Causal Inference – Final Project

Estimating the causal effect of wearing glasses on school achievements among myopic students

Ariel Kreisberg Nitzav 207541293

GitHub repository: https://github.com/ArielNitzav/CausalInferenceFinalProject

Introduction

Poor vision is a common phenomenon that affects children at young age at significant rates. In the US alone, it is estimated that around 6.8% percent of the children aged under 18 have a diagnosed eye and vision condition[1]. It has been reported in the past that children with uncorrected refracted errors have lower scores on various cognitive tests and may experience an improvement in reading capability when vision problems are treated. One common sight condition is myopia, or short-sightedness, which is an eye disorder that causes distant objects to appear blurry while close objects appear normal[2]. Myopia is more common among children in Asia and specifically in rural China, where the awareness to poor vision and its implications is relatively low. This, in addition to low financial and social resources, often leave various vision problems including myopia untreated or poorly treated at best. The high prevalence of myopia among Chinese children seems to be due in part to high levels of near work related to school and limited time spent outdoors[3].

In this project, I use data collected for Ma et al.[4] study "Effect of providing free glasses on children's educational outcomes in China: cluster randomized controlled trial" from 2014 which examined the effect of provision of free glasses on academic performance in rural Chinese children with myopia between October of 2012 and June of 2013 (one school year). In my project, I set to examine the **causal effect** of sight correction on school academic performance – mathematical performance, particularly. Specifically, the project examines the causal effect by estimating it using four different approaches as taught in class, as well as by a graph model of the problem.

[1] https://www.cdc.gov/visionhealth/basics/ced/fastfacts.htm

[2] https://en.wikipedia.org/wiki/Myopia

[3] Morgan IG, Ohno-Matsui K, Saw SM. Myopia. Lancet 2012; 379:1739-48

[4] Ma X, Zhou Z, Yi H, Pang X, Shi Y, Chen Q et al. Effect of providing free glasses on children's educational outcomes in China: cluster randomized controlled trial BMJ 2014; 349: g5740 doi:10.1136/bmj.g5740

The main contribution and importance of this study is because in contrary to prevalent situation in the west, in China it is a common misconception that wearing glasses harms vision, instead of improves it. Thus, it is important to show what effect wearing glasses has on the quality of living, and its specific effect on school achievements. Wearing glasses may help a myopic student become more concentrated while in class and may leave them with more cognitive resources for learning itself rather than for trying to figure out what is written on the board in class.

In this study, two groups are filtered out from the original study's participants – a treatment group of myopic children who were given free access to prescribed glasses and reportedly wore them, and a control group – a group of myopic children who reportedly did not wear glasses throughout the study. The participants took two math tests: one at baseline (start of the study) and one at the endline of the study. The latter is considered the outcome of the study, by which the causal effect is estimated.

The data

The data contains records of hundreds of students mostly in the ages of 8-12, all visually impaired and diagnosed with myopia, in different primary schools in two prefectures in China – Tianshui in Gansu, which considered to be poor, and Yulin in Shaanxi, which considered wealthier. Each school taking part in the study is the only school within its township to take part in the study, to reduce contamination between treatment groups. The study ran as a randomized clustered trial, meaning that **schools** were randomly allocated to the different treatment arms, instead of the students themselves being randomly allocated to treatments as in a regular RCT. In the original study, 3 treatment arms were used:

1. **Free glasses:** Free glasses were given by prescription based on the child’s measured refractive power and dispensed at school by the study optometrist.
2. **Vouchers:** Vouchers bearing the child’s name, school, and glasses prescription, exchangeable for free glasses at the local county hospital, at a median distance from children’s townships of 30 km (range 1-105 km).
3. **Control group:** A glasses prescription and letter to the parents informing them of the refractive status of their child, without a ready access to acquiring glasses.

Of all the children participating in the study (n=3177), 1153 (36.3%) were randomized to the free glasses group, 988 (31.1%) to the vouchers group and 1036 (32.6%) to the control group. 15% percent of them wore glasses at baseline. 96% of the participants completed the study. Some of the school were randomized for having a series of educational lectures about the importance of wearing glasses.

Data Features

The features of the data considered in this study appear in the following table. Other features that appeared in the original study's data and were omitted were either negligible or redundant.

Feature name	Summary	Range
age	Age of participant	[8, 15]
baseusage	Does the participant already wear glasses at baseline	0 – no 1 – yes
aware0	Is the participant aware of being myopic?	0 – no 1 – yes
harmvision0	Does the participant believe that wearing glasses harms the vision? (a common misconception in rural China)	0 – no 1 – yes
RE2	Refractive Error (Diopters) based on SE2, most myopic eye	1 – <-2.00 2 – >=-2.00 <-0.50 3 – >=-0.50 <0.50 4 – >=0.50
SE2	Most myopic eye SE (Spherical Equivalent)	[-14,9]
va18	Visual acuity of Better Seeing Eye – being able to identify 8 out of ten characters on a board	0 – no 1 – yes
pedu12	One or both parents with 12 years of education or more	0 – no 1 – yes
hhwealth	Household wealth	1 – poor tertile 2 – median tertile 3 – rich tertile

bboarding	Boarding at school	0 – no 1 – yes
pamigrant	Both parents out-migrated for work (in rural China it is common that parents at the age of work leave for most of the year to work in a bigger city for the hope of better income opportunities; the children are usually left with their grandparents who are often illiterate[5])	0 – no 1 – yes
edu	Educational training dummy – whether the participants went through a series of educational lectures about the importance of wearing glasses	0 – no 1 – yes
seatchange	Seat change between baseline and endline	1 – moved forward 2 – no change 3 – moved backwards
emtea_14	Proportion of test questions given by teachers in class were written on a black board	1 – All 2 – Most 3 – About half 4 – Rarely 5 – None
blackboard	Proportion of material taught on blackboard	1 – less than half 2 – half 3 – more than half

[5] Hu, X., Cook, S., & Salazar, M. A. (2008)

zmath	Baseline standardized math scores	$[-2.6227272, 2.5632114]$ mean: 0.216402 std. dev: 0.989274
wear2	Does the participant wear glasses – random check seven months later	0 – no 1 – yes
use2	Does the participant use glasses – random check seven months later (self-report)	0 – no 1 – yes
major [T, Treatment]	Major treatment arms	1 – Control 2 – Voucher 3 – Free
ezmath [Y, Outcome]	Endline standardized math scores	$[-2.4761298, 2.383235]$ mean: 0.343585 std. dev: 0.983479

Data adaptation to this project

In this project, to try and better isolate and assess the causal effect of the treatment, the original study's data was narrowed down and filtered according to these steps:

1. **Control group:** Includes only participants who were assigned to the original control group, completed the study, did not wear glasses at baseline and did not appear to wear glasses at a random check ['wear2'], as well as did not self-report on wearing glasses ['use2'] in the end of the study. This filtering process was designed to try and isolate participants who did not use glasses before and throughout the study. However, it is not certain that these participants did not wear glasses for at least short period of the study, since these participants were given prescription for glasses (but not the glasses themselves, or a voucher for their acquirement). After the filtering, the control group contained 555 participants.
2. **Treatment group:** The two original treatment groups (free glasses, vouchers) were merged into one treatment group. The resulting treatment group contained only participants who were assigned to either the free glasses or vouchers groups,

completed the study, did not wear glasses at baseline and did appear to wear glasses at a random check ['wear2'], as well as self-reported on wearing glasses ['use2'] in the end of the study. This filtering process was intended to try and isolate participants who did not use glasses before the study but persisted on wearing them throughout the study after being given them for free or after buying them using a free voucher. However, it is possible that participants in this group did not wear glasses for at least short part of the study. After the filtering, the treatment group contained 482 participants.

Identification assumptions and limitations

At class, we encountered 4 assumptions necessary for valid causal effect identification. In this section we'll go through them and discuss whether they hold, and if so, to what extent.

1. **Stable Unit Treatment Value Assumption (SUTVA):** SUTVA implies that potential outcomes for each participant are unrelated to the treatment status of other participants. It is plausible that this assumption holds in this study. Each student taking part in the study has their own myopia severity, and every student in the study received the same treatment as their fellow participating classmates who suffer from myopia and has their own educational level. Moreover, each school participating in the study is the only school from its township participating in the study. It is unlikely that one's math level will be affected by other student wearing or not wearing glasses.
2. **Consistency:** The consistency assumption is often stated as such that an individual's potential outcome under their observed exposure history is precisely her observed outcome[6]. This assumption is often problematic in RCTs as each participant is assigned to only one treatment arm, and in our case, there exists a relatively loose supervision and control on whether students who were given glasses did in fact wore glasses during the study's period, and if they did, to what extent. The same goes for students in the control group, who were given a personalized glasses prescription, but without a direct financial mean of acquiring them as in the treatment group. Thus, some students from the control group did end up wearing glasses at the endline. To try to address this shortcoming, in my study, only students who reported or were seen not wearing glasses at a random check were included in the control group and only students who reported or were seen wearing glasses at a random check were included

[6] Cole, S. R., & Frangakis, C. E. (2009). Commentary: The Consistency Statement in Causal Inference: A Definition or an Assumption? *Epidemiology*, *20*(1), 3–5.

in the treatment group. This, obviously, does not assure that the consistency assumption holds, but rather helps us get close to it.

3. **Ignorability:** According to the ignorability assumption, potential outcomes are independent of treatment assignment – this assumption holds in RCTs in general, and it is safe to assume it holds in this clustered random trial as well.
4. **Common support:** Based on the propensity scores diagram shown later in this report, it can be clearly seen that there is a significant overlap among the two groups, and that each participant had the chance of potentially appearing in the other group.

Correlations within the data

As can be seen from the following heatmap shown in figure 1, strong correlations can be found mostly among dummies of the same categorical features. Specifically, most of the features show no correlation with either the treatment ('major') or the outcome ('ezmath') – except for 'zmath' (baseline math scores) showing strong correlation with 'ezmath' (endline math scores). This is predictable, and mostly shows that stronger students remain stronger and weaker students remain weaker in terms of their math achievements as measured in the study. The lack of correlation also indicates good randomization to treatment groups as expected in a RCT.

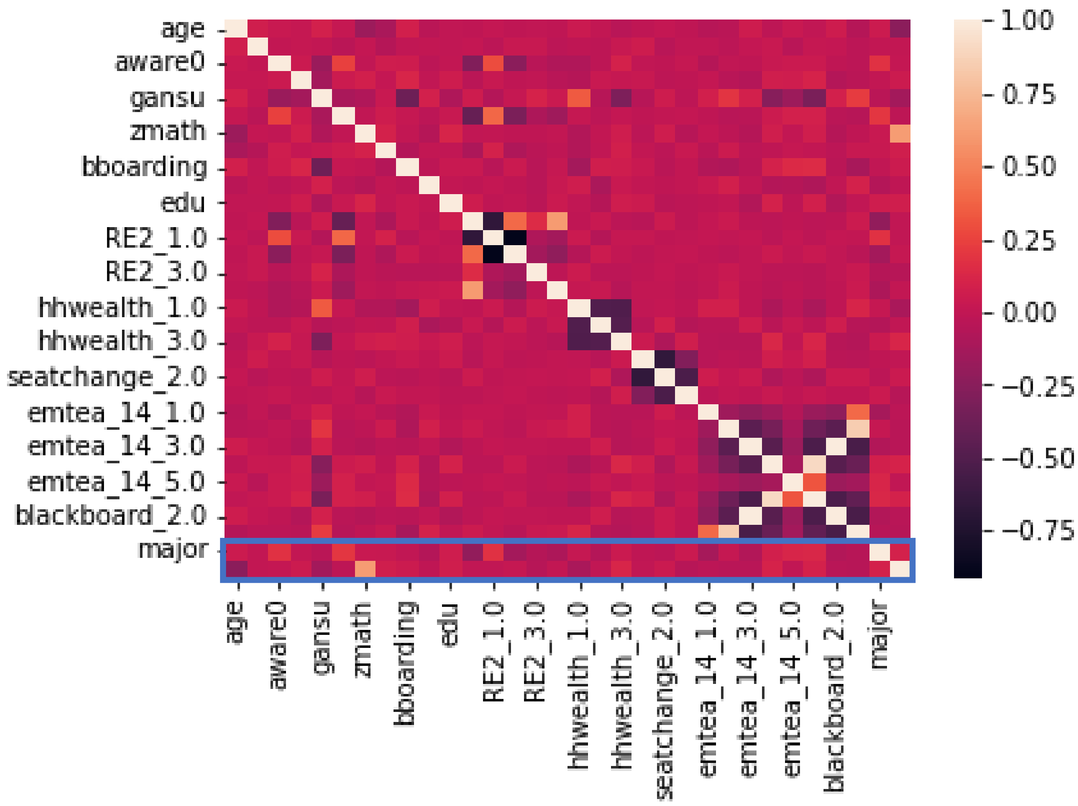


Figure 1: correlation heatmap between the data features. The last two lines marked by the blue rectangle show the correlation with the treatment assignment ('major', the first row) and the outcome row ('ezmath', the second row)

Propensity scores and overlap

Using logistic regression calculated on the treatment groups as target as taught in class, we get the following propensity scores figure (figure 2), which shows that the desired overlap between the treatment and the control group holds to a great extent. This shows that the common support assumption does, in fact, hold in this case.

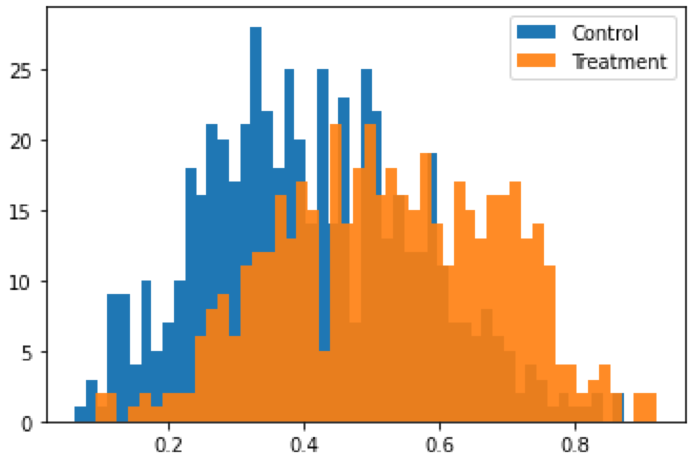


Figure 2: propensity scores for the two treatment groups

Causal effect estimation

To estimate the causal effect of wearing glasses on school achievements among myopic students using ATE, four approaches that were taught in class were used:

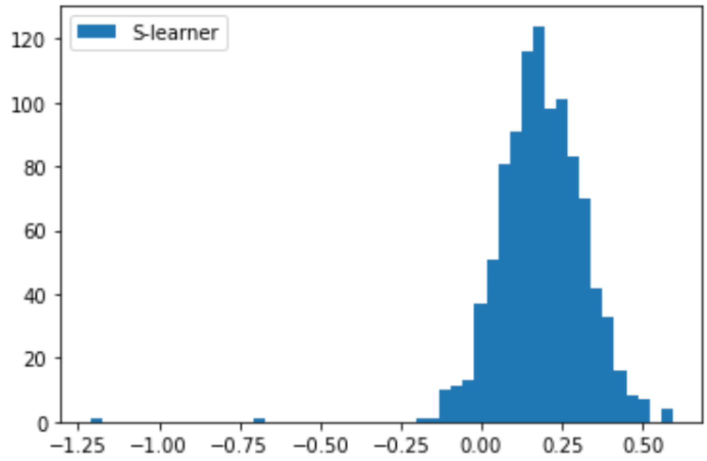
1. S-learner
2. T-learner
3. 1-NN matching
4. IPW estimation

To show statistical significance, a bootstrap process was used for each of the four approaches:

1. for 1, …, 1000:
    a. sample 100 participants from the control and from the treatment, resulting in 200 participants in total;
    b. calculate ATE based on the chosen approach and sampled participants
2. calculate the average ATE, std and confidence intervals from the resulting distribution

<u>S-learner</u>

For the S-learner, a linear regression model f was trained on the whole sample in each bootstrap iteration, with the treatment variable as feature. Then, the ATE was estimated as explained in the tutorial: $\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} f(x_i, 1) - f(x_i, 0)$. The ATE distribution from the bootstrap process appears in figure 3.



<u>Figure 3:</u> ATE empirical bootstrap distribution based on the S-learner

T-learner

For the T-learner, two linear regression models f_0, f_1 were trained on the sampled control and treatment groups, respectively, in each bootstrap iteration. Then, ATE was estimated as explained in the tutorial: $\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} f_1(x_i) - f_0(x_i)$. The ATE distribution from the bootstrap process appears in figure 4.

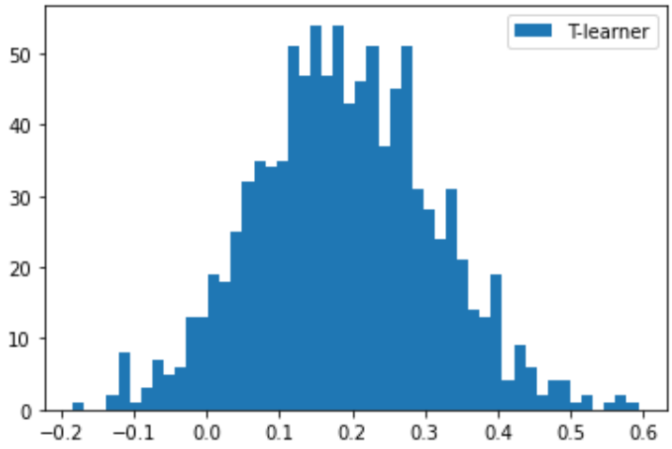


Figure 4: ATE empirical bootstrap distribution based on the T-learner

1-NN matching

For the matching approach, for each participant in the sample from one group, its "closest" neighbor (in terms of cosine similarity) from the other group was found in order to calculate ITE values which were later averaged into a final ATE score at each iteration. The resulting empirical ATE distribution appears in figure 5.

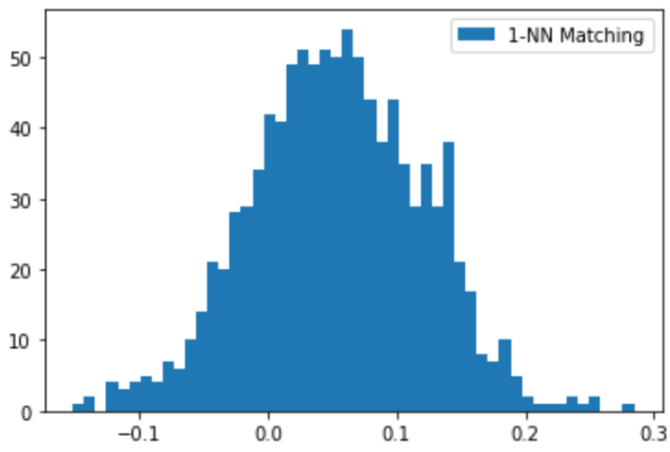


Figure 5: ATE empirical bootstrap distribution based on the 1-NN matching approach

IPW

For the IPW approach, at each bootstrap iteration, propensity scores were calculated by training logistic regression model on the treatment group assignment, and the ATE value was estimated using the formula shown in tutorial 8. The resulting empirical ATE distribution appears in figure 6.

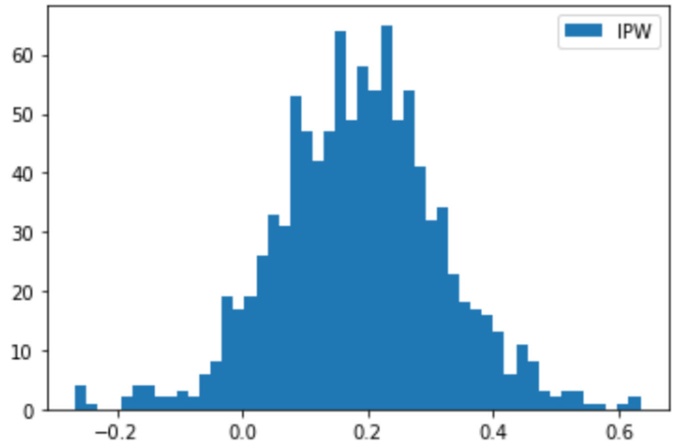


Figure 6: ATE empirical bootstrap distribution based on the IPW approach

Figure 7 shows the different distributions resulting from the different approaches stacked on top of each other for easier comparison.

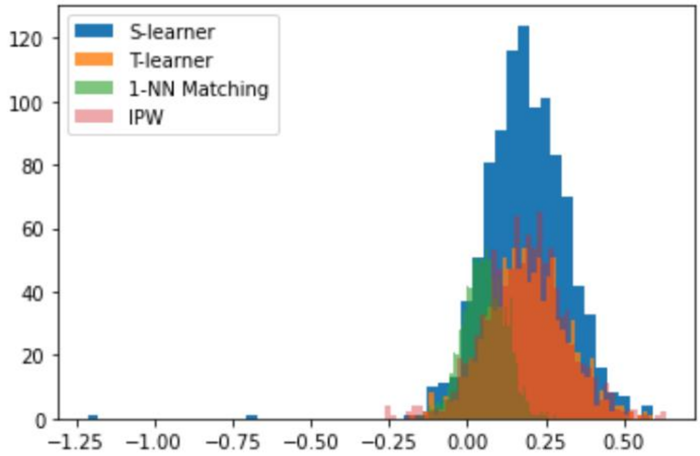


Figure 7: different ATE empirical bootstrap distributions, based on the four different approaches mentioned above

Results

	S-learner	**T-learner**	**1-NN matching**	**IPW**
Average value and standard deviation	0.187 ± 0.133	0.189 ± 0.122	0.054 ± 0.066	0.187 ± 0.131
95% CI	$(-0.056, 0.433)$	$(-0.046, 0.433)$	$(-0.079, 0.18)$	$(-0.061, 0.454)$

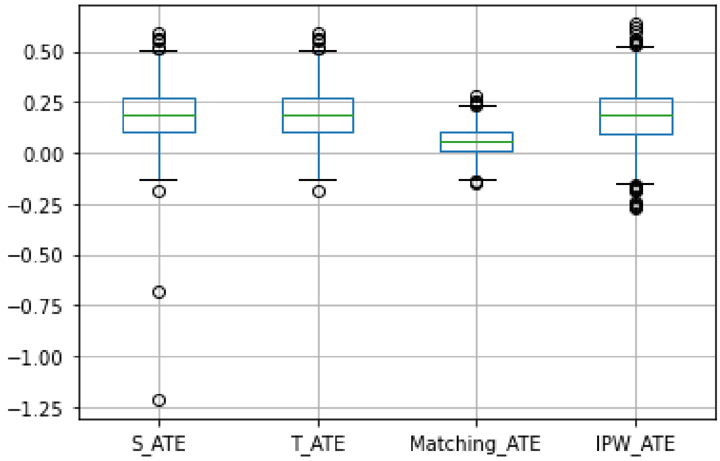


Figure 8: Boxplot of the resulting empirical distributions, according to the different four estimation approaches

As can be seen in the results above, the ATE across the different approaches lies around 0.187. In the original data, the median endline math score ('ezmath', the outcome of the study) was 0.5414, the 75% percentile was 1.123 and the 90% percentile was 1.483, 0.187 is a significant leap that **can set apart average students from exceling students**. Although the standard deviation values across the different approaches were relatively low (around 0.13 for 3 out 4 approaches), all the confidence intervals at significance level of 95% included 0 within their range, making these results somewhat statistically inconclusive. To conclude, there appears to be a causal effect of wearing glasses when visually impaired (myopic), but the effect is not founded enough, statistically.

Causal graph approach

Another approach that was considered to the problem was modeling the data and its dependencies using causal graph. The graph and its dependencies included some of the features mentioned earlier in this work (for example, age was excluded for logical reasons), and were constructed according to prior knowledge about Chinese socioeconomics as well as plain common sense. The dependencies and their justifications appear in the appendix, in the end of this work. After the graph was manually constructed, it was analyzed using the DoWhy Python library[7].

1. The causal graph

   The resulting graph appears in figure 9. It contains logical dependencies, for example: if parents are out of town, migrating for work in a bigger city ('pamigrant'), it may affects the student's boarding status and their household wealth status, etc.; if parents lack 12 years of formal education, the student may be more likely to think that wearing glasses harms the vision (a common misconception in China ['harmvision0']) or may affect the chance of them migrating out of the village for work, leaving the child to stay with their grandparents who may be illiterate.

   RE2 (refractive error of the most myopic eye) serves as a confounder, affecting both the treatment and the outcome.

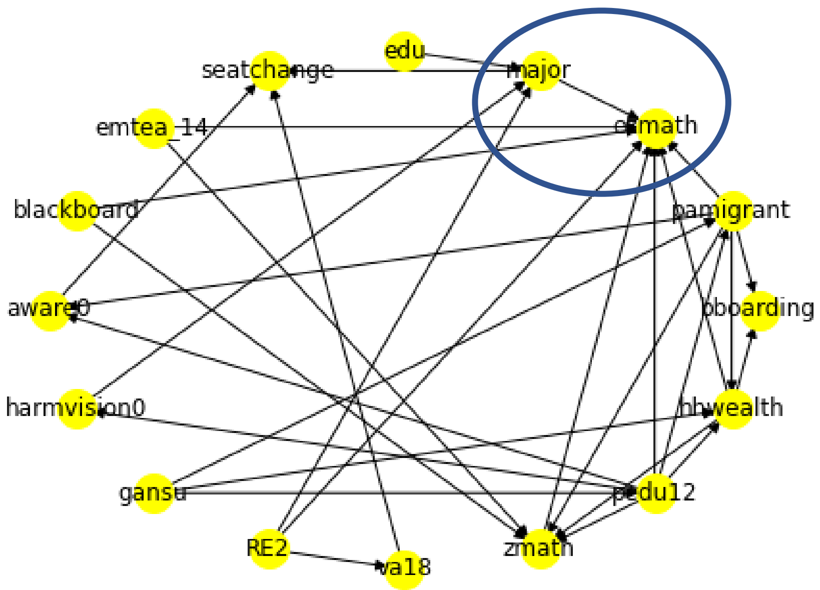


Figure 8: the causal graph of the problem; 'major' (treatment assignment) and 'ezmath' (outcome) are marked by a circle

Amit Sharma, Emre Kiciman, et al. DoWhy: A Python package for causal inference. 2019 [7]

2. Identifying the effect

   Using the backdoor criterion, it is possible to identify the effect based on the graph by conditioning on 'pedu12' (one or both parents with 12 years of education or more) and 'RE2' (the confounder). It is important to note that an unconfoundedness assumption is assumed in this case, meaning that if there exists another confounder U that affects both the treatment and the outcome, it holds that $P(ezmath|major, pedu12, RE2, \boldsymbol{U}) = P(ezmath|major, pedu12, RE2)$.

3. Estimating the effect

   After identifying the effect, DoWhy allows for its estimation based on the data and the causal graph using various approaches. Using the backdoor propensity score weighting, we get that the ATE is equal to 0.189, with confidence intervals equal to $(0.084,\ 0.28)$. This is estimation is very close to the previous mentioned estimations, which were mostly around ATE of 0.187. This further validates the causal effect of wearing glasses on math performance among myopic students.

4. Refuting the estimation

   DoWhy allows for various methods of trying to refute the causal model based on the data and the causal graph. I used two of them: the first – creating a synthetic random common cause, a confounder that "affects" both the treatment and the outcome. If the assumption was correct, the estimation should not change. Using this method, the new estimated effect (ATE) was 0.189376, while the "real" estimated effect was 0.189387. This resulted in a p-value of 0.96, meaning with our estimation wasn't refuted. The second refutation method considered was to use only subset of the data when estimating the causal effect; if the assumption was correct, the estimation should not change that much. In this case, the "new" estimated effect was 0.1869 with p-value of 0.98, which leads to refuting the assumption that the causal model does not hold.

Discussion and possible weaknesses

In this study, I examined the causal effect of wearing glasses on math achievements among myopic students. Via various estimation approaches, it can be concluded that such causal effect does exist and caused myopic students to perform better on math tests. This aligns with previous literature. The nature of the original study and the way it was conducted allowed for great randomization that helped in reducing socioeconomic differences and school achievements among the different students that participated in the study but suffered from a few shortcomings as well. Despite being a clustered randomized controlled trial, this study suffers from multiple weaknesses and shortcomings. One fundamental shortcoming is that most of the data collected in this study was self-reported (by the students participating) or reported by the staff of the different schools which took part in the study, as it may lead to inaccuracies in the data, and may cause it not to reflect the reality as desired. Specifically, this is crucial in terms of the consistency of the study. As explained previously, the assignment to groups in this study was done based on filtering students by their original study's group assignment (control, free glasses, voucher for free glasses), their self-reports on whether they wore glasses or not and based on random check conducted by the school staff on whether the students did, in fact, wear glasses (for the unified treatment group) or did not wear glasses (for the control group). This makes the treatment assignment somewhat inconclusive, as the original study's researchers could not make sure that each student wore their glasses in school or didn't, respective to their treatment assignment. Another very important shortcoming is the lack of information about the school achievements of the participating students, as well as whether they received any additional or external help with their studies, such as private tutoring, homework assistance, extracurricular activities and so on. In addition, not much is known about the schools themselves and their financial and educational resources, which are known to have a great effect on the students' achievements – beside the prefecture in which they are located.

Future work may include collecting data in a more supervised manner with greater control on the students and on whether they wear glasses or not during the study, as well as collecting more data regarding their academic status, to better isolate the effect of wearing glasses apart from scholar resources which may or may not be provided to the students.

References

1. https://www.cdc.gov/visionhealth/basics/ced/fastfacts.htm
2. https://en.wikipedia.org/wiki/Myopia
3. Morgan, I. G., Ohno-Matsui, K., & Saw, S. M. (2012). Myopia. Lancet (London, England), 379(9827), 1739–1748. https://doi.org/10.1016/S0140-6736(12)60272-4
4. Ma X, Zhou Z, Yi H, Pang X, Shi Y, Chen Q et al. Effect of providing free glasses on children’s educational outcomes in China: cluster randomized controlled trial BMJ 2014; 349: g5740 doi:10.1136/bmj.g5740
5. Hu, X., Cook, S., & Salazar, M. A. (2008). Internal migration and health in China. Lancet (London, England), 372(9651), 1717–1719. https://doi.org/10.1016/S0140-6736(08)61360-4

6. Cole, S. R., & Frangakis, C. E. (2009). Commentary: The Consistency Statement in Causal Inference: A Definition or an Assumption? Epidemiology, 20(1), 3–5. http://www.jstor.org/stable/25662662
7. Sharma, A., & Kiciman, E. (2020). DoWhy: An End-to-End Library for Causal Inference. arXiv preprint arXiv:2011. 04216.

Appendix

<u>Causal graph:</u>

Vertices (data features): aware0; harmvision0; gansu; RE2; va18; zmath; pedu12; hhwealth; bboarding; pamigrant; edu; seatchange; emtea_14; blackboard; ezmath (outcome); major (treatment)

Dependencies:

aware0 → seatchange;

harmvision0 → major;

gansu → hhwealth;

gansu → pamigrant;

gansu → pedu12;

pedu12 → aware0;

pedu12 → harmvision0;

pedu12 → pamigrant;

pedu12 → hhwealth;

pedu12 → zmath;

pedu12 → ezmath;

hhwealth → bboarding;

hhwealth → zmath;

hhwealth → ezmath;

pamigrant → bboarding;

pamigrant → aware0;

pamigrant → hhwealth;

pamigrant → zmath;

pamigrant → ezmath;

edu → major;

emtea_14 → zmath;

emtea_14 → ezmath;

zmath → ezmath;

blackboard → zmath;

blackboard → ezmath;

major → ezmath;

major → seatchange;

RE2 → ezmath;

RE2 → major;

RE2 → va18;

va18 → seatchange