ISSN 2959-6157

IPTW and DR Estimators: Advancing towards the Adoption of Methodologies for Estimating Causal Treatment Effects in Observational Studies

Pengbo Xia[1, *]

[1]Department of Yizhong, High School, Tianjin, 300051, China
*Corresponding author: chrisx@ldy.edu.rs

Abstract:

The propensity score refers to the likelihood of a subject being selected for intervention, given their observed baseline covariates. By assigning different weights to academic disciplines based on the reciprocal probability of receiving treatment, there will be a synthetic sample where the assignment of treatment is not influenced by baseline factors that were measured. Utilizing inverse probability of treatment weighting (IPTW) through the propensity score enables people to assemble unbiased or objective estimates regarding average treatment effects. In order to estimate the causal relationship between an exposure and an outcome, the second notion of doubly robust estimation combines an outcome regression with a propensity score model. Only when the statistical model is appropriately stated can unbiased estimates be obtained using the propensity score approach and outcome regression separately. However, even if only one of the two models is correctly specified, an unbiased estimator for the effect can be obtained by employing the doubly robust estimator. This introduction to inverse probability of treatment weighting and doubly robust estimators includes conceptual overviews, an application to students' performance in high school, as well as discussions based on the project.

Keywords: Average treatment effect; inverse probability of treatment weighting; doubly robust; causal inference.

1. Introduction

Observational data provides ample opportunities to explore the impact of exposure, for instance, when examining the comparative efficacy and security of various therapies. Therefore, the utilization of observational studies is becoming increasingly prevalent among researchers for the estimation of the effect of interventions, exposures, and therapies on the resultant health. The effects of exposure or causality can be expressed using potential outcomes within Rubin's framework for causal modeling [1]. In the presence of two mutually exclusive exposures within a specific time interval, patients possess a set of potential outcomes, with each exposure corresponding to an outcome. However, in reality, patients are allocated to only one exposure, resulting in the observation of only one potential outcome. In this instance, a contrast (such an absolute discrepancy) between the possible results under each exposure might be used to characterize the causative impact for this patient. The average causal effect (ACE) is then calculated by averaging these differences across all patients in a relevant population. Given the existence of a single possible outcome in actuality, it presents difficulties to directly assess the causal impact at an individual patient level. Therefore, most studies focus on estimating ACE instead. Interestingly, ACE can also be defined without explicitly considering individual-level causal effects. For instance, when dealing with binary outcomes, risk differences or ratios may serve as alternative measures of ACE.

In randomized controlled trials, the process of randomization guarantees that there are no systematic differences comparing the measured and unmeasured baseline characteristics of treated and control participants. This ensures a dependable foundation for assessing the efficacy of interventions. However, in non-randomized studies, the presence of treatment-selection bias may arise due to systematic variations between control and treatment participants. Consequently, estimating the effect of treatment solely via contrasting the results of treatment groups is not feasible.

Propensity score techniques are being utilized more frequently to estimate treatment effects from observational data. The propensity score is defined as the probability of treatment assignment conditional on measured baseline covariates [2-4]. Rosenbaum and Rubin showcased a key attribute of the propensity score: when provided with the propensity score, treatment status becomes uncorrelated with measured baseline covariates [2]. Therefore, the propensity score acts as an equalizing measure, ensuring that participants in treatment and control possessing identical propensity scores display similar distributions of baseline factors that were observed.

There are four different approaches discussed in the statistical literature for utilizing the propensity score: covariate adjustment using the propensity score, stratification or subclassification on the propensity score, matching on the propensity score, and inverse probability of treatment weighting (IPTW) [2, 5]. Rubin suggested that propensity score techniques offer the benefit of enabling observational research to be structured in a manner similar to randomized experiments. This approach allows for the separation of study design from the analysis of exposure's impact on the outcome [6].

Moreover, Robins and colleagues introduced the concept of doubly robust estimators (DR) that necessitate a model for estimating both the propensity score and the outcome model within the same estimator. The benefit of these estimators is that they provide researchers two chances to get accurate findings since they provide unbiased estimates of treatment effects even in cases when only one or both of these constituent models are accurately described [7].

This study introduces a more accurate and resilient approach for inferring causality from observational data, along with a novel analytical instrument for researchers in relevant fields. This will not only improve the comprehension of particular treatment impacts but also foster progress in social science and medical research.

2. Estimators

2.1 Framework and Average Treatment Effects

Considering a scenario where there exists a choice between two alternatives and the existence of two potential treatments is assumed. (e.g., active treatment versus control treatment). Within the framework of potential outcomes, it is postulated that each individual possesses a pair of potential outcomes: $Y_i(0)$ and $Y_i(1)$, representing the outcomes under the control treatment and the active treatment, respectively, when administered in identical circumstances. Nevertheless, the assignment of treatment is dichotomous, where each participant is randomly assigned to either the control group or the active treatment group. Let $Y_i(1)$ and $Y_i(0)$ denote the hypothetical posttest outcomes for the treatment and control groups respectively. For each subject, the effect of treatment is defined as $Y_i(1)-Y_i(0)$: the difference between the two potential outcomes. The actual observation of these outcomes totally dependent upon the treatment variable Z_i, such that $Y_i = Z_i Y_i(1)+(1-Z_i)Y_i(0)$, where Y_i represents the observed continuous posttest outcome variable. The Xi vector includes all initial variables, encompassing the measurement of the outcome before the test. The average treatment effect (ATE) is defined to be: $E[Y_i(1)-Y_i(0)]$, with the expectation taken across the population of interest [8]. The ATE represents the mean impact, on a population-wide scale, of transitioning a complete populace from the control group to the treated group. The average causal inference in this context is $\tau = E(Y_i(1) \mid X_i) - E(Y_i(0) \mid X_i)$.

2.2 IPTW Estimator

The propensity score, which indicates the probability of receiving treatment depending on the characteristics of the individuals, is the basis for all of the computations shown in this section, i.e. $P(Z_i = 1 \mid X_i)$. To generate the inverse probability of treatment weighting, the estimated propensity scores $\left(\widehat{p_i}\right)$ are utilized by applying weighting techniques. These scores are estimated values that come from a probit or logistic model.

By applying inverse weights of $1/\left(\widehat{p_i}\right)$ when Z_i equals 1, or $1/(1-\widehat{p_i})$ when Z_i equals 0, the IPTW estimator can be derived.

$$\hat{\tau}_{IPTW} = N^{-1}\sum_{i=1}^{N}\frac{Z_i Y_i}{\widehat{p_i}} - N^{-1}\sum_{i=1}^{N}\frac{(1-Z_i)Y_i}{1-p_i} \tag{1}$$

However, if the model is misspecified, then the weighting will be inappropriate and the IPTW estimator may be biased [9].

2.3 DR Estimator

Compared with doubly robust (DR) estimator, the IPTW estimators are part of a group of unbiased estimation methods. By removing any term with an expected value of zero from the estimation equation, unbiased estimates could still be obtained. This extra term can be utilized to enhance the efficiency of the estimators and safeguard against model misrepresentation. The most efficient estimator in this category is known as the semiparametric efficient estimator, which is demonstrated to be the doubly robust estimator.

Lunceford and Davidian provide the formula for a double-robust estimation [8]:

$$\hat{\tau}_{DR} = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_iY_i-(Z_i-\widehat{p_i})m_1(X_i)}{\widehat{p_i}} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i+(Z_i-\widehat{p_i})m_0(X_i)}{1-\widehat{p_i}} \tag{2}$$

where $m_Z X_i = E(Y_i \mid Z_i = Z, X_i)$ for $Z=0$ or $Z=1$, i.e., these predicted values are derived from separate regressions conducted for each group, using the same model. The regressions include baseline covariates and the initial outcome measure to estimate coefficient values and generate predicted outcomes [10].

The doubly robust estimator becomes the semi-parametric efficient estimator when both models are well-specified. In the case of a correct exposure model, the double-robust estimator demonstrates reduced variance compared to the IPTW estimator. On the contrary, in case the outcome model is precise, it demonstrates increased variability compared to the standard regression model. Nevertheless, it offers safeguard against potential misrepresentation in this specific model. If both exposure and outcome models are misspecified, then resulting estimates may be biased. Notably, because they depend on these models, other techniques like regression or IPTW would likewise show same bias.

3. Methods

3.1 Data Source

The dataset used in this study consists of information from 2,392 students, including details of learning habits and academic achievements (Table 1). This study aims to examine the correlation between academic performance of high school students and their engagement in extracurricular activities, utilizing the provided dataset. The treatment group, also known as the independent variable, refers to the involvement of students in extracurricular activities(Z) (mean= 0.38; standard deviation,0.49). The dependent variable centers around students' GPA(Y) (mean= 1.91; standard deviation, 0.91) as an indicator of their academic performance. Furthermore, the covariate considers the age(X) (mean= 16.5; standard deviation, 1.12) among students. (The Kaggle website offers specifics on the data generation procedure.)

Table 1. Lists of Variables

Variable	Logogram	Meaning
Age	X	The age of the high school students.
Extracurricular	Z	A binary variable indicating whether the student participates in extracurricular activities (1 for yes, 0 for no).
GPA	Y	The GPA of the student, which represents their academic performance.

3.2 Application

Calculating Propensity Scores: The project will utilize a logistic regression model to determine the propensity scores of students, where Extracurricular (Z) is considered as the response variable and Age (X) as the predictor variable. This step aims to estimate the predicted probability of participating in extracurricular activities based on a student's age. Figure 1 is a histogram about propensity score.

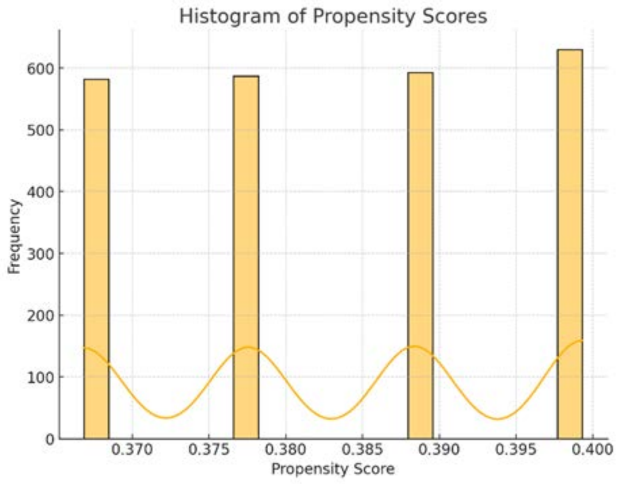

Fig. 1 Histogram of Propensity Scores

Calculating ATE using IPTW and DR Methods: In the IPTW approach, propensity scores are utilized to assign weights to individual observations within the regression command. These assigned weights facilitate the adjustment of observed outcomes, enabling estimation of the Average Treatment Effect. The Doubly Robust approach utilizes both a propensity score model and an outcome regression (logistic regression) model to estimate the Average Treatment Effect, ensuring a more resilient estimation even in cases of potential misspecification in either of the models.

4. Results and Discussion

4.1 ATE Estimation

IPTW Method: The estimated ATE obtained through IPTW approach was 2.277, suggesting that engagement in extracurricular pursuits is linked to an average GPA increment of 2.277.

DR Method: The estimated ATE obtained through DR approach was 3.854, suggesting that engagement in extracurricular pursuits is linked to an average GPA increment of 3.854.

4.2 Data Analysis

The figure (Figure 2) presented below demonstrate the comparison of ATE estimates using various approaches: unadjusted, IPTW and DR. The unadjusted approach fails to consider potential confounding factors, whereas IPTW and DR approaches incorporate the covariate (age) in order to enhance the precision of estimating the causal effect. Furthermore, boxplots (Figure 3) are included to visually represent the distribution of GPA among treatment groups following adjustment.

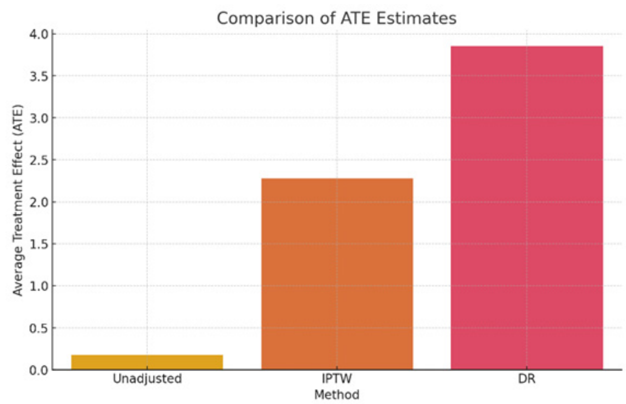

Fig. 2 Comparison of ATE Estimates

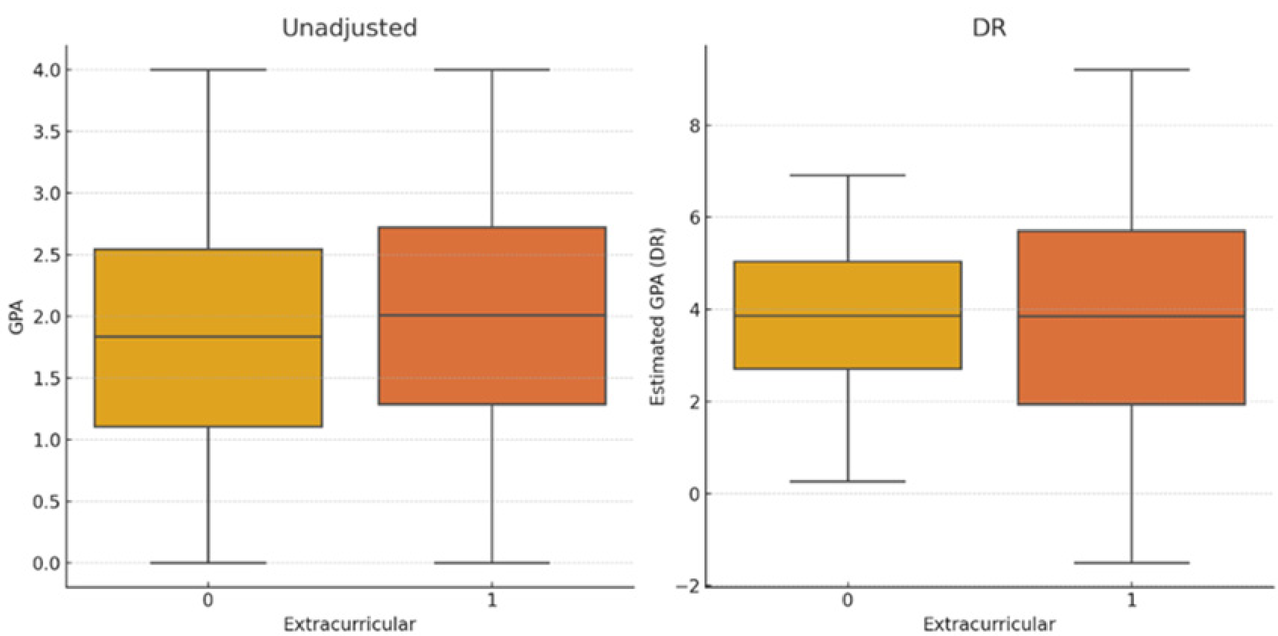

Fig. 3 Comparison of GPA

4.3 Discussion

The histogram reveals a reasonably equitable dispersion of propensity scores, indicating the presence of diversity in the probability of engagement in extracurricular activities among distinct age cohorts.

When it comes to comparison of ATE, the unadjusted ATE is found to be lower compared to both the IPTW and DR estimates, indicating that not considering factors like age may result in underestimating the impact of extracurricular activities on GPA. It is worth noting that utilizing the DR method, which incorporates propensity score and outcome models, produces the highest ATE estimate, suggesting a strong positive influence of extracurricular activities. Moreover, the unadjusted GPA boxplot indicates a noticeable difference between students involved in extracurricular activities and those who are not. However, the gap becomes more prominent in the DR-adjusted GPA distribution, thereby reinforcing the conclusion that engagement in extracurricular activities has a positive impact on academic performance when considering age.

Suggested approaches for tackling this phenomenon are presented herein. Encouraging diverse involvement in extracurricular pursuits is crucial due to empirical evidence indicating a noteworthy positive link between students' engagement in these activities and their academic success. Education policymakers and school administrators should prioritize promoting such involvement as it not only enhances scholastic accomplishments but also fosters comprehensive growth.

The analysis of this study is subject to certain limitations. Although efforts have been made to improve the inference of causal relationships through methodologies in the project, it is important to acknowledge the presence of potential confounding factors, such as study habits, or parental involvement, that could influence the association between extracurricular activities and GPA. Future studies may consider employing more sophisticated statistical techniques or experimental designs in order to further validate this relationship. Furthermore, it is important to note that the sample used in this study may not fully represent the entire population of high school students. As a result, there is a possibility that the research findings may have limited applicability. To improve the generalizability and relevance of these results, future investigations could be conducted across various geographical regions and diverse educational institutions.

5. Conclusion

The data unequivocally shows that there is a substantial and direct correlation between students' GPA and extracurricular activity involvement. By utilizing IPTW and DR estimators, the study confirms the significance of the relationship between extracurricular activity involvement and GPA. These approaches integrate propensity scores and outcome models, offering more rigorous statistical evidence to establish causal connections. Importantly, the doubly robust (DR) method highlights an even more significant impact, emphasizing the importance of incorporating both propensity scores and outcome models in studies that aim to establish causal relationships. These findings emphasize the necessity of promoting extracurricular activities as a means to enhance academic performance among high school students, which could be explained by the following facts. Firstly, effective time management and a strong sense of responsibility can be developed through participation in extracurricular activities. These skills can then be applied to academic pursuits, allowing students to efficiently allocate their study time and improve their GPA. Additionally, engaging in extracurricular activities that require problem-solving and critical thinking can enhance cognitive abilities, which are essential for better understanding and absorption of course content. Furthermore, teamwork and social interaction in extracurricular activities contribute significantly to the development of social skills and a cooperative mindset among students, which are equally important within an academic setting during collaborative projects and discussions. Moreover, participating in extracurricular activities provides stress relief for students while also improving their mental well-being. This positive state of mental health helps students maintain focus on their studies, ultimately enhancing their academic performance.

Even though this study shows a correlation between participation in extracurricular activities and scholastic performance, more investigation is required to pinpoint the exact causal mechanisms. Subsequent investigations could provide a more holistic evaluation of how engagement in extracurricular activities influences students' scholastic performance, encompassing factors related to cognition, emotions, and social interactions. The implementation of these measures will enhance the education system's efficiency and promote students' comprehensive growth.

References

[1] Sekhon J. The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. The Oxford Handbook of Political Methodology. Oxford: Oxford University Press, 2009.

[2] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects. Biometrika, 1983, 70: 41-55.

[3] Rosenbaum P R, Rubin D B. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 1984, 79: 516-524.

[4] Austin P C. An introduction to propensity-score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 2011, 46: 399-424.

[5] Rosenbaum P R. Model-based direct adjustment. Journal of the American Statistical Association, 1987, 82: 387-394.

[6] Rubin D B. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Services & Outcomes Research Methodology, 2001, 2: 169-188.

[7] Rubin D B. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 1974, 66: 688-701.

[8] Imbens G W. Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics, 2004, 86: 4-29.

[9] Michele Jonsson Funk, Daniel Westreich, et al. Doubly Robust Estimation of Causal Effects, American Journal of Epidemiology, 2011, 761-767.

[10] Emsley R, Lunt M, Pickles A, Dunn G. Implementing Double-robust Estimators of Causal Effects. The Stata Journal, 2008, 8(3): 334-353.