

Propensity Score Matching

Vincent Vu and Kenta Yashiro

Trier University

Motivation

Table: New Drug Treatment and Propensity Scores
(Generated Data from R)

ID	Treat.	Age	Income	Educ.	Health Score	Employ.	Pre-existing Condition
1	1	65	30k	HS	60	Retired	Hypertension
2	0	50	50k	BSc	75	Full-time	None
3	1	72	25k	ND	50	Retired	Diabetes
4	0	45	60k	MSc	80	Full-time	None
5	1	68	28k	HS	65	Part-time	Hypertension
6	0	52	32k	HS	62	Part-time	Hypertension
7	1	70	29k	HS	66	Part-time	Hypertension
8	0	54	35k	HS	64	Part-time	Hypertension
9	1	67	27k	ND	55	Retired	Diabetes
10	0	60	26k	ND	52	Retired	Diabetes

We would like to match units to compare the effects of a treatment on their condition. However, matching completely identical individuals is unrealistic due to the many variables.

► Adjustment for confounding in observational studies

- Confounders influence both the outcome as well as the ability/decision to receive treatment, making it difficult to find feasible control groups in many instances.

► Question

- How do we compare the control and treatment groups when they have dissimilar characteristics due to confounders (Dehejia and Wahba, 1999)?

Matching methods

Hansen and Klopfer (2006) provide an outline on drawbacks of the Nearest Neighbor method as well as the alternative stratified methods:

Nearest Neighbor Matching

- Also known as greedy matching, a treated unit is matched with the untreated unit with the closest propensity score
- Repeated without replacement until every treated unit is 1:1 matched to a control unit
- Order in which units are matched can be set by the one performing the match (highest to lowest, lowest to highest, randomized order, etc.)
- As mentioned by Hansen and Klopfer (2006), this method fails to take advantage of populations that are plentiful with untreated units, so some data is lost.

Optimal Full Matching

- Untreated units are matched to treated units with the closest propensity score until no untreated units are unmatched
- Aims to minimize the SDM (standardized mean difference) of covariates between the treated and untreated matches

Subclass/Strata Matching

- Treated population is divided into equally sized subclasses based on their propensity scores
- Untreated population is sorted into subclasses with similar propensity scores

Once a matching method is chosen, covariate balance must be checked to ensure the quality of the match.

Applied Example

ID	Treat.	Age	Income	Educ.	Health Score	Employ.	Pre-existing Condition	Propensity Score
1	1	65	30k	HS	60	Retired	Hypertension	0.85
2	0	50	50k	BSc	75	Full-time	None	0.40
3	1	72	25k	ND	50	Retired	Diabetes	0.90
4	0	45	60k	MSc	80	Full-time	None	0.30
5	1	68	28k	HS	65	Part-time	Hypertension	0.78
6	0	52	32k	HS	62	Part-time	Hypertension	0.78
7	1	70	29k	HS	66	Part-time	Hypertension	0.76
8	0	54	35k	HS	64	Part-time	Hypertension	0.75
9	1	67	27k	ND	55	Retired	Diabetes	0.89
10	0	60	26k	ND	52	Retired	Diabetes	0.88

- The table has been updated with an additional column including a propensity score, a balancing score ranging from 0 to 1 that takes into account all confounders.
- The propensity score represents the likelihood a patient received treatment, adjusting for differences between the treatment and control groups.
- Matching, stratifying, or weighting based on propensity scores enables comparison between groups with similar covariate distributions.

Average Treatment Effect

The treatment effect for each unit is usually obtained using

$$E[Y_1 - Y_0|X]$$

Since only Y_1 or Y_0 is available for each unit, this is generally not feasible.

Griener and Stuart (2023) give three common methods to choose the average effect of treatment:

Average Treatment Effect in the Treated (ATT)

- After matching treated units are assigned a weight of 1, while untreated units are assigned a weight of

$$w_i^u = \frac{b_i}{1 - b_i}$$

- ATT is obtained using $E[Y_1 - Y_0|Z = 1]$

Average Treatment Effect in the Untreated (ATU)

- After matching untreated units are assigned a weight of 1, while untreated units are assigned a weight of

$$w_i^t = \frac{1 - b_i}{b_i}$$

- ATU is obtained using $E[Y_1 - Y_0|Z = 0]$

Average Treatment Effect in the Population (ATE)

- Inverse probability weighting is used after matching, where

$$w_i = \frac{Z}{b_i} + \frac{1 - Z}{1 - b_i}$$

- ATE is obtained using $E[Y_1 - Y_0]$

NOTE: Linear regression for estimating the treatment effect must be used in combination with g-computation (Griener and Stuart, 2023).

Theory

Propensity Score

The propensity score e for the i th unit is a probability measure that the i th unit will receive treatment given a vector of covariates X_i (Rubin, 2001):

$$e_i = e(X_i) = P(Z_i = 1|X_i)$$

The logit model suggested by Rosenbaum and Rubin (1983) follows the conditional likelihood model,

$$e_i = \frac{P(Z = 1)P(X|Z = 1)}{P(Z = 1)P(X|Z = 1) + P(Z = 0)P(X|Z = 0)} \quad (1)$$

While this is sufficient in many cases, a probit model can also fit populations where the propensity score distribution is normal.

Reduction in Bias

Given bias B in the unmatched population and bias B_m in the matched population, Rosenbaum and Rubin (1983) give us the reduction in bias from propensity score matching

$$B - B_m = \int E(\mathbf{X}|b)(P_m(b|Z = 0) - P(b|z = 0)) db$$

Implementation in this project

The Lalonde dataset, provided by the National Supported Work Demonstration (Lalonde, 1986), estimates the impact of attending a work training on the participant's 1978 income using the following covariates:

Covariates	Description
RE74	Unit's income in 1974
RE75	Unit's income in 1975
Race	Black, Hispanic, or White
Age	Unit's age in 1978
Married	1 = married, 0 = unmarried
Educ	Unit's years of education
No Degree	1 = no degree, 0 = degree

The steps taken to simulate a propensity score match in R on the Lalonde dataset were as follows:

1. Calculate the propensity score for each unit using the MatchIt package (fitted with probit link and glm distance).
2. Perform a matching method.
3. Plot the covariate balance.
4. Obtain the ATT using a regression model fitted with the matched data.

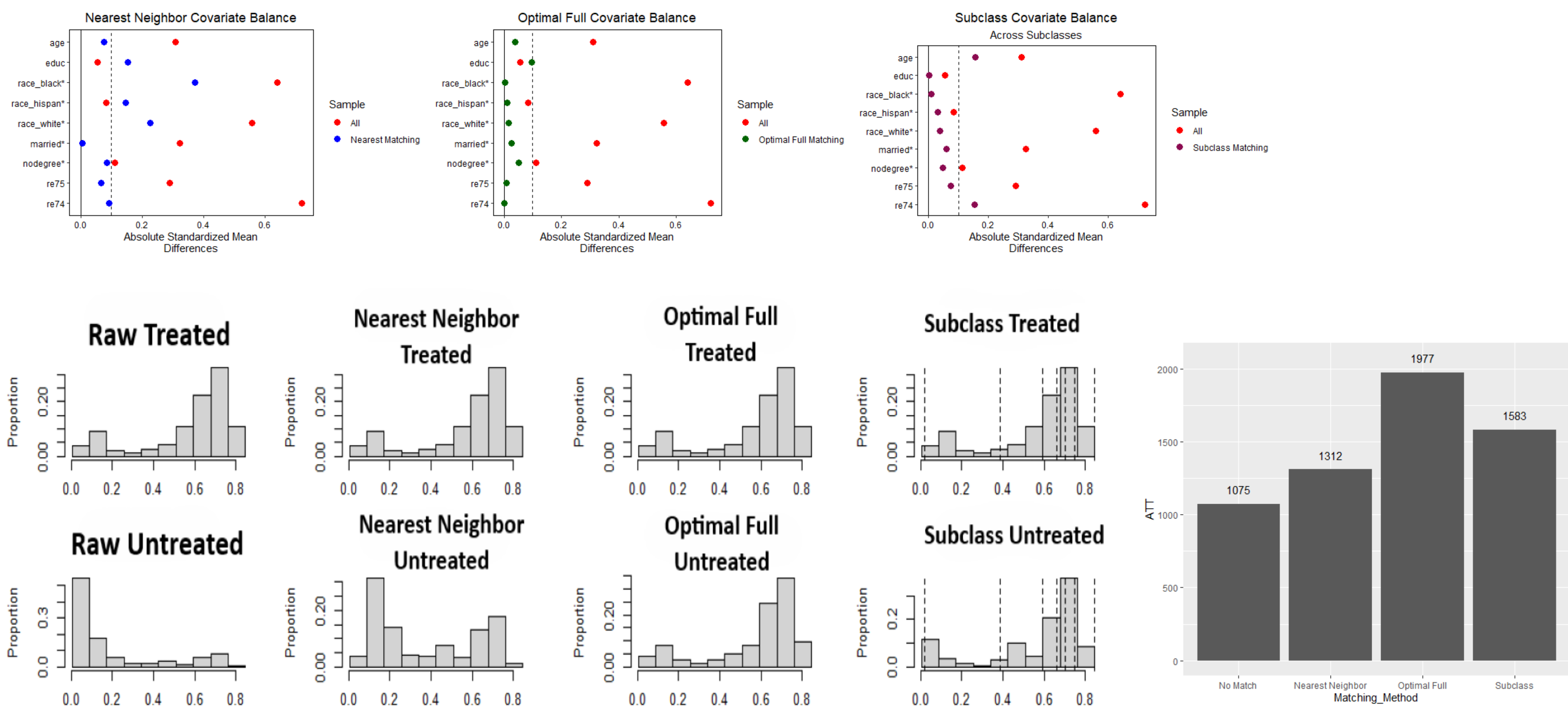
Nearest neighbor, optimal full, and subclassification matching were performed using the MatchIt package (Ho, Imai, King, and Stuart, 2011) in R, which calls functions from the optmatch package (Hansen and Klopfer, 2006).

Conclusion and Summary

- From our simulation, the individuals that attended the training on average received an increase in income between \$1,583 and \$1,977 compared to if they had not attended the job training.
- Optimal Full and Subclassification Matching had the best covariate balance as well as similar distributions between the treated and untreated populations, so we take the ATTs from both methods.
- Stratified methods, while providing a better match, are computationally more expensive than the greedy method.

- Dehejia, R. H., and Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94(448), 1053â€“1062.
- Dehejia, R. H. and Wahba, S. (2002). Propensity Score Matching Methods for Non-Experimental Causal Studies. *Review of Economics and Statistics*, Vol. 84, pp. 151-161.
- Griener, S. and Stuart, E. A. (2023). Choosing the Causal Estimand for Propensity Score Analysis of Observational Studies. <https://arxiv.org/abs/2106.10577>
- Hansen, B. B., Klopfer, S. O. (2006). Optimal Full Matching and Related Designs via Network Flows. *Journal of Computational and Graphical Statistics*, 15(3), 609â€“627.
- Ho, D., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1â€“28.
- Robert Lalonde, "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, Vol. 76 (1986), pp. 604-620.
- Rosenbaum, P. R., and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41â€“55. <https://doi.org/10.2307/2335942>
- Rubin, D. (2001): Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. In: *Health Services & Outcomes Research Methodology*, 2, p. 169-188.

Simulation Results



FOR ADDITIONAL INFORMATION



Universität Trier

Vincent Vu and Kenta Yashiro

Economic and Social Statistics Department

Universitätsring 15

D-54286 Trier

