

Propensity Score Weighting

Walter L. Leite

Research and Evaluation Methodology
Program

University of Florida

How do weights work?

Score	Weight
4	1
2	2
1	4
5	1
2	2

Simple mean:

$$\frac{(4 + 2 + 1 + 5 + 2)}{5} = 2.8$$

Weighted mean:

$$\frac{(4 \times 1) + (2 \times 2) + (1 \times 4) + (5 \times 1) + (2 \times 2)}{10} = 2.1$$

Reasons for weighting

1. Eliminate bias due to unequal probability of selection
2. Compensate for non-response error due to unequal response rates
3. Reduce frame error (unequal coverage of the population by the sampling frame).
4. Improve precision of the estimates through the use of auxiliary information.

Types of Weights

- Sampling weights
- Non-response weights
- Post-stratification weights
- Precision weights
- Propensity score weights

Interpretation of the weight as a balancing adjustment

- Sampling weights are frequencies of each observation in the population.
- Propensity score weights are frequencies of each observation in a pseudo-population where the treatment and control group have equal distributions of covariates.
- If the treatment assignment is random, the distributions of covariates for the treatment group is similar to the distribution for the control group.
- In quasi-experimental designs, propensity score weighting adjusts the distribution of covariates so they are similar across groups.

Reducing bias by weighting based on the propensity score

Two traditions:

- Statistics: Propensity score weighting in cross-sectional studies Rosenbaum (1987)
- Epidemiology: Inverse probability of treatment weighting in longitudinal studies (Robins and Hernan, 2000).

Weight for estimating the ATT

Propensity Score

$$w_i = Z_i + (1 - Z_i) \frac{\hat{e}_i}{1 - \hat{e}_i}$$

Condition: Treated or untreated.

- The weight is 1 for treated individuals and the odds of treatment for untreated individuals.

Summary of ATT weights for career academy example

treat: 0

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.001527	0.032260	0.057220	0.085960	0.099450	2.247000

treat: 1

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	1	1	1	1	1

Weight for estimating the ATE

Condition: Treated or untreated.

$$w_i = \frac{Z_i}{\hat{e}_i} + \frac{1 - Z_i}{1 - \hat{e}_i}$$

Propensity Score

Summary of ATE Weights for Career Academy Example

treat: 0

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.002	1.032	1.057	1.086	1.099	3.247

treat: 1

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	5.036	9.478	14.200	16.640	187.800

Extreme weights

- Common with ATE but rare with ATT weights
- May be caused by propensity score model misspecification or lack of common support
- Result in biased estimates and/or inflated standard errors.

Options to Deal with Extreme Weights

- Re-specify the propensity score model
- Switch to another propensity score method
- Truncate weights
- Use stabilized weights

Weight Truncation

- Assign the weight at a cutoff percentile to observations with weights above the cutoff
- Lee et al. ([2011](#)) found that truncation of weights reduced treatment effect estimate bias and standard errors with weights estimated with propensity scores but not with data mining methods.

Summary of truncated ATE Weights at 99% percentile for Career Academy Example

treat: 0

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.002	1.032	1.057	1.086	1.099	3.247

treat: 1

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	5.036	9.478	11.390	16.640	24.760

Stabilized Weights

- Multiply the weights by constants equal to the expected values of being in the treatment or control group:

Without Sampling Weights

$$c_1 = \frac{\sum_{i=1}^{n_1} \hat{e}_i}{n_1}$$

$$c_0 = \frac{\sum_{i=1}^{n_0} (1 - \hat{e}_i)}{n_0}$$

With Sampling Weights

$$c_1 = \frac{\sum_{i=1}^{n_1} w_{si} \hat{e}_i}{\sum_{i=1}^{n_1} w_{si}}$$

$$c_0 = \frac{\sum_{i=1}^{n_0} w_{si} (1 - \hat{e}_i)}{\sum_{i=1}^{n_0} w_{si}}$$

Summary of stabilized ATE Weights for Career Academy Example

treat: 0

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1584	0.1632	0.1672	0.1717	0.1738	0.5134

treat: 1

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.9202	4.6340	8.7220	13.0700	15.3100	172.8000

Covariate Balance Evaluation

- Comparison between standardized differences between weighted means or proportions of treated and control group for each covariate.
- The *bal.stat* function of the *twang* package computes the weighted and proportions, and standardized differences, using the standard deviation of the treated group rather than the pooled standard deviation.

Covariate balance for English fluency (categorical) and Number of siblings at home (numeric)

	tx.mn	tx.sd	ct.mn	ct.sd	std.eff.sz	stat	p
BYSTLNG2:0	0.017	0.131	0.018	0.133	-0.003	0.158	0.944
BYSTLNG2:1	0.009	0.095	0.013	0.113	-0.040	NA	NA
BYSTLNG2:2	0.119	0.323	0.119	0.323	0.000	NA	NA
BYSTLNG2:3	0.033	0.179	0.031	0.174	0.010	NA	NA
BYSTLNG2:4	0.822	0.383	0.819	0.385	0.006	NA	NA
bysibhom	1.914	1.617	1.899	1.625	0.009	0.214	0.831

Criteria for covariate balance evaluation

Austin (2011): $|d| < 0.1 \text{ SD}$

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.

Stuart (2010): $|d| < 0.25$

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.

What Works Clearinghouse (2013): $|d| < 0.05$ or $|d| < 0.25 + \text{covariate adjustment}$

U.S. Department of Education, Institute of Education Sciences, & What Works Clearinghouse. (2013). *What Works Clearinghouse: Procedures and standards handbook (Version 3.0)*. Washington, DC: Author.

Summary of covariate balance across PS estimation methods for ATT weights

Propensity Score Estimation	Minimum	Mean	Maximum
Logistic regression	0.000	0.010	0.057
Random forest	0.000	0.035	0.157
GBM	0.000	0.033	0.125

Horvitz and Thompson Estimator of the treatment effect (Rosenbaum, 1987)

This estimator is the difference between weighted means:

$$\Delta = \frac{\sum_{i=1}^{n_1} w_{i1} y_{i1}}{\sum_{i=1}^{n_1} w_{i1}} - \frac{\sum_{i=1}^{n_0} w_{i0} y_{i0}}{\sum_{i=1}^{n_0} w_{i0}}$$

Estimation with weighted regression:

$$Y_i = \beta_0 + B_1 T_i + e_i$$

Treatment effect estimates for career academy on income

Weighted means and mean difference:

treat	outcome	se
0	0 4.282430	0.08755116
1	1 4.408251	0.15992330

contrast	SE
0.28878	0.0927

Weighted regression:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.35565	0.04079	82.272	< 2e-16 ***
Treat	0.28878	0.09348	3.089	0.00215 **

Glass Delta = 0.140

Design-based Methods to obtain correct standard errors

- Taylor series linearization
- Replication techniques:
 - Jackknife
 - Bootstrap

Taylor Series Linearization

- Taylor Series Linearization can be used to obtain the variance of statistics which are a non-linear combination of means and totals by finding linear functions that approximate the statistics.
- Linearization depends on finding partial derivatives analytically or numerically.
- The survey package in R obtains variance estimates for many statistics (means, sums, proportions) using linearization.

The Jackknife Repeated replication

1. Divide the full sample into random groups of size 1 or larger.
2. In turn, each group is removed from the full sample in order to create a subsample.
3. Recalculate the weights taking into the account the observations still in the sample (replicate weights).
4. The parameter of interest is estimated with each jackknifed sample, using the replicate weights.
5. The standard deviation of the multiple estimated parameters is the standard error of the estimate.

The bootstrap (more specifically, the nonparametric bootstrap)

1. Take K samples with replacement of the same size of the original sample from the original sample, mimicking the original sample design.
2. Estimate the parameter of interest with each bootstrapped sample.
3. The standard deviation of the multiple estimated parameters is the standard error of the estimate.

Other types of Bootstrap: Residual bootstrap, Parametric bootstrap.

Combining treatment effect estimates from multiple imputed datasets

- Rubin's (1987) rules are used to combine estimates and standard errors:
- Combine estimates by taking the mean:

$$\Delta = \sum \Delta_m / m$$

- Obtain standard errors by combining the within-imputation and between-imputation variances:

$$\text{var}(\Delta)_W = \sum \text{var}(\Delta)_m / m \quad \text{var}(\Delta)_B = \sum (\Delta_m - \Delta)^2 / (m - 1)$$

$$SE(\Delta) = \sqrt{\text{var}(\Delta)_W + (1 + (1 / m)) \text{var}(\Delta)_B}$$

Treatment Effect Estimates of Career Academy on Income with Multiple Imputation

	results	se	(lower	upper)	missInfo
(Intercept)	3.3754822	0.03656966	3.30379330	3.4471711	3 %
treat1	0.2722613	0.09304645	0.08988992	0.4546328	1 %

Doubly-Robust Estimators of Treatment Effects

Doubly-robust estimation consists in removing selection bias due to covariates using both the propensity score model and the outcome model.

The treatment effect estimate will be unbiased if either the propensity score model or the outcome model are correct.

There are several double-robust estimators of treatment effect.

Steps of Doubly-Robust Regression Estimation of Treatment Effects

1. Fit separate weighted regression models to the outcomes of treated and untreated:

$$y_{i1} = \beta_{01} + \beta_{11}X_{i1} + \varepsilon_{i1} \quad y_{i0} = \beta_{00} + \beta_{10}X_{i0} + \varepsilon_{i0}$$

2. For the ATE, obtain predicted outcomes in the presence and absence of the treatment for the entire sample. For the ATT, obtain predicted outcomes for the treated only.
3. Obtain the differences between predicted outcomes in the presence and absence of treatment. The treatment effect is the mean of those differences.

Doubly-robust estimate of the effect of career academy on income

Statistic:

0.3159268

SE:

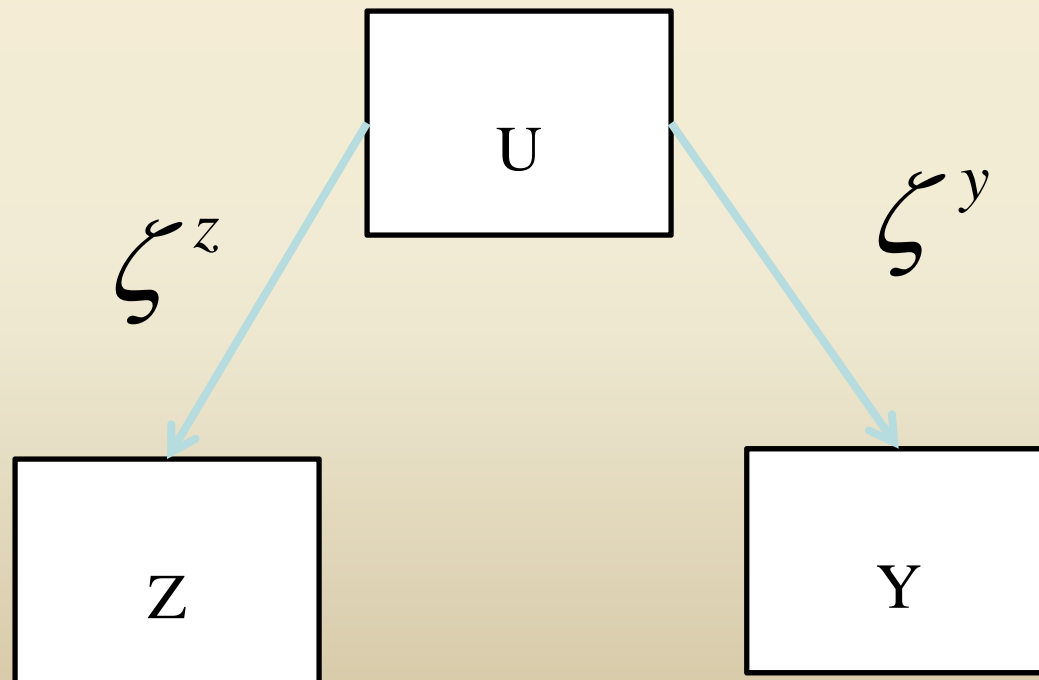
0.0824568

Sensitivity Analysis

- It asks the question: Would the conclusion change if an important covariate was omitted?
- Goals:
 - Determine how strong the effect of an omitted covariate would have to be for the significance test of the treatment effect to change
 - Determine the degree of robustness of treatment effects to hidden bias, which is the part of the selection bias due to omitted confounders.

Sensitivity Parameters

- The sensitivity parameters set the relationships of the confounding variable U with the treatment Z and outcome Y



Steps of Carnegie, Harada and Hill (2016) sensitivity analysis

- 1) Define sensitivity parameters based on the observed data. Multiple values should be selected for both ζ^Z and ζ^Y .
- 2) For each pair of values of ζ^Z and ζ^Y and the vector of observed values of Y, Z, and X, simulate U from the conditional distribution of U given Y, Z, and X.
- 3) Fit the outcome model with Z, X and U as predictors and obtain the treatment effect and its standard error.
- 4) Repeat steps 2 and 3 for k iterations with same pair of sensitivity parameters, and average the treatment effects across iterations. The standard error is obtained by squaring the sum of the within-replication and between-replication variances.
- 5) Select another pair of sensitivity parameters and repeat steps 2 to 4.

Step 1: Define ranges for sensitivity parameters

1.1 - Standardize all continuous observed predictors.

1.2 - Obtain standardized coefficients for the regression of Z on X, and the regression of Y on X.

1.3 - Define multiple values of sensitivity parameters ζ^Z and ζ^Y around the values of observed regression coefficients.

Step 5: Select another pair of sensitivity parameters and repeat steps 2 to 4

Repeating the process for all pairs of sensitivity parameters will result in a grid of treatment effects and standard errors. For example:

	$\zeta^Y:1$	$\zeta^Y:2$	$\zeta^Y:3$
$\zeta^Z:1$	$\tau_{11}(\sigma_{11})$	$\tau_{12}(\sigma_{12})$	$\tau_{13}(\sigma_{13})$
$\zeta^Z:2$	$\tau_{21}(\sigma_{21})$	$\tau_{22}(\sigma_{22})$	$\tau_{23}(\sigma_{23})$
$\zeta^Z:3$	$\tau_{31}(\sigma_{31})$	$\tau_{32}(\sigma_{32})$	$\tau_{33}(\sigma_{33})$

Decisions on sensitivity are based on how large ζ^Z and ζ^Y need to be for the treatment effect to become non-significant.

Example results of sensitivity analysis

Coefficients on U where significance level 0.05 is lost:

Y	Z
0.981	2.109
0.998	0.870
1.033	0.000
1.113	-1.610
1.423	-2.418
1.897	-2.574

R packages useful for propensity score weighting

Package	Function	Objective
survey	sydesign, svymean, svyby, svycontrast, svyglm	Estimate weighted means, proportions and regressions with propensity score weights
Twang	bal.stat	Evaluate covariate balance with propensity score weights
treatSens	treatSens, sensPlot	Obtain a sensitivity analysis