[SI] [YokeHong] - ECO372 Assignment 3
Student Nb: [1005815806]

a.

(i) The unit of observation in the dataset are secondary school students in Colombia during the mid-90s who maintained satisfactory academic progress. In the following questions, we would be exploring the relationship between voucher provision to these students and their test scores.

(ii) There are 283 observations

(iii) The variable is "vouch0".

(iv) The variables are "math", "reading", "writing", "totalpts" respectively.

b.

	Sample Average	Sample Standard Deviation
Math	0.0028718	1.002821
Reading	0.0006201	1.003512
Writing	0.0049475	1.001712

As mentioned in the article, the "test scores are in standard deviation units". By observing the sample average and the standard deviation of the math, reading and writing test scores, we observe that they can be interpreted as a standard normal distribution. The mean and standard deviation have values close to a mean of 0 and a standard deviation of 1, implying a standard normal distribution.

c.

All Applicants	(1) Total Points	(2) Math Scores	(3) Reading Scores	(4) Writing Scores
Student won voucher	0.217	0.178	0.204	0.126
	(1.79)	(1.50)	(1.66)	(1.03)
Observations	282	282	283	283

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

By comparing the values obtained in part (c) from STATA to the estimates in table 5 of the paper, we do not find large differences between the estimates.

d.

All Applicants	(1) Total Points
Student won voucher	0.0236
	(0.17)
Observations	189

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Large differences are present between the estimates I obtain to those in the table. The paper indicates an OLS estimate with covariates of 0.205 for the total points of all applicants whereas the estimate obtained from STATA is determined to be 0.0236. This may be due to the fact that since there are 189 observations from the STATA data whereas there are 282 observations from the paper, it implies a difference in the sample which would affect the estimates determined. Additionally, the lower sample size in the STATA data would also indicate a higher standard error and lower precision of estimates when compared to the estimate from the paper.

e.

Table 5 column (1) represents the effect of winning the voucher lottery on the total points of subtests consisting of mathematics, reading and writing for all applicants, derived using an OLS estimator.

The number 0.217 standard deviation has a t-statistic of 0.217/0.116 which equals to 1.87 which is larger than 1.65 thus, it is marginally significant at the 10% level. Therefore, we interpret the number 0.217 as that if the students were to win the voucher lottery, these winners will score 0.217 standard deviations greater than lottery losers on average.

This estimate is interesting as we can observe that by providing the students with vouchers to cover the cost of their school, these students are able to perform better in school as they tend to achieve higher test scores.

f.

(i) The unit of observation in the dataset are secondary school students in Colombia during the mid-90s who maintained satisfactory academic progress. In the following questions, we would be exploring the relationship between voucher provision to these students and their length of schooling. There are 1135 observations.

(ii) The variable “vouch0”.

(iii) The variable “usesch”.

(iv) The variables “scyfnsh” and “inschl” respectively.

g.

	(1) Highest Grade Achieved by Student	(2) School Status at Time of Survey
Student won voucher	0.126*	0.00764
	(2.45)	(0.38)
Observations	1135	1135

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From columns (1) and (2) in my data generated from STATA, we can compare these values to Table 3, column (3), rows "Highest grade completed" and "Currently in school".

For "Highest grade completed", the paper obtained an estimate of 0.130 and a standard error of 0.051 for the coefficient of WonVoucher. In my data from STATA, I generated an estimate of 0.126 and a standard error of 0.0515738 which are both very similar to the data in the paper.

For "Currently in school", the paper obtained an estimate of 0.007 and a standard error of 0.020 for the coefficient of WonVoucher. In my data from STATA, I generated an estimate of 0.00764 and a standard error of 0.020327 which are also both very similar to the data in the paper.

h.

No, estimating (2) does not estimate the causal effect of actually using the voucher. The estimate is done via an IV estimate where the assignment of the voucher is seen as an instrument for estimating a causal effect of using the voucher on the outcome.

We observe that not all students that are allocated vouchers actually used the vouchers. Allocating the vouchers is done using a random lottery process which would solve part of this issue. Therefore, it could imply that there is no selection bias and the OLS yields unbiased causal estimate of using a voucher on the desired outcome.

However, the decision to actually use the voucher may not be random. For example, the private school might be further than a nearby public school from the students hence, the students would prefer to attend the public school instead, even if they had won a voucher. Additionally, non-observable differences could exist which could affect the outcome as the students use the vouchers. Since the decision to attend might not be random as well as being unable to control for the unobservable variables, estimating (2) will yield a biased estimator that is unable to estimate the causal effect of using the voucher.

i.

The causal chain the authors have in mind is as follows.

$$Voucher\ Allocation\ Z_i\ (vouch0) \rightarrow UsedVoucher\ D_i\ (usesch) \rightarrow Outcome\ Y_i$$

The first requirement of an IV approach is the first stage requirement where Z_i (voucher allocation) must have a causal effect on D_i (using the voucher). As quoted in the paper, "lottery winners were 15 percentage points more likely to attend private schools" therefore, it satisfies the first stage requirement.

The second requirement is that the voucher allocation must be uncorrelated with the unobservable difference that might affect the outcome. This requirement is satisfied since the voucher allocation is conducted through a lottery which is a random process.

Lastly, the third requirement states that the impact of voucher allocation on the outcome must occur only through the students using the voucher. This requirement is also satisfied since the paper mentions "The assumption that a scholarship use dummy satisfies the exclusion restriction in an instrumental variables (IV) setup". This also largely makes sense since students who have won the vouchers are able to utilize them for only their schooling expenses but not anything else.

j.

	(1) OLS with controls	(2) IV with controls
Used School Voucher	0.171**	0.191*
	(3.14)	(2.48)
Observations	1135	1135

t statistics in parentheses
Dependent variable: Highest Grade Completed
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From columns (1) and (2) in my data generated from STATA, we can compare these values to Table 7, columns "OLS" and "2SLS", on row "Highest grade completed".

For the OLS estimate of the "Highest grade completed", the paper obtained an estimate of 0.167 and a standard error of 0.053 for the coefficient of using a school voucher. In my data from STATA, I generated an estimate of 0.171 and a standard error of 0.0544356 which are both very similar to the data in the paper. The estimate from my data is only slightly higher (0.171 – 0.167 = 0.004) than the estimate from the paper.

For the 2SLS estimate of the "Highest grade completed", the paper obtained an estimate of 0.196 and a standard error of 0.078 for the coefficient of using a school voucher. In my data from STATA, I generated an estimate of 0.191 and a standard error of 0 .0769695 which are also both very similar to the data in the paper. The estimate from my data is only slightly lower (0.191 – 0.196 = –0.003) than the estimate from the paper.

k.

Yes, there is an effect of school vouchers on the highest grade achieved. The t-statistic is determined to be 3.14 on STATA for the coefficient of using the school vouchers which is greater than 1.96. Therefore, it is statistically significant at the 0.2% level as indicated by the P-Value of 0.002. Therefore, we reject the null hypothesis of having no effect of school vouchers on highest grade achieved.

Thus, we can conclude that students who use school vouchers have on average a 0.171 greater highest grade achieved than students who do not. For example, a student without using the school voucher may have a highest grade completed of grade 10 whereas a student that used the voucher would complete grade 10.171.

l.

No, there is no effect of school vouchers on the likelihood of being in school by the time of the survey. The coefficient of using the school voucher provides a t-statistic of 1.04 on STATA, which is less than 1.96. Therefore, it not statistically significant at the 5% level. Therefore, we do not reject the null hypothesis that is having no effect of school vouchers on the likelihood of being in school.

Thus, we can conclude that while using the school voucher will increase the likelihood of being in school by 0.0222886 on average in our sample, we cannot rule out that this is due to sampling variability. Thus, the difference is not sufficient to conclude it at the population level.