# Analysis of Financial Incentives on Grades in College

Junhao Yu

January 14, 2023

**Abstract**

This project analyzes the causal effect of academic achievement rewards(financial incentives) on higher average grades in college according to the OK randomize experimental data[2]. By utilizing methods like Fisher's Exact P-Values, Neyman's Repeated Sampling Approach, Regression and Model-Based Inference for Completed Randomized Experiment(CRE), Stratified Randomize Experiment(SRE) and Pairwise Randomized Experiment(PRE) under Rubin's framework[4], I found academic achievement rewards have not significant positive effects on higher average grades in overall population and have slightly significant positive effects on higher average grades only in certain stratification.

# 1 Introduction

Whether academic supports have positive effect on higher grades is long-researching problem because the concerns about the completion of college have been rising since the enrollment rates have increased. Such supports can be divided into two categories: support services and financial incentives. For the former one, Scrivener *et al.*[6] proved it has positive causal effects while Angrist *et al.*[1] and MacDonald *et al.*[5] found no significant positive causal effects based on the same experimental trial results. For the latter one, Dynarski *et al.*[3] presented a great positive effects of financial incentives on higher grades while Sioquist *et al.*[7, 8] proved quite the opposite result both according to Georgia's Helping Outstanding Pupils Educationally (HOPE) program observational results. Therefore, concrete and valid analysis based on randomized experiment trials, especially about financial incentives, is demanded to provide more credible causal effect results. Based on this need, Angrist *et al.*[2] conducted "Opportunity Knocks"(OK) experiment in Canadian commuter university. As a randomized experimental study, the OK program randomly assigned those who had applied(first and second year students) into control and treatment group. The treated students could get \$100 for every course they attained a grade of 70 or more and additional \$20 for each percent over 70 while the controlled students could not get paid. The high reward levels, high program engagement and randomized experimental design made trial results more valid.

This report will give an analysis on this randomized experimental data. The potential outcome(Y) we cast focus on is the average grade for the whole year of 2008(continuous variables ranging from 0 to 100). Covariates we care about includes the gender, the first/second year, whether students' first language is English and students' high school grade average(continuous variables ranging from 0 to 100). The goal is to figure out whether or not academic achievement rewards(financial incentives) have a significantly positive causal effect on higher average grades.

# 2 Method

In this report I will estimate the Average Causal Effect(ACE) $\tau$ under Rubin's framework with methods like Fisher's Exact P-Values, Neyman's Repeated Sampling Approach, Regression and Model-Based Inference for Completed Randomized Experiment(CRE), Stratified Randomize Experiment(SRE) and Pairwise Randomized Experiment(PRE) since this is a randomized experiment. First I perform all four methods on the data for CRE to get a baseline results after I check the balance of covariates. Then according to the baseline results, using Neyman's Approach and Regression for SRE and Neyman's Approach for PRE to get more precise estimate of ACE.

The variables in this report I use include the average grade for the whole year of 2008(Y) and covariates like the gender, the first/second year, whether students' first language is English and students' high school grade average. Also, for the reasonableness of assumption(the details will be shown afterwards), I also create a new variable called *delta*, which is the difference between college grades and high school grades(i.e. $delta = Grade_{college} - Grade_{high\ school}$).

The assumptions we need in this report are as follows:

- SUTVA
  - No Interference: the potential outcomes for any unit do not vary with the treatments assigned to other units.(i.e. $Y_i(\vec{t}) = Y_i(t_i)$).
  - No Hidden Variations of Treatments: for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.(i.e. $Y_i(t_i) \in Y_i(t), Y_i(c)$)
- Indivisualistic: $p_i(\mathbf{X}, \mathbf{Y(0)}, \mathbf{Y(1)}) = q(X_i, Y_i(0), Y_i(1))$
- Probabilistic: $0 < p_i(\mathbf{X}, \mathbf{Y(0)}, \mathbf{Y(1)}) < 1$
- Unconfounded: The assignment mechanism does not depend on the potential outcomes.

## 2.1 CRE

### 2.1.1 Balance of Covariates

In this part, I use both quantitative and qualitative methods to check whether the covariates are balanced. For covariates like gender, the vital traits are their means since they are binary variables. For covariates like high school grades, the vital traits are their kernal density plots since they are continuous variables. Therefore, I use summary() in R package **base** and geom_density() in R package **ggplot2**.

### 2.1.2 Fisher's Exact P-Value(FEP)

The $T^{dif}$ I use is as follows:

$$T^{dif} = |\bar{Y}_t^{obs} - \bar{Y}_c^{obs}| = |\frac{\sum_{i:W_i=1} Y_i^{obs}}{N_t} - \frac{\sum_{i:W_i=0} Y_i^{obs}}{N_c}|$$

Hence, the null hypothesis $H_0$ is $Y_i(0) = Y_i(1)$ while the alternative hypothesis $H_1$ is $Y_i(1) = Y_i(0) + C$.

In this report, I simulate sampling 3000 times. Since FEP cannot calculate the ACE precisely but only serve as a benchmark, I do not explore much about this method.

### 2.1.3 Neyman's Approach

In CRE, Neyman's approach is rather straightforward. If we further assume that $Y_i(1) - Y_i(0) = \tau_{fs}$, we have the following equations:

$$\hat{\tau} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}, \hat{V}(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$$

### 2.1.4 Regression

In order to incorporate covariates, I use linear regression method, based on lm() in R package **base**. Specifically, I treat all variables except $Y$ as explanatory variables and regress $Y$ on them.

### 2.1.5 Model-Based Inference

Since in this problem, there is one obviously violated assumption, which is $E(\mu_t) = 0$ and $E(\mu_c) = 0$, where $\mu_c$ and $\mu_t$ are the expectations of $Y_i(0)$ and $Y_i(1)$, respectively. Hence, if I imitate the existing conclusions deducted in the lecture note, there could be mistakes. Therefore, I create a new variable *delta*, whose prior expectation can be reasonably seen as 0. Combined with other priori, I simulate the model-based inference 1000 times. The other priori are mostly based on intuitions. For instance, the fluctuation of one person's grades cannot be too large.

## 2.2 SRE

### 2.2.1 Neyman's Approach with Regression

In 2.1.4, I find that covariates like gender, year and high school grades have significant influence on the outcome. Also, due to the fact that gender and year are all binary variables, I use them as the standards of stratification. As for high school grades, I incorporate it in the model by regression method. Specifically, in each strata, I first use regression method to get $\tau(\hat{j})$ and $\hat{V}(\tau(\hat{j}))$. Then, based on stratified Neyman's Approach, I use the following equations to get $\hat{\tau}$ and $\hat{V}(\hat{\tau})$ on overall population.

$$\hat{\tau}^{strata} = \sum_{j=1}^{J} \frac{N(j)}{N} \tau(\hat{j}), \hat{V}(\hat{\tau}^{strata}) = \sum_{j=1}^{J} (\frac{N(j)}{N})^2 \hat{V}(\tau(\hat{j}))$$

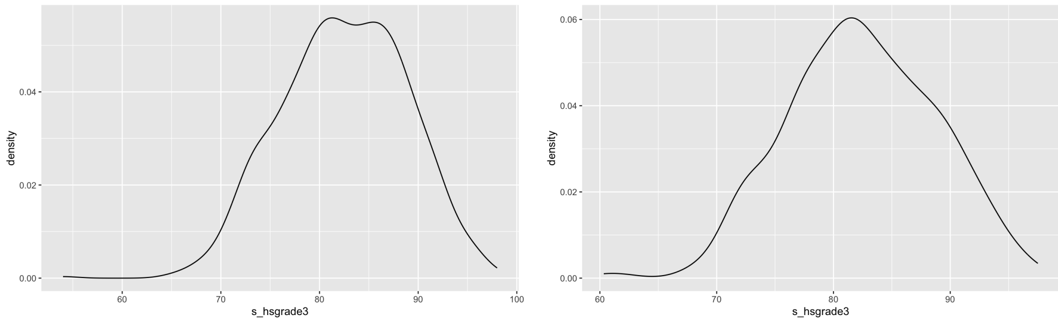


Figure 1: The kernal density of high school grades in control(left) and treatment(right) group

## 2.3 PRE

### 2.3.1 Neyman's Approach

In order to get more accurate estimate, I use Neyman's Approach for PRE. In this report, in order to find pairs with similar covariates as many as possible, I use random sampling with replacement when I already find several units with similar covariates. If we further assume constant treatment effect within pairs and that treatment effect is constant across pairs, we can calculate the estimates of ACE and its variance based on the following equations:

$$\hat{\tau} = \frac{2}{N}\sum_{j=1}^{N/2} \hat{\tau}^{pair}(j)$$

$$\hat{V}^{pair}(\hat{\tau}) = \frac{4}{N \times (N-2)}\sum_{j=1}^{N/2} (\hat{\tau}^{pair}(j) - \hat{\tau})^2$$

# 3 Result

From the table 1 and table 2, several statistics of covariates are similar in control and treatment group. The only large difference occurs in the mean of gender, suggesting some kinds of imbalance, which leads me to use SRE and PRE. From the figure 1, we can see qualitatively that the distributions of high school grades are similar in control and treatment group.

| | **High School Grades** | **Gender** | **Year** |
|---|---|---|---|
| **1st Q** | 78.00 | 0.0 | 0.0000 |
| **Median** | 82.17 | 0.5 | 1.0000 |
| **Mean** | 82.39 | 0.5 | 0.5209 |
| **3rd Q** | 87.00 | 1.0 | 1.0000 |

Table 1: The statistics in treatment group

| | **High School Grades** | **Gender** | **Year** |
|---|---|---|---|
| **1st Q** | 78.17 | 0.0000 | 0.0000 |
| **Median** | 82.67 | 0.0000 | 1.0000 |
| **Mean** | 82.49 | 0.2826 | 0.5895 |
| **3rd Q** | 87.17 | 1.0000 | 1.0000 |

Table 2: The statistics in control group

## 3.1 CRE Results

From the table 3, we can see that all methods give the same results: academic achievement rewards have no significant positive effects on higher average grades in overall population.

| | **FEP** | **Neyman's Approach** | **Regression** | **Model-Based(delta)** |
|---|---|---|---|---|
| $\hat{\tau}$ | NaN | 0.487178 | -0.02409 | -25.09619 |
| $\hat{\sigma}(\hat{\tau})$ | NaN | 0.6230992 | 0.51042 | 18.4618 |
| **t-statistic** | NaN | NaN | -0.047 | NaN |
| **p-value** | 0.4323333 | NaN | 0.9624 | NaN |
| **Confidence Interval(CI)** | NaN | (-0.734, 1.708) | (-1.025, 0.976) | (-61.281, 11.089) |

Table 3: The results of four methods for CRE

## 3.2 SRE Results

It can be seen clearly that after stratification but without regression, the estimated variance of estimated $\hat{\tau}$ have remained the same. But the $\hat{\tau}$ has dropped a lot. I reckon this is because the stratification provides additional explanatory ability to explain $Y$, while such ability has been attributed to treatment. This phenomenon partly proves that the stratification is useful and necessary, correcting imbalance errors in CRE.

Also, by comparing the results with and without regression, we can see that the estimated variances of estimated $\hat{\tau}$ with regression drop conspicuously. Also, regression has made several stratas more significant compared with no regression. This phenomenon proves that high school grades have great impact on average grades in college. Only after we strip this influence can we accurately estimate ACE. Though the $\hat{\tau}$ in overall population is not significant at all, some stratas show a little significance, such as first year female students and second year female students.

However, it can be seen that the causal effect is still not significant. Hence, we need to utilize more strict and more accurate methods to estimate it so that we can get a valid result.

| | **First Year** | | **Second Year** | | **Whole Population** |
|---|---|---|---|---|---|
| | **Female** | **Male** | **Female** | **Male** | |
| $\hat{\tau}$ | 0.484502 | -0.2127408 | 0.4703871 | -0.9562276 | 0.07375274 |
| $\hat{\sigma}(\hat{\tau})$ | 0.9372345 | 1.192675 | 1.331536 | 1.404488 | 0.637105 |
| **CI** | (-1.171, 3.111) | (-2.550, 2.125) | (-2.139, 3.080) | (-3.709, 1.797) | (-1.175, 1.322) |

Table 4: Neyman's Approach without regression for SRE

| | **First Year** | | **Second Year** | | **Whole Population** |
|---|---|---|---|---|---|
| | **Female** | **Male** | **Female** | **Male** | |
| $\hat{\tau}$ | 0.65503 | -0.47052 | 0.65169 | -0.64411 | -0.01942201 |
| $\hat{\sigma}(\hat{\tau})$ | 0.89379 | 0.90886 | 1.14662 | 1.17301 | 0.5097877 |
| **CI** | (-1.097, 2.407) | (-2.252, 1.311) | (-1.596, 2.899) | (-2.943, 1.655) | (-1.019, 0.980) |

Table 5: Neyman's Approach with regression for SRE

## 3.3 PRE Results

After matching pairs and conducting Neyman's Approach, we get the following results:

$$\hat{\tau} = 0.305334$$
$$\hat{\sigma}(\hat{\tau}) = 0.4014235$$
$$CI : (-0.481, 1.092)$$

Very clearly, the estimated variance of estimated $\hat{\tau}$ drops furthermore, suggesting superiority of PRE. However, even under this framework, the estimated causal effect is still not significant.

Thus, we are at least 95% confident to say that there is no causal effect at all in this problem. Combined with the results in 3.2, our final conclusion is that academic achievement rewards have no significant positive effects on higher average grades in overall population and have slightly significant positive effects on higher average grades only in certain stratification.

# 4 Discussion

Refer to the results in this OK experimental study, academic achievement rewards have no significant positive effects on higher average grades. Specifically, the p-value, the confidence interval of different methods for CRE, SRE and PRE provide the same conclusion above.

However, while I was conducting the above methods, I noticed that there could be violations in assumptions, which might lead to the invalidity of these results I get.

First of all, according to Angrist *et al.*[2], they provided students with additional $20 for each percent over 70. This policy actually contrasts with the assumption that there are no hidden variations in treatment(SUTVA). Different units with different scores do have different kinds of treatments. One possible consequence of this violation could be that the best students will work even harder to get such additional money, which they will not do when there is no additional money, while those students who are only fairly good will only try to get 70 scores and no more since the marginal utility of additional money is rather small because they have to struggle a lot to get additional points. This phenomenon could lead to the increasing variability of treatment group compared with control group, further impairing the validity of the results.

Second of all, there were actually more covariates that affect treatment group or control group only. For instance, Angrist *et al.*[2] in fact pointed out that treated students also had the opportunity to interact with randomly assigned peer advisors. This is a confounding factor which is not displayed in the data. Without doubt, this extra help from others will have positive impact on units' grades and I fail to incorporate such factors into the analysis since there is no such data.

Third of all, doubts can be casted on the assumption that treatment effect within pairs is constant and that treatment effect is constant across pairs. This additional assumption in PRE can be easily violated if we consider the above two kinds of violations. Therefore, the superiority of PRE might be only due to the power of this extra assumption rather than the framework itself.

Based on the discussions about the above violations, I reckon in the future we need to take all the covariates into account and make sure that the assumptions of Rubin framework actually hold in the experiments devised. Also, if I have more data about this experiment in the future, like the help from peer students, I might reconduct my analysis by incorporating such important yet missing covariates.

Based on the current results of analysis, I present a possible causal DAG for OK experiment as follows:

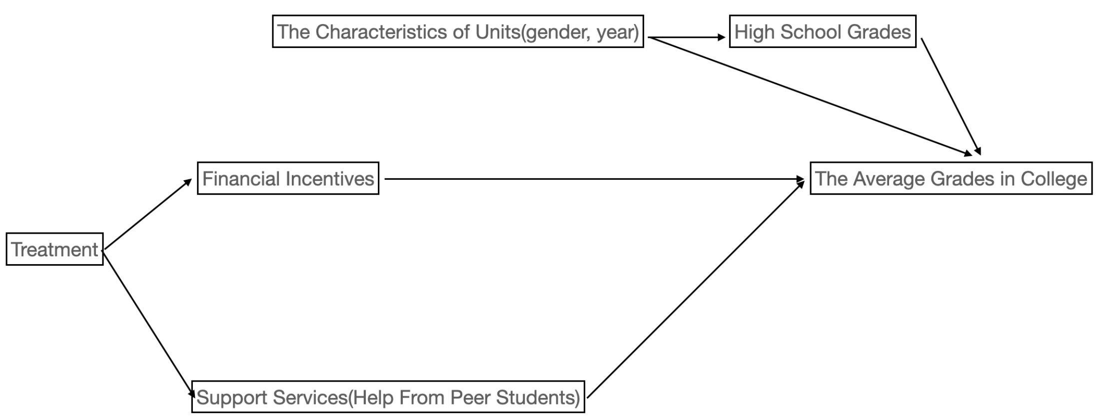


Figure 2: A Possible Causal DAG

# References


[1] Joshua Angrist, Daniel Lang, and Philip Oreopoulos. Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63, 2009.

[2] Joshua Angrist, Philip Oreopoulos, and Tyler Williams. When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources*, 49(3):572–610, 2014.

[3] Susan Dynarski. Building the stock of college-educated labor. *Journal of human resources*, 43(3):576–610, 2008.

[4] Guido W Imbens and Donald B Rubin. Causal inference in statistics, social, and biomedical sciences (2015). *Google Scholar Google Scholar Digital Library Digital Library.*

[5] Heather MacDonald, Lawrence Bernstein, and Cristofer Price. *Foundations for Success: Short-Term Impacts Report.* Canada Millennium Scholarship Foundation, 2009.

[6] Susan Scrivener and Michael J Weiss. More guidance, better results? three-year effects of an enhanced student services program at two community colleges. *Three-Year Effects of an Enhanced Student Services Program at Two Community Colleges (August 1, 2009). New York: MDRC*, 2009.

[7] David L Sjoquist and John V Winters. Building the stock of college-educated labor revisited. *Journal of Human Resources*, 47(1):270–285, 2012.

[8] David L Sjoquist and John V Winters. State merit-based financial aid programs and college attainment. *Journal of Regional Science*, 55(3):364–390, 2015.