IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 006, LIR475
DOI: https://doi.org/10.52731/lir.v006.475

Evaluating the Effects of Pre-University Education Using Propensity Score Matching

Keita Nishiyama *

Abstract

This study uses propensity score matching to explore the causal relationship between pre-university education and first-year GPA. Given the challenges of conducting randomized controlled trials in educational settings due to ethical concerns and practical limitations, observational studies often become necessary. Propensity scores, initially proposed by Rosenbaum and Rubin, enable a more reliable estimation of causal effects by simulating an experimental framework in observational data. This method adjusts for covariate distributions between treatment and control groups, allowing for statistically comparable conditions. The findings support that pre-university education significantly improves first-year GPA, demonstrating the method's effectiveness in educational research and highlighting its potential for broader application across various academic disciplines.

Keywords: Causal Inference, Pre-University Education, Propensity Score Matching

1 Introduction

The evaluation of the effects of university education is conducted through various methods. Common approaches include visualizing primary data on dashboards and observing longitudinal changes in metrics such as GPA. However, the most crucial aspect in evaluating university education effects is how students change due to educational interventions and whether these changes are beneficial [1]. In natural sciences, particularly in basic medical research, experimental studies are conducted.

Randomized Controlled Trials (RCT) is a research design that randomly assigns participants to an experimental group or a control group to evaluate the causal relationship of an intervention by comparing the outcomes between these groups. This method minimizes selection bias and provides high reliability in estimating causal relationships. Generally, conducting RCTs in educational research takes a lot of work. Experiments involving children or students often raise ethical issues and create difficulties in establishing conditions that produce high-quality research outcomes. On the other hand, there are "observational studies," which do not involve researchers manipulating independent variables (factors or conditions). Observational studies are frequently conducted in the social sciences, where pure experimental research through random allocation is often tricky. However, observational studies carry the risk of issues in the distribution of covariates (confounders) that may affect the dependent variables. To control for the effects of covariates, statistical analysis methods such as analysis of covariance and panel data analysis have

* Senshu University Kanagawa, Japan

been developed and used. However, these methods have limitations, such as mathematical constraints and the absence of longitudinal data for some subjects.

In recent years (past 20 years), research has been actively conducted on methods using propensity scores to adjust for covariates, circumventing these limitations and gaining attention. Usually, in observational data, it is only possible to observe the presence or absence of an intervention for the same individual. Methods using propensity scores enable inference using a counterfactual assumption: "if an intervention had been administered or not." Propensity scores, proposed by Rosenbaum and Rubin[2], theoretically enable observational data to approximate experimental research. These scores are calculated from a model that estimates the probability of an individual receiving an intervention based only on observed covariates. Using these scores, it is possible to effectively adjust for differences in covariate distributions between treatment and control groups, making the groups statistically comparable. Specific methods include propensity score matching, stratification based on propensity scores, and covariate adjustment, achieving balance between groups and enabling more accurate estimation of intervention effects.

The advantage of this method lies in significantly reducing the issues of selection bias and confounding factors in observational studies. Moreover, analyses using propensity scores effectively clarify the effects of specific interventions in naturally occurring situations and are an essential tool for analyzing complex causal relationships in reality. Thus, they are used across various fields, including medicine, psychology, and economics. The Rubin Causal Model (RCM), proposed by Donald Rubin, is a framework for statistical causal inference that uses the concept of "potential outcomes" to estimate the effects of treatments on individuals. Specifically, it considers the outcomes of individuals under both the presence and absence of an intervention, treating the unobserved outcome as a potential outcome [3].

In the Rubin Causal Model, each subject's treatment effect (ITE) differs between the outcomes with and without the treatment. However, since it is impossible to simultaneously observe both states in the same individual, this "counterfactual" approach is used to infer causal relationships. By employing propensity scores, it is possible to balance potential outcomes in the dataset of an observational study, allowing for more reliable causal inference. The Rubin Causal Model is widely used in medicine, social sciences, and economics and is an essential framework for accurately estimating causal relationships from observational data [4][5][6].

At Senshu University's Faculty of Network Information, pre-university education is conducted for students admitted through recommendation-based admission exams. Specifically, tasks in general (academic skills), English, and mathematics are assigned before enrollment, and scoring, feedback, and advice are provided. The Faculty of Network Information requested this pre-university education to the IR Department to verify the program's effectiveness. This study introduces a case where the IR department effectively evaluated pre-university education using related and implementation data. The details of the datasets used are described later, but all these data are observational, and thus, the treatment effect could not be evaluated within an RCT framework. Therefore, the study was conducted using propensity scores. This study aimed to clarify the causal relationship between pre-university education and first-year GPA using propensity scores.

2 Material and Methods

The data for pre-university education includes information from students who took recommendation-based admission exams and enrolled in the Faculty of Network Information at Senshu University between 2020 and 2023. Table 1 summarizes this data.

Table 1: Summary of Pre-University Education

Variable	N = 361[1]
Entrance Examination System	
Admission Office	33 (9.1%)
Affiliated Schools	91 (25%)
Designated schools	170 (47%)
Partner Schools A	45 (12%)
Partner Schools B	14 (3.9%)
Sports	8 (2.2%)
Entrance Year	
2020	81 (22%)
2021	94 (26%)
2022	84 (23%)
2023	102 (28%)
[1] n (%)	

Variable	Academic Skills, N = 361[1]	English, N = 361[1]	Mathematics, N = 361[1]
Letter Grades			
A	145 (40%)	57 (21%)	255 (94%)
B	137 (38%)	124 (46%)	6 (2.2%)
C	58 (16%)	53 (20%)	4 (1.5%)
D	16 (4.4%)	17 (6.3%)	2 (0.7%)
F	5 (1.4%)	19 (7.0%)	3 (1.1%)
(NA)	0	91	91
[1] n (%)			

The recommendation-based admission exams included are AO (Admissions Office), affiliated school recommendation, designated school recommendation, sports recommendation, designated educational collaboration recommendation, and educational exchange partnership recommendation. Pre-university tasks included general (academic skills), mathematics, and English, each evaluated with letter grades A to D and F according to their engagement. Additionally, students admitted through affiliated school recommendations only undertook the general (academic skills) task, not mathematics or English tasks.

This study aims to estimate the causal effects of educational interventions through propensity score matching, including other students from the same faculty in the dataset. Covariates selected such as high school GPA, numerical representation of the level of the high school (high school deviation value), first-year GPA, and various first-year scores from the GPS-Academic assess-ment test (thinking ability, resilience, leadership, collaboration, academic motivation, experi-ence) were selected. After removing missing data, such as those who did not take the GPS-Academic, a final dataset of 681 individuals was created, as shown in Table 2.

Table 2: The Covariates by Control and Treated Groups

Variable	Control, N = 432 [1]	Treated, N = 249 [1]
GPA(Original)	2.66 (0.69)	2.81 (0.60)
GPA(Deviation Value	51 (9)	53 (8)
High School levels	58 (7)	52 (6)
High School GPA	3.70 (0.53)	4.21 (0.39)
Thinking	50 (8)	45 (8)
Resilience	48 (9)	49 (9)
Leadership	47 (9)	47 (9)
Collaboration	49 (8)	49 (8)
Experience	56 (15)	58 (13)
Motivation	68 (18)	67 (16)

[1] Mean (SD)

Using this data, an effect evaluation of pre-university education through propensity score matching was conducted. First, Preprocessing was conducting the R language environment [7] and tidyverse package[8]. Then, propensity score matching was performed using the Matchit package[9]. This package allows propensity score matching to be executed similarly to regression analysis. The Average Treatment Effect (ATE) was estimated in this case, and the data was divided into five strata using stratified analysis. The distance estimation method for approximating covariates was binomial logistic regression.

3 Results

First, the propensity score matching results are shown in Figure 1 and Table 3. These results indicate that the covariates of the treatment and control groups have changed to a level that can be considered identical through propensity score matching[10]. Therefore, it was judged possible to infer the educational effects using the counterfactual framework.

Next, the average treatment effect (ATE) was calculated from the model obtained through propensity score matching. The calculation used the bootstrap BCa method (bias-corrected and accelerated percentile method). ATE's effect size and 95% confidence interval were 1.978 (0.681, 3.244). Figure 3 shows the distribution of these results.

Table 3: Propensity Score Matching Result and Covariates

Variables/Subclass	**1**, N = 136[1]	**2**, N = 136[1]	**3**, N = 137[1]	**4**, N = 135[1]	**5**, N = 137[1]
GPA(Original)	2.48 (0.76)	2.64 (0.67)	2.79 (0.63)	2.76 (0.62)	2.89 (0.53)
GPA(Deviation Value)	49 (10)	51 (9)	53 (8)	53 (8)	54 (7)
High School levels	63 (5)	60 (5)	57 (5)	53 (5)	48 (5)
High School GPA	3.30 (0.38)	3.57 (0.35)	3.91 (0.32)	4.19 (0.32)	4.47 (0.33)
Thinking	57 (6)	49 (7)	48 (6)	47 (7)	42 (9)
Resilience	47 (9)	49 (10)	50 (9)	48 (9)	49 (9)
Leadership	46 (9)	46 (9)	48 (9)	48 (10)	47 (10)
Collaboration	48 (8)	49 (8)	50 (8)	50 (9)	49 (8)
Experience	54 (15)	55 (14)	59 (14)	60 (15)	58 (14)
Motivation	68 (17)	66 (17)	69 (16)	68 (17)	65 (18)
Group					
Control	131 (96%)	119 (88%)	89 (65%)	62 (46%)	31 (23%)
Treated	5 (3.7%)	17 (13%)	48 (35%)	73 (54%)	106 (77%)
distance	0.04 (0.02)	0.14 (0.04)	0.31 (0.05)	0.53 (0.08)	0.80 (0.08)
weights	1.00 (1.75)	1.00 (0.73)	1.00 (0.03)	1.00 (0.35)	1.00 (0.98)

[1] Mean (SD); n (%)

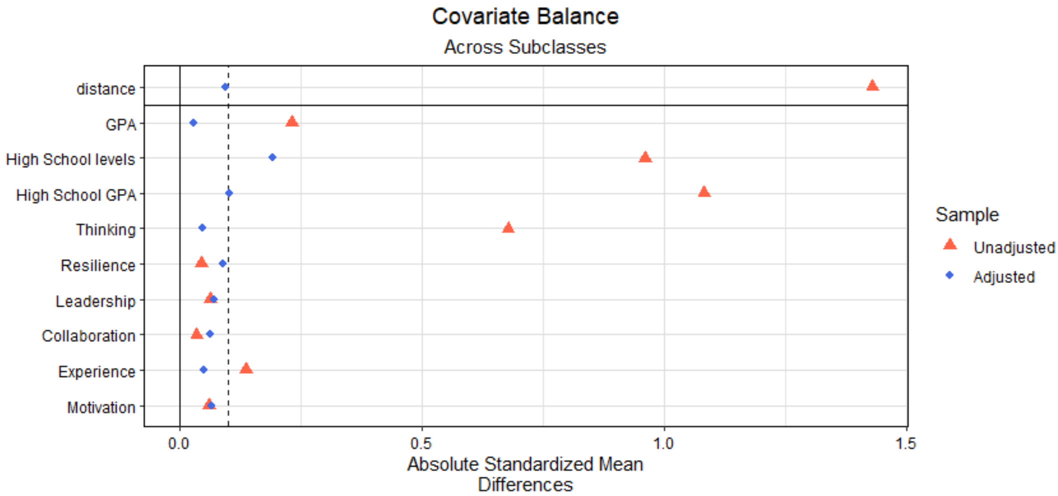

Figure 2: Propensity Score Matching Result and Covariates Balance

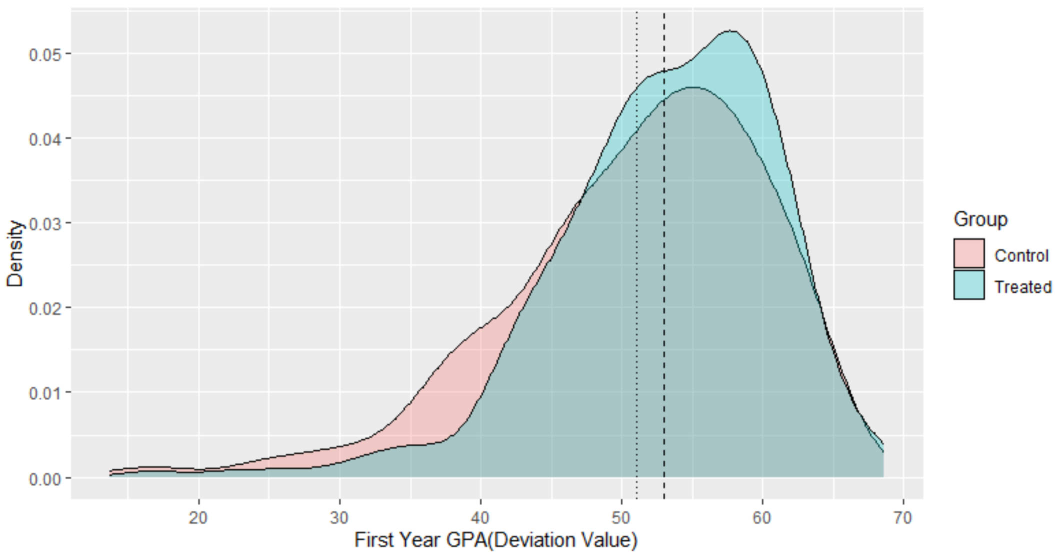

Figure 1: GPA Density from Propensity Score Matching Result

4 Discussion

Regarding the propensity score matching balance adjustment, the Absolute Standardized Mean Difference shouldbe below 0.1. In this study, all variables except the High School Level variable fell within this range. The Absolute Standardized Mean Difference for the High School Level variable was approximately 0.2. These findings suggest that the covariate bias in this study's propensity score matching was sufficiently corrected. Furthermore, the ATE was 1.978 (0.681, 3.244), indicating that the GPA of the treatment group was higher. Therefore, the hypothesis of this study that pre-university education raises first-year GPA was supported.Additionally, Figure 3 shows that compared to the control group, the treated group's distribution is more centralized, suggesting that students predicted to have lower basic academic skills and consequently lower GPAs were upwardly adjusted by the effects of pre-university education. This can enhance basic academic skills, which is one of the primary purposes of pre-university education for students admitted through recommendation-based exams, and the results support its functionality. Additionally, the Faculty of Network Information decided to continue the existing pre-university education methods based on the findings of this study. Demonstrating evidence using accumu-lated educational data in support of such decision-making is a good practice in the role and ac-tivities of Institutional Research. Confirming the causal relationship between academic content and its outcomes is essential in verifying educational effects. Even in the field of university ed-ucation, where generally only observational data is available, adopting propensity score match-ing methods has demonstrated that causal inference regarding educational content and its out-comes is possible.

5 References

[1] K. Iwasaki, “Evidence-based Policies in Education: New Directions and Issues,” *Japanese J. Eval. Stud.*, vol. 10, no. 1, pp. 1_17–1_29, 2010.

[2] P. R. Rosenbaum and D. B. Rubin, “Reducing bias in observational studies using subclassification on the propensity score,” *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 516–524, 1984.

[3] D. B. Rubin, “Estimating causal effects of treatments in randomized and nonrandomized studies,” *J. Educ. Psychol.*, vol. 66, no. 5, pp. 688–701, Oct. 1974.

[4] F. J. Thoemmes and E. S. Kim, “A Systematic Review of Propensity Score Methods in the Social Sciences,” *Multivariate Behav. Res.*, vol. 46, no. 1, pp. 90–118, Feb. 2011.

[5] M. Li, “Using the Propensity Score Method to Estimate Causal Effects,” *Organ. Res. Methods*, vol. 16, no. 2, pp. 188–226, Apr. 2013.

[6] M. H. Clark and N. L. Cundiff, “Assessing the Effectiveness of a College Freshman Seminar Using Propensity Score Adjustments,” *Res. High. Educ.*, vol. 52, no. 6, pp. 616–639, Sep. 2011.

[7] R Core Team, “R: A Language and Environment for Statistical Computing.” Vienna, Austria, 2018.

[8] H. Wickham, “tidyverse: Easily Install and Load the ‘Tidyverse.’” 2017.

[9] D. E. Ho, K. Imai, G. King, and E. A. Stuart, “**MatchIt** : Nonparametric Preprocessing for Parametric Causal Inference,” *J. Stat. Softw.*, vol. 42, no. 8, 2011.

[10] N. Greifer, “cobalt: Covariate Balance Tables and Plots.” 2024.