = The matching ratio, must be a number from 1 to N for N:1 control to patient matching
- score_diff = A number between 0 and 1 that gives the allowable absolute difference the treated and control patients' matched propensity scores.
- seed  = An optional input parameter, which is the seed for the random number generator
Estimating Treatment Effect in Matched Sample
Estimating Treatment Effects

- Run the same outcome analyses you would have done on the original data
  - Double robust: regression adjustment for confounders can reduce residual effects, increase precision
- If matching done with replacement, need to use weight to reflect the fact that controls used more than once
Some Considerations
 › PS model:
   › Non-parsimonious model to estimate PS
   › Include covariates that are associated with outcome, or with both outcome and treatment; do **NOT** include covariates that are strongly correlated with treatment, but not directly associated with outcome
   › Can include interaction terms and higher order to improve PS estimation and matching
Sample size
- At least 1,000 – 1,500 (Shadish 2013)

Missing data
- List-wise deletion
Thanks!

Questions?
Comments?

Further questions: yduan@uchc.edu
References:

Overview/tutorial of Propensity Score method:
Others:
Materials from Other Presentations:
References (cont.):

Macros for propensity score matching: