WANG Liangjiayi - ECO372 Assignment 1

Student Nb: [1004405789]

a. This research conducts a randomized controlled trial to evaluate the efficiency of village-based schools on children's academic performance and participation. All villages were group into village groups first, and then the villages are randomly selected to be the treatment or control group. All treatment villages receive a village-based school and the control villages do not receive one. In conclusion, the village-based school program remarkably improves the enrollment and test scores among all children, especially for girls.

b.

(i) The level of observation in the dataset is a child who live in the villages and participate in the experiment.

(ii) There are 1,728 observations in the dataset.

(iii) The variable 'Treatment' denotes whether participants were in a village where a school was placed or not. (Treatment =1 if village group assigned to treatment; Treatment =0 if village group assigned to control)

(iv) There were 892 participants in a village where a school was placed and 836 participants in a village where a school was not placed.

(v) The variable 'f07_formal_school' records whether a student is enrolled in a formal school in Fall 2007

The variable 'f07_both_norma_total' records the total normalized test score for each child in fall 2007

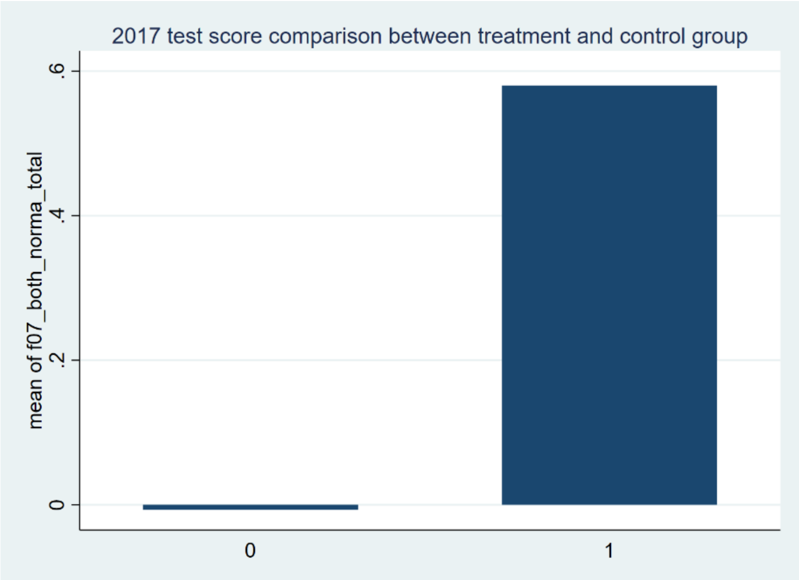
The variable 's08_formal_school' records the total normalized test score for each child in spring 2008

(vi) The variable 'f07_girl_cnt' records whether the participant is a boy or girl in 2007. (f07_girl_cnt=1 girl, f07_girl_cnt=0 boy)

The variable 's08_girls_cnt' records whether the participant is a boy or girl in 2007. (s08_girls_cnt =1 girl, s08_girls_cnt =0 boy)

c. Yes, we can interpret such differences as the effect of placing that school. The villages are randomly chosen to receive the school, the Randomized Controlled Trial holds which removes the selection bias. The difference in outcomes between groups is unbiased estimator of treatment effect. Moreover, the paper indicated that the shorter period of instruction in the village-based schools do not influence the overall treatment effect. Thus, the random implementation of schools has a causal effect on a student's academic performance.

d.



e.

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	653	-.0068562	.0389941	.996449	-.0834252	.0697129
1	721	.5798561	.039723	1.066619	.5018694	.6578427
combined	1,374	.3010183	.0289809	1.074249	.2441667	.3578699
diff		-.5867122	.0558514		-.6962755	-.4771489

```
    diff = mean(0) - mean(1)                                      t = -10.5049
Ho: diff = 0                                     degrees of freedom =     1372

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

$H_0: score\ 2007_{control} - score\ 2007_{treatment} = 0$

$H_a: score\ 2007_{control} - score\ 2007_{treatment} \neq 0$

$$T - test\ statistics: t = \frac{\overline{score\ 2007_{control}} - \overline{score\ 2007_{treatment}}}{SE(\overline{score\ 2007_{control}} - \overline{score\ 2007_{treatment}})} = -10.5049$$

$p - value = 0.0000 < 0.05$

Since p-value <0.05, we reject null hypothesis H0 that score for students in treatment group is the same for students in control group. At 5% significance level, we have evidence that there is a statistically significant difference between 2007 test score for students in treatment group ($score\ 2017_{treatment}$) and 2007 test score for students in control group ($score\ 2017_{control}$).

Thus, we conclude that the test score of children in village-based schools (treatment group) is higher than test score of children in villages where a school is not placed (control group) by 0.59 standard deviation on average.

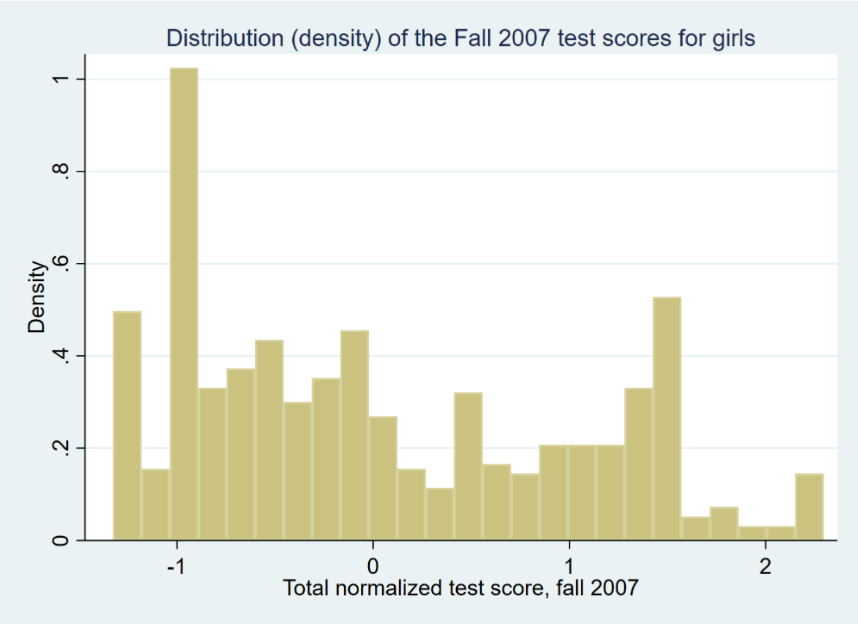
f. In fall 2007, without the control of demographic characteristics of families, the treatment groups girls' normalized test score was 0.691 standard deviations higher than the control groups girls' normalized test score on average.

In fall 2007, without the control of demographic characteristics of families, the treatment groups boys' normalized test score was 0.424 standard deviations higher than the control groups boys' normalized test score on average.

In general, girls will gain more benefit (i.e. better academic performance) from the village-based school program compared to boys.

g. For all gender students, the difference of mean between treatment group students' normalized test score and control group students' normalized test score is 0.59 and it falls between 0.691 and 0.424. Meanwhile, 0.691 and 0.424 measure the difference for girls and boys respectively. Thus, the two groups' average score should between each group's average score.

h.



The majority of density located at left side of graph (<0) which represent that the girls' normalized test score is relatively low. In other words, girls' overall academic performance is unsatisfactory. This situation may because of the social concerns like early marriage, gender discrimination, which result in girls spend less time in school education.

i.

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	310	-.3582405	.0431518	.7597654	-.4431489	-.273332
1	357	.388164	.0530896	1.003099	.2837553	.4925727
combined	667	.0412594	.0376297	.9718381	-.0326278	.1151465
diff		-.7464044	.0684148		-.8807439	-.612065

diff = mean(0) - mean(1)   t = -10.9100
Ho: diff = 0   Satterthwaite's degrees of freedom = 653.27

Ha: diff < 0   Ha: diff != 0   Ha: diff > 0
Pr(T < t) = 0.0000   Pr(|T| > |t|) = 0.0000   Pr(T > t) = 1.0000

The difference of mean (0.746) in girls' test score between treatment and control group is statistically significant since p-value <0.05. The standard error is 0.068 which smaller than the standard error in table 4 (0.13).

j. In this case, the parents' education level is confounder, which leads to the selection bias. This will affect our previous finding of question c and e because the Randomized Controlled Trial is not hold and the difference in outcomes between groups become a biased estimator of treatment effect.

k.

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	716	3.09148	.1327689	3.552653	2.830817	3.352144
1	785	3.301911	.1261432	3.534261	3.054292	3.549529
combined	1,501	3.201532	.0914603	3.543425	3.022129	3.380936
diff		-.2104304	.1831384		-.5696679	.1488071

```
    diff = mean(0) - mean(1)                                      t =  -1.1490
Ho: diff = 0                     Satterthwaite's degrees of freedom =  1484.95

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1254         Pr(|T| > |t|) = 0.2507          Pr(T > t) = 0.8746
```

f07_yrs_ed_head_cnt: Years of education of the head of the household, fall 2007

$H_0: year\_edu\ 2007_{control} - year\_edu\ \ 2007_{treatment} = 0$

$H_a: year\_edu\ \ 2007_{control} - year\_edu\ \ 2007_{treatment} \neq 0$

$T - test\ statistics: t = \frac{\overline{year\_edu\ \ 2007_{control}} - \overline{year\_edu\ \ 2007_{treatment}}}{SE(\overline{year\_edu\ \ 2007_{control}} - \overline{year\_edu\ \ 2007_{treatment}})} = -1.1490$

$p - value = 0.2507 > 0.05$

Since p-value > 0.05, we fail to reject H0 that the education level of head of the household in treatment group is the same for students in control group. At 5% significance level, we have no evidence that there is a statistically significant difference between 2007 education level of the head of the household in treatment group ($year\_edu\ 2007_{treatment}$) and 2007 education level of the head of the household in control group ($year\_edu\ 2007_{control}$).

Based on this result, the parents' education level is not a confounder and the problem occurring in question j is solved. The random assignment of schools in question c is appropriate and the estimate in question e is unbiased.

l.

```
. tab var1
```

var1	Freq.	Percent	Cum.
0	867	50.17	50.17
1	861	49.83	100.00
Total	1,728	100.00	

When randvar =0, there are 867 individuals in group-0.

When randvar =1, there are 861 individuals in group-1.

m.

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	696	.2835437	.0406869	1.073395	.2036597	.3634277
1	678	.3189569	.041309	1.075623	.2378476	.4000661
combined	1,374	.3010183	.0289809	1.074249	.2441667	.3578699
diff		-.0354132	.0579816		-.1491554	.078329

```
    diff = mean(0) - mean(1)                                      t =  -0.6108
Ho: diff = 0                     Satterthwaite's degrees of freedom =   1370.9

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2707         Pr(|T| > |t|) = 0.5415          Pr(T > t) = 0.7293
```

$H_0: score\ 2007_{group-0} - score\ 2007_{group-1} = 0$

$H_a: score\ 2007_{group-0} - score\ 2007_{group-1} \neq 0$

$$t = \frac{\overline{score\ 2007_{group-0}} - \overline{score\ 2007_{group-1}}}{SE(\overline{score\ 2007_{group-0}} - \overline{score\ 2007_{group-1}})} = -0.6108$$

$p - value = 0.5415 > 0.05$

Since the p-value > 0.05, fail to reject H0 that the fall 2007 test scores in treatment group is the same for students in control group. At 5% significance level, we have no evidence that there is a statistically significant difference between group-0 2007 test score $(score\ 2007_{group-0})$ and group-1 2007 test score $(score\ 2007_{group-1})$.

This is what I expected. RCT will remove the selection bias and the difference in outcomes between groups is unbiased estimator of treatment effect. Randomizing into the treatment and control group mechanically ensures the CIA holds. The potential outcome (test score) on average same in both groups.

n.

$\alpha = 5\% = \text{P(type 1 error)} = P(reject\ H_0|H_0\ is\ True)$

Since RCT holds, we know that the null hypothesis (i.e. no statistically difference between treatment and control group test scores) is true. If we find a statistically significant difference, we falsely reject the H0. Thus, the type 1 error occurs. The probability of making a type I error is represented by the α, which is the *p*-value (level of significance for hypothesis test). The α=0.05 represents it is acceptable to have a 5% chance that we wrongly reject the true null hypothesis. Thus, we expect there are 5% * 180 = 9 students to find a statistically significant difference.