

A review of propensity score: principles, methods and application in Stata

Alessandra Grotta and Rino Bellocco

Department of Statistics and Quantitative Methods
University of Milano–Bicocca

&

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

Italian Stata Users Group Meeting - Milano, 13 November
2014

Theoretical background

Application in Stata

The central role of the propensity score in observational studies for causal effects

BY PAUL R. ROSENBAUM

Departments of Statistics and Human Oncology, University of Wisconsin, Madison, Wisconsin, U.S.A.

AND DONALD B. RUBIN

University of Chicago, Chicago, Illinois, U.S.A.

SUMMARY

The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. Both large and small sample theory show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates. Applications include: (i) matched sampling on the univariate propensity score, which is a generalization of discriminant matching, (ii) multivariate adjustment by subclassification on the propensity score where the same subclasses are used to estimate treatment effects for all outcome variables and in all subpopulations, and (iii) visual representation of multivariate covariance adjustment by a two-dimensional plot.

Causal inference framework

ID	T	Y
1	0	21
2	1	31
...
n	1	15

$$T \rightarrow Y$$

Potential outcomes

ID	T	Y	Y(t=0)	Y(t=1)
1	0	21	21	22
2	1	31	16	31
...
n	1	15	15	15

$$T \rightarrow Y$$

Potential outcomes

ID	T	Y	Y(0)	Y(1)
1	0	21	21	22
2	1	31	16	31
...
n	1	15	15	15

$$T \rightarrow Y$$

Potential outcomes

ID	T	Y	Y(0)	Y(1)
1	0	21	21	22
2	1	31	16	31
...
n	1	15	15	15

$$T \rightarrow Y$$

Potential outcomes

ID	T	Y(0)	Y(1)
1	0	21	22
2	1	16	31
...
n	1	15	15

$$T \rightarrow Y$$

Individual treatment effect

ID	T	Y(0)	Y(1)
1	0	21	22
2	1	16	31
...
n	1	15	15

$$\tau_i = Y_i(1) - Y_i(0)$$

Fundamental problem of causal inference

ID	T	Y(0)	Y(1)
1	0	21	.
2	1	.	31
...
n	1	.	15

$$\tau_i = Y_i(1) - Y_i(0)$$

Average treatment effect (ATE)

ID	T	Y(0)	Y(1)
1	0	21	.
2	1	.	31
...
n	1	.	15

$$ATE = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

Average treatment effect among treated (ATT)

ID	T	Y(0)	Y(1)
1	0	21	.
2	1	.	31
...
n	1	.	15

$$ATT = E[Y(1) - Y(0) / T = 1] = E[Y(1) / T = 1] - E[Y(0) / T = 1]$$

Potential outcome means

ID	T	Y(0)	Y(1)
1	0	21	.
2	1	.	31
...
n	1	.	15

$$ATE = E[Y(1)] - E[Y(0)]$$

$$ATT = E[Y(1)/T = 1] - E[Y(0)/T = 1]$$

Potential outcome means

ID	T	Y(0)	Y(1)
1	0	21	.
2	1	.	31
...
n	1	.	15

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

Potential outcome means

ID	T	Y(0)	Y(1)	X ₁	X ₂	...	X _p
1	0	21	.				
2	1	.	31				
...				
n	1	.	15				

$$(Y(1), Y(0)) \not\perp T$$

Potential outcome means

ID	T	Y(0)	Y(1)	X
1	0	21	.	
2	1	.	31	
...	
n	1	.	15	

$(Y(1), Y(0)) \not\perp T$

Strong ignorability assumption

ID	T	Y(0)	Y(1)	X
1	0	21	.	
2	31	.	1	
...	
n	1	.	15	

$$(Y(1), Y(0)) \perp T | \mathbf{X}$$

$$0 < P(T = 1 | \mathbf{X}) < 1$$

Adjusting for X

- ▶ Regression
- ▶ Matching
- ▶ Stratification

ID	T	Y(0)	Y(1)	X
1	0	21	.	
2	1	.	31	
...	
n	1	.	15	

$$(Y(1), Y(0)) \perp T | \mathbf{X}$$

$$0 < P(T = 1 | \mathbf{X}) < 1$$

Then...

ID	T	Y(0)	Y(1)	X	b(X)
1	0	21	.		
2	1	.	31		
...		
n	1	.	15		

$$(Y(1), Y(0)) \perp T | b(\mathbf{X})$$

$$0 < P(T = 1 | b(\mathbf{X})) < 1$$

Balancing score

ID	T	Y(0)	Y(1)	X	b(X)
1	0	21	.		
2	1	.	31		
...		
n	1	.	15		

$$\mathbf{X} \perp T | b(\mathbf{X})$$

Propensity score

ID	T	Y(0)	Y(1)	X	b(X)	e(X)
1	0	21	.			
2	1	.	31			
...			
n	1	.	15			

$$e(\mathbf{X}) = P(T = 1|\mathbf{X})$$

$$\mathbf{X} \perp T | e(\mathbf{X})$$

Propensity score

ID	T	Y(0)	Y(1)	\mathbf{X}	$b(\mathbf{X})$	$e(\mathbf{X})$
1	0	21	.			
2	1	.	31			
...			
n	1	.	15			

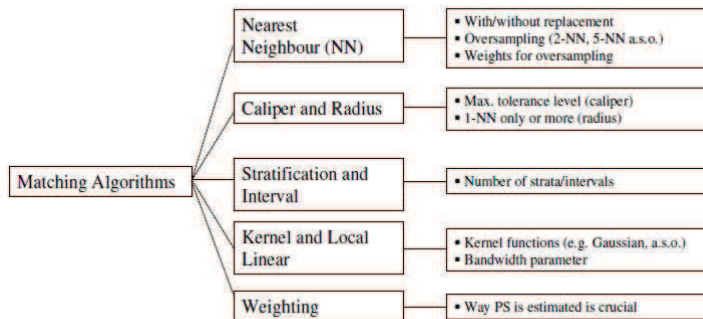
$$e(\mathbf{X}) = P(T = 1|\mathbf{X})$$

Propensity score is the coarsest balancing score: $e(\mathbf{X})=f(b(\mathbf{X}))$

- ▶ Matching
- ▶ Stratification
- ▶ Regression

- ▶ **most popular** propensity score based method
- ▶ we match subjects from the treatment groups by $e(\mathbf{X})$
- ▶ subjects who are unable to be matched are **discarded** from the analysis

Different matching algorithms have been proposed



Some practical guidance for the implementation of propensity score matching (Caliendo, 2005)

Nearest neighbor matching

$$ATT = \frac{1}{N^T} \sum_{i \in T} [Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^C]$$

- ▶ N^T number of treated units
- ▶ $C(i)$ set of controls matched to treated unit i
- ▶ N_i^C number of controls matched to treated unit i
- ▶ $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$; 0, otherwise

- ▶ using $e(\mathbf{x})$, we stratify the entire sample into quantiles
- ▶ within each stratum, we assess the treatment effect
- ▶ we compute an overall treatment effect by averaging the results for each stratum

- ▶ $e(\mathbf{x})$ is included in the **outcome** regression model
- ▶ **with/without** other covariates
- ▶ we assume a **linear** relationship between $e(\mathbf{x})$ and Y

We can estimate propensity score using logistic regression

$$P(T = 1 | X_1, \dots, X_p) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}$$

Adequacy of propensity score model can be checked evaluating the **balance** of X across treatment groups.

$$\mathbf{X} \perp T | \widehat{e(\mathbf{X})}$$

- ▶ Statistic tests
- ▶ Standardized differences
- ▶ Graphical methods

Theoretical background

Application in Stata

Does homelessness affect physical health?

- ▶ **treatment**: homelessness
- ▶ **outcome**: physical health, measured by the Physical Component Summary (PCS) score
- ▶ **confounders**: age, sex, alcohol intake, Mental Component Summary (MCS) score

HELP (Health Evaluation and Linkage to Primary Care) dataset

- ▶ PSCORE (Becker and Ichino, 2002)
- ▶ estimation of propensity score

```
. ssc install pscore  
. help pscore  
  
. pscore exposure varlist  
  
. pscore homeless age gender alc mcs, ///  
logit pscore(mypscore)
```

PSCORE - output

The treatment is homeless

homeless	Freq.	Percent	Cum.
0	244	53.86	53.86
1	209	46.14	100.00
Total	453	100.00	

Logistic regression

Number of obs = 453
LR chi2(4) = 32.75
Prob > chi2 = 0.0000
Pseudo R2 = 0.0524

Log likelihood = -296.26879

homeless	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.013659	.0129651	1.05	0.292	-.0117521 .0390701
female	-.4545305	.2370108	-1.92	0.055	-.9190632 .0100022
alc	.024878	.0057815	4.30	0.000	.0135464 .0362096
mcs	-.0099835	.0077682	-1.29	0.199	-.0252089 .005242
_cons	-.6582588	.518953	-1.27	0.205	-1.675388 .3588703

PSCORE tests the **balancing hypothesis** through this algorithm:

1. Split the sample in k equally spaced intervals of $e(\mathbf{x})$
2. Within each interval test that the average $e(\mathbf{x})$ of treated and untreated do not differ

PSCORE tests the **balancing hypothesis** through this algorithm:

1. Split the sample in k equally spaced intervals of $e(\mathbf{x})$
2. Within each interval test that the average $e(\mathbf{x})$ of treated and untreated do not differ
3. If the test fails, split the interval and test again
4. Continue until, in all intervals, the average $e(\mathbf{x})$ of treated and untreated units do not differ

PSCORE - balance checking

Test for block 3

Two-sample t test with equal variances

```
-----+-----
      Group |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
          0 |      105   .4751954   .0052945   .0542521   .4646963   .4856945
          1 |      113   .4848887   .0054481   .0579139   .4740941   .4956834
-----+-----
combined |      218   .4802199   .0038105   .0562608   .4727097   .4877302
-----+-----
      diff |           -.0096933   .0076152                   -.0247029   .0053163
-----+-----
      diff = mean(0) - mean(1)                                t =   -1.2729
Ho: diff = 0                                                degrees of freedom =    216

      Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
Pr(T < t) = 0.1022              Pr(|T| > |t|) = 0.2044              Pr(T > t) = 0.8978
```

PSCORE tests the **balancing hypothesis** through this algorithm:

1. Split the sample in k equally spaced intervals of $e(\mathbf{x})$
2. Within each interval test that the average $e(\mathbf{x})$ of treated and untreated do not differ
3. If the test fails, split the interval and test again
4. Continue until, in all intervals, the average $e(\mathbf{x})$ of treated and untreated units do not differ
5. Within each interval, test that the means of each characteristic do not differ between treated and untreated

PSCORE - balance checking

Testing the balancing property for variable age in block 3

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	105	36.22857	.7236529	7.415235	34.79354	37.6636
1	113	36.47788	.7777827	8.267943	34.9368	38.01895
combined	218	36.3578	.5317885	7.851764	35.30967	37.40593
diff		-.2493047	1.06662		-2.35162	1.853011

diff = mean(0) - mean(1) t = -0.2337
Ho: diff = 0 degrees of freedom = 216

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.4077 Pr(|T| > |t|) = 0.8154 Pr(T > t) = 0.5923

The balancing property is satisfied

- ▶ PSMATCH2 (Leuven and Sianesi, 2003)
- ▶ propensity score matching

PSMATCH2 - syntax

```
. ssc install psmatch2
. help psmatch2

. psmatch2 depvar [indvar], ///
outcome(varname) pscore(varname)

. psmatch2 homeless, ///
outcome(pcs) pscore(mypscore) neighbor(1)

. psmatch2 homeless age gender alc mcs, ///
outcome(pcs) neighbor(1)
```

PSMATCH2 - output

```
. psmatch2 homeless, outcome(pcs) pscore(mypscore) neighbor(1)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	46.9367801	47.8530158	-.916235746	1.48207493	-0.62

Note: S.E. does not take into account that the propensity score is estimated.

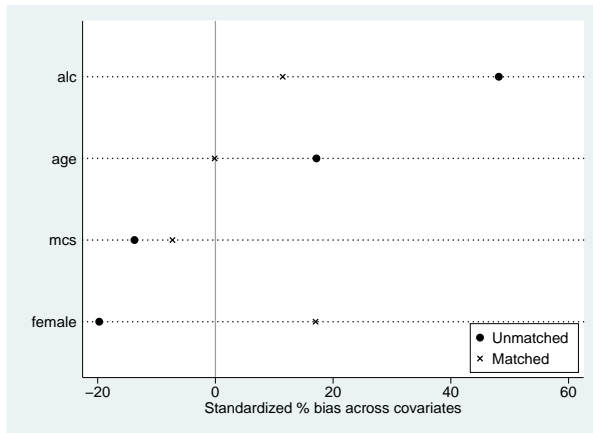
PSTEST assesses balance in the matched samples

```
. pstest age female alc mcs, both graph
```

Variable	Unmatched Matched	Mean		%bias	%reduct bias	t-test		V(T)/ V(C)
		Treated	Control			t	p> t	
age	U	36.368	35.041	17.2		1.83	0.068	1.33*
	M	36.368	36.378	-0.1	99.3	-0.01	0.990	1.24
female	U	.19139	.27459	-19.7		-2.08	0.038	.
	M	.19139	.11962	17.0	13.7	2.03	0.043	.
alc	U	23.038	13.512	48.1		5.19	0.000	2.38*
	M	23.038	20.775	11.4	76.2	1.06	0.291	1.35*
mcs	U	30.731	32.487	-13.7		-1.45	0.147	0.81
	M	30.731	31.663	-7.3	46.9	-0.78	0.435	0.97

* if variance ratio outside [0.76; 1.31] for U and [0.76; 1.31] for M

PSTEST - output



PSMATCH2 - output

```
. psmatch2 homeless, outcome(pcs) pscore(mypscore) radius caliper(0.03)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	47.172174	48.4370461	-1.26487214	1.07716821	-1.17

Note: S.E. does not take into account that the propensity score is estimated.

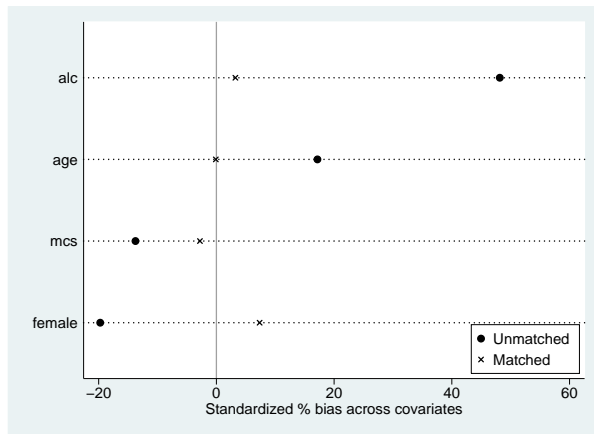
PSTEST assesses balance in the matched samples

```
. pstest age female alc mcs, both graph
```

Variable	Unmatched Matched	Mean		%bias	%reduct bias	t-test		V(T)/ V(C)
		Treated	Control			t	p> t	
age	U	36.368	35.041	17.2		1.83	0.068	1.33*
	M	36.167	36.175	-0.1	99.4	-0.01	0.992	1.25
female	U	.19139	.27459	-19.7		-2.08	0.038	.
	M	.19704	.16609	7.3	62.8	0.81	0.420	.
alc	U	23.038	13.512	48.1		5.19	0.000	2.38*
	M	20.443	19.804	3.2	93.3	0.36	0.716	1.04
mcs	U	30.731	32.487	-13.7		-1.45	0.147	0.81
	M	30.772	31.129	-2.8	79.7	-0.29	0.772	0.88

* if variance ratio outside [0.76; 1.31] for U and [0.76; 1.32] for M

PSTEST - output



PSMATCH2 - output

```
. gen femmcs=female*mcs
. gen femalc=female*alc
. gen alcmcs=alc*mcs

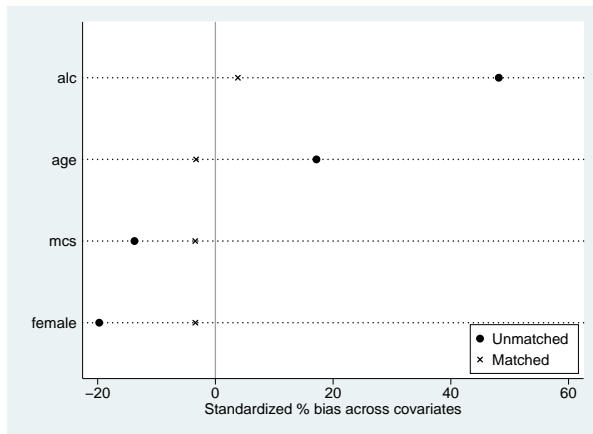
. pscore homeless age female alc mcs femmcs femalc alcmcs, logit pscore(mypscore)

. psmatch2 homeless, outcome(pcs) pscore(mypscore) neighbor(1)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	46.9367801	47.5441296	-.607349496	1.41574607	-0.43

Note: S.E. does not take into account that the propensity score is estimated.

PSTEST - output



PSMATCH2 - output

```
. pscore homeless age female alc mcs femmcs femalc alcmcs, logit pscore(mypscore)
```

```
. psmatch2 homeless, outcome(pcs) pscore(mypscore) neighbor(1)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	46.9367801	47.5441296	-.607349496	1.41574607	-0.43

Note: S.E. does not take into account that the propensity score is estimated.

Standard error:

- ▶ Lechner (2001)
- ▶ Abadie et al. (2004)
- ▶ Abadie and Imbens (2006)

PSMATCH2 - output

```
. pscore homeless age female alc mcs femmcs femalc alcmcs, logit pscore(mypscore)
```

```
. psmatch2 homeless, outcome(pcs) pscore(mypscore) neighbor(1)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	46.9367801	47.5441296	-.607349496	1.41574607	-0.43

Note: S.E. does not take into account that the propensity score is estimated.

Source of variability:

- ▶ propensity score estimation
- ▶ matching on the common support
- ▶ order in which treated individuals are matched

Set of commands to estimate ATE and ATT (ATET) through:

- ▶ regression adjustment
- ▶ inverse probability weighting
- ▶ PS matching
- ▶ ...

TEFFECTS - syntax

```
. teffect psmatch (outcome variable) ///  
(treat treatvarlist [, treatmodel])
```

```
. teffects psmatch (pcs)    ///  
(homeless age female alc mcs  ///  
femalc femmcs alcmcs, logit), atet
```


TEFFECTS - output

```
. teffects psmatch (pcs) (homeless age female alc mcs femalc femmcs alcmcs, logit), atet
```

```
Treatment-effects estimation      Number of obs      =      453
Estimator      : propensity-score matching      Matches: requested =      1
Outcome model  : matching                      min =      1
Treatment model: logit                        max =      1
```

```
-----
            pcs |           Coef.      AI Robust
                   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
ATET
  homeless |
(1 vs 0) | -0.6073495   1.335721   -0.45   0.649   -3.225315   2.010616
-----
```

Standard error:

- ▶ Abadie and Imbens (2012)

TEFFECTS vs. PSMATCH2

```
. teffects psmatch (pcs) (homeless age female alc mcs femalc femmcs alcmcs, logit), atet
```

```
Treatment-effects estimation          Number of obs      =      453
Estimator      : propensity-score matching  Matches: requested =      1
Outcome model  : matching                  min =              1
Treatment model: logit                    max =              1
```

	pcs	Coef.	AI Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ATET							
homeless (1 vs 0)		-.6073495	1.335721	-0.45	0.649	-3.225315	2.010616

```
. psmatch2 homeless, outcome(pcs) pscore(mypscore) neighbor(1)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
pcs	Unmatched	46.9367801	49.000829	-2.06404896	1.0129221	-2.04
	ATT	46.9367801	47.5441296	-.607349496	1.41574607	-0.43

Note: S.E. does not take into account that the propensity score is estimated.

- ▶ ATT* (Becker and Ichino, 2002)
- ▶ PS **matching**: ATTND, ATTNW (nearest neighbor) ATTR (radius), ATTK (kernel)
- ▶ PS **stratification**: ATTS

```
. att* depvar [indvar], ///  
outcome(varname) pscore(varname)  
  
. attnw pcs homeless, pscore(mypscore)
```

ATT* - output

```
. attnw pcs homeless, pscore(mypscore)
```

```
ATT estimation with Nearest Neighbor Matching method  
(equal weights version)  
Analytical standard errors
```

```
-----  
n. treat.   n. contr.       ATT   Std. Err.       t  
-----  
          209           116    -0.607     1.416     -0.429  
-----
```

```
Note: the numbers of treated and controls refer to actual  
nearest neighbor matches
```

- ▶ SENSATT (Nannicini, 2007)
- ▶ after ATT*
- ▶ assesses the robustness of ATT with respect to unmeasured confounding

References I

- ▶ Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 79:516-24.
- ▶ Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 2008; 22(1):31-72.
- ▶ Sweeney LP, Samet JH, Larson MJ, et al. Establishment of a multidisciplinary health evaluation and linkage to primary care (HELP) clinic in a detoxification unit. *J Addict Dis* 2004; 23:33-45.
- ▶ Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2002; 2(4):358-377.
- ▶ Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003.

- ▶ Lechner M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, in: Lechner, M., Pfeiffer, F. (eds), *Econometric Evaluation of Labour Market Policies*, Heidelberg ca/Springer, 2001.
- ▶ Abadie A, et al. Implementing matching estimators for average treatment effects in Stata. *Stata journal* 2004; 4:290-311.
- ▶ Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. *Econometrica* 2006; 74(1):235-267.
- ▶ Abadie A, Imbens GW. Matching on the estimated propensity score. Harvard University and National Bureau of Economic Research. 2012.
- ▶ Nannicini T. A Sensatt: a simulation-based sensitivity analysis for matching estimators. *The Stata Journal* 2007; 7(3):334-350.