Problem Set 3: Analysis of Experiment Data

Kai Chiu Yang (20-740-734) Jingyang Wang (19-765-270)
Xuanyi Wang (20-743-852) Tinghuan Liu (20-742-391)
Xuefei Xia (20-739-777)

March 30, 2021

1 Summary statistics

```
library(haven)
library(stargazer)
library(tidyverse)
library(ivpack)
re_d = read_dta("./STAR.dta")
summary(re_d$girl)
re_d %>%
  group_by(sck) %>%
  summarise(avg=mean(tscorek))
re_d %>%
  summarise( mean_all = mean(tscorek),
             median_all = median(tscorek),
             max_experience = max(totexpk_m)/12)
```

(a) Among 5,749 observations, there are 2,795 girls (48.6%) and 2,954 boys (51.4%).

(b) Average: 461.2
Median: 457.5
Average in treatment group (D=1): 466.0
Average in control group (D=0): 459.1

(c) 27 years (324 months).

2 OLS Regression

```
model1 <- lm(tscorek ~ sck, re_d)
print(paste("The standard deviation of test score is ", sd(re_d$tscorek)))
```

[1] "The standard deviation of test score is 36.9351445563753"

```
library("lfe")
model2 <- felm(tscorek ~ sck + as.factor(schidkn), re_d)
stargazer(model1,model2, type = "latex", header=FALSE,
          keep = c("sck"))
```

Table 1:

	Dependent variable:	
	tscorek	
	OLS	*felm*
	(1)	(2)
sck	6.913***	7.563***
	(1.058)	(0.949)
Observations	5,749	5,749
R^2	0.007	0.231
Adjusted R^2	0.007	0.221
Residual Std. Error	36.802 (df = 5747)	32.606 (df = 5669)
F Statistic	42.714*** (df = 1; 5747)	
Note:		*p<0.1; **p<0.05; ***p<0.01

(a) **Interpretation:**
holding other conditions fixed, compared with kids in regular class, kids in small class averagely get 6.913 higher in standardized tests.
Economic significance:
Change from regular class to small class is averagely associated with a increase in **tscorek** by 6.91, 18.75% of one standard deviation of **tscorek**. We think **sck** has an economically significant effect on **tscorek**.

(b) We prefer the one with school fixed effects.
Reason:
Firstly, fixed effects consider individual heterogeneity (in this case, it is the hetero-

geneity across different schools). Since treatment is randomly assigned, theoretically, school characteristics cannot correlate with treatment. However, due to sample error, they may correlate in specific sample. So it is better to include school dummy.
Secondly, school characteristics exhibit explanation power for outcome variable, test score. Including it could decrease the variation in error term and decrease standard error of estimate of β_D, which increases the precision of estimate.

3 Standard Errors

```
library(sandwich)
compute_rob_se = function(lm) {
  vcov = vcovCL(lm, type="HC1")
  se = sqrt(diag(vcov))
}

compute_clu_se = function(lm) {
  vcov = vcovCL(lm, type="HC1", cluster=~schidkn)
  se = sqrt(diag(vcov))
}

rob_se = compute_rob_se(model2)
clu_se = compute_clu_se(model2)
stargazer(model2,model2,model2,
          se = list(NULL,rob_se,clu_se),
          type = "latex", header=FALSE,
          keep = c("sck"),
          title = "Regression results with different standard errors",
          column.labels = c("Wild SE", "Robust SE", "Cluster SE"))
```

(a) We do not think it correct. That "D is independent of the error term" only means that we could get unbiased estimation of β_D, which is about consistency. And on the other hand, decision about robust standard errors and cluster robust standard errors is about efficiency of β_D. Essentially, "D is independent of the error term" cannot prevent using cluster robust standard errors or not.
Secondly, we could expect autocorrelation among error terms within same school. For example, due to better teaching quality, smarter students or other unobserved char-

Table 2: Regression results with different standard errors

	Dependent variable:		
	tscorek		
	Wild SE	Robust SE	Cluster SE
	(1)	(2)	(3)
sck	7.563***	7.563***	7.563***
	(0.949)	(0.989)	(1.821)
Observations	5,749	5,749	5,749
R^2	0.231	0.231	0.231
Adjusted R^2	0.221	0.221	0.221
Residual Std. Error (df = 5669)	32.606	32.606	32.606
Note:			*p<0.1; **p<0.05; ***p<0.01

acteristics, the overall test scores could be higher in a certain school. So, we need a standard error considering heteroscedasticity and autocorrelation at the same time. A robust standard error could only solve heteroscedasticity problem; cluster-robust standard error could solve both, and lead to unbiased estimation of $var(\beta_D)$. Hence, we prefer using cluster-robust standard error.

(b) **Reason why allowing for correlated errors within schools is important:**
Due to sharing same characteristics within same school, error terms are highly possible correlate. Considering this and using cluster-robust standard error could get a consistent estimation of $var(\beta_D)$.
Comment:
Regression results with different standard errors are in table 2. The cluster standard error in column (1), 1.821, is larger than robust standard error in column (2), 0.989, which implies a positive autocorrelation within clusters.

4 Testing whether randomization worked

```
ts_g <- lm(girl ~ sck , re_d)
ts_f <- lm(freelunk ~ sck  ,re_d)
ts_t <- lm(totexpk_m ~ sck  ,re_d)
stargazer(ts_g,ts_f,ts_t,
          type = "latex", header=FALSE)
```

Table 3:

	Dependent variable:		
	girl	freelunk	totexpk_m
	(1)	(2)	(3)
sck	−0.0004	−0.014	-5.427^{***}
	(0.014)	(0.014)	(1.988)
Constant	0.486^{***}	0.487^{***}	113.316^{***}
	(0.008)	(0.008)	(1.092)
Observations	5,749	5,749	5,749
R^2	0.00000	0.0002	0.001
Adjusted R^2	−0.0002	−0.00001	0.001
Residual Std. Error (df = 5747)	0.500	0.500	69.171
F Statistic (df = 1; 5747)	0.001	0.934	7.451^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

(a) If randomization worked, we should expect treatment group and control group is balanced, i.e., averagely, there should be no significant difference among pretreatment characteristics between small class group and regular class group.
If we run regression with pretreatment characteristics on treatment, β_D should be not statistics significant. From table 3, we could see it is not significant for **girl** and **freelunk** , while significant for **totexpk_m** at 1% significance level. Consider the significant difference between treatment group and control group in total experience of teachers, we conclude randomization might not work well.

```
ts_girl <- lm(sck ~ girl, re_d)
ts_freelunk <- lm(sck ~ freelunk,re_d)
ts_totexpk_m <- lm(sck ~ totexpk_m,re_d)
stargazer(ts_girl,ts_freelunk,ts_totexpk_m,
          type = "latex", header=FALSE)
```

(b) Regression results are in table 4. The coefficients of girl and freelunk are not significant, so we cannot reject the hypothesis that they could explain sck. However, the coefficient of totexpk_m is significant, which implies that totexpk_m could statistically significantly explain treatment status.

Table 4:

	Dependent variable:		
	sck		
	(1)	(2)	(3)
girl	-0.0004		
	(0.012)		
freelunk		-0.012	
		(0.012)	
totexpk_m			-0.0002^{***}
			(0.0001)
Constant	0.302^{***}	0.307^{***}	0.328^{***}
	(0.008)	(0.008)	(0.011)
Observations	5,749	5,749	5,749
R^2	0.00000	0.0002	0.001
Adjusted R^2	-0.0002	-0.00001	0.001
Residual Std. Error (df = 5747)	0.459	0.459	0.459
F Statistic (df = 1; 5747)	0.001	0.934	7.451^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01		

```
model_add_cov <- lm(tscorek ~ sck
                    + girl + freelunk
                    + totexpk_m, re_d)

stargazer(model_add_cov,
          type = "latex", header=FALSE,
          keep = c("sck","girl","totexpk_m",
                   "freelunk"))
```

Table 5:

	Dependent variable:
	tscorek
sck	6.912^{***}
	(1.009)
girl	6.902^{***}
	(0.927)
freelunk	-19.549^{***}
	(0.929)
totexpk_m	0.049^{***}
	(0.007)
Observations	5,749
R^2	0.098
Adjusted R^2	0.097
Residual Std. Error	35.091 (df = 5744)
F Statistic	156.008^{***} (df = 4; 5744)
Note:	*p<0.1; **p<0.05; ***p<0.01

(c) If randomization works well, treatment status should not correlate with other pretreatment characteristics, and according to omitted variable bias formula, estimate of β_D should almostly stay unchanged. In our case, the new estimate is 6.912 which is slightly lower than 2(a), 6.913. This slight change might be caused by positive correlation with *totexpk_m* and *sck* (recall 4(a)). Due to a positive bias for *totexpk_m*, the new estimate become smaller.

Parameter estimates on totexpk_m, freelunk and girl are all not causal, since they are correlated with error term.

5 Heterogeneity in Treatment Effects

```
re_dm <- re_d %>%
    mutate(sck_girl = sck*girl,
           sck_freelunk = sck*freelunk,
           sck_totexpk_m = sck*totexpk_m)
library(plm)
model_inter_g <- plm(tscorek ~ sck
                    + girl + sck_girl
                    , re_dm, index = c("schidkn"), model = "within")
model_inter_f <- plm(tscorek ~ sck
                    + freelunk + sck_freelunk
                    , re_dm, index = c("schidkn"), model = "within")
model_inter_e <- plm(tscorek ~ sck
                    + totexpk_m + sck_totexpk_m
                    , re_dm, index = c("schidkn"), model = "within")
model_inter_d <- plm(tscorek ~ sck
                    + girl + sck_girl
                    + freelunk + sck_freelunk
                    + totexpk_m + sck_totexpk_m
                    , re_dm, index = c("schidkn"), model = "within")

rob_se_5 <- list(sqrt(diag(vcovHC(model_inter_g, type = "HC1"))),
               sqrt(diag(vcovHC(model_inter_f, type = "HC1"))),
               sqrt(diag(vcovHC(model_inter_e, type = "HC1"))),
               sqrt(diag(vcovHC(model_inter_d, type = "HC1"))))
stargazer(model_inter_g,model_inter_f,model_inter_e,model_inter_d,
          type = "latex", header=FALSE,keep.stat = c("n", "adj.rsq"),
          se = rob_se_5,
          keep = c("sck","girl","sck_girl",
                   "freelunk","sck_freelunk",
                   "totexpk_m",  "sck_totexpk_m"))
```

Table 6:

	Dependent variable:			
	tscorek			
	(1)	(2)	(3)	(4)
sck	8.782***	6.880***	11.910***	12.543***
	(1.823)	(2.127)	(3.874)	(3.690)
girl	6.795***			6.915***
	(0.992)			(0.943)
sck_girl	−2.502			−3.069*
	(1.812)			(1.748)
freelunk		−19.053***		−19.034***
		(1.506)		(1.453)
sck_freelunk		1.134		1.076
		(2.882)		(2.730)
totexpk_m			0.043***	0.039**
			(0.016)	(0.016)
sck_totexpk_m			−0.038	−0.037
			(0.029)	(0.028)
Observations	5,749	5,749	5,749	5,749
Adjusted R^2	0.006	0.055	0.002	0.068
Note:				*p<0.1; **p<0.05; ***p<0.01

```
linearHypothesis(model_inter_d, test = "F", c("sck_freelunk", "sck_girl","sck_totexpk_m"
                 vcov. = vcovHC, type = "HC1")
```

Linear hypothesis test

Hypothesis: sck_freelunk = 0 sck_girl = 0 sck_totexpk_m = 0

Model 1: restricted model Model 2: tscorek ~ sck + girl + sck_girl + freelunk + sck_freelunk + totexpk_m + sck_totexpk_m

Note: Coefficient covariance matrix supplied.

Res.Df Df F Pr(>F) 1 5666
2 5663 3 1.9097 0.1257

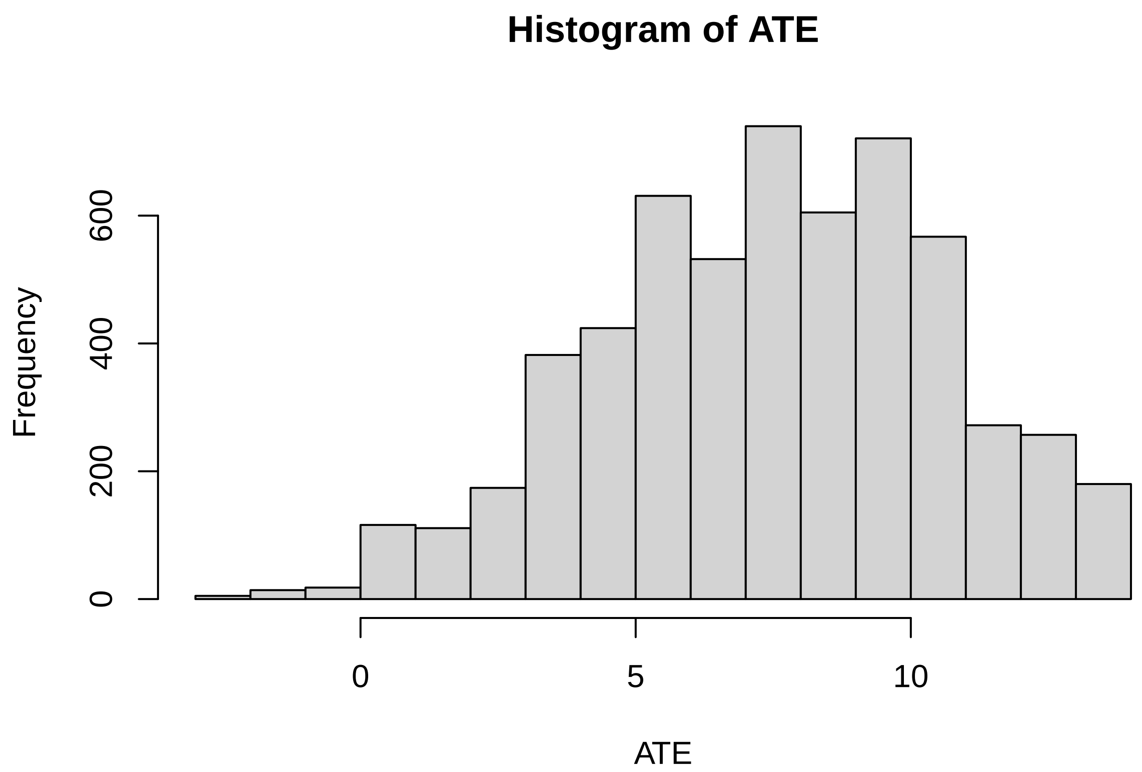
(a) **Interpretation**:
Sck: if students is a boy, without free lunch and taught by a teacher with 0 month teaching experience, he will averagely get 12.543 higher in test score in small class, compared with his counterparts in regular class. (this is our benchmark group) It is statistically significant at 1% significance level;
sck girl: compared with boy in small class, girl averagely will get 3.07 lower score in small class, and it is statistically significant at 10% significance level;
sck freelunk: compared with students who don't rely on free lunch in small class, students relying on free lunch averagely will get 1.08 more in small class, and it is not statistically significant;
sck totexpk m: one month more experience of teacher will averagely lead to 0.04 decrease in test score for students in small class.
Largest group
Given above estimation, treatment effect is largest among students who are boy, enjoying free lunch and taught by a teacher with 0 month teaching experience. The treatment effect is $12.543 + 1.076 = 13.618$.
F-test
A F-test with the joint null hypothesis that **sck_girl = sck_freelunk = sck_totexpk_m =0** shows that F(3, 78)=1.91[1], and Prob>F=0.1257. We cannot reject the null hypothesis that the treatment effect is the same for everyone.

[1]We have no idea why R doesn't show the adjusted degree of freedom, we use the adjusted one here. We also offer a stata code.

```
coe <- model_inter_d$coefficients
ATE <- coe[1:1] + coe[3:3]*re_dm$girl + coe[5:5]*re_dm$freelunk + coe[7:7]*re_dm$totexpk
mean(ATE)
```

[1] 7.484765

```
hist(ATE)
```



(b) ATE = 7.48, which is slightly lower than question (3) (7.56, without girl, freelunk and toexpk_m) but higher than question (4) (6.91, with girl, freelunk and toexpk_m).

Reason

For gender as an example. If the treatment effect is heterogeneous between boys and girls, and boys and girls are not 50 to 50. In this case, the unconditional effect estimated when not controlling gender will be different from the ATE after considering gender heterogeneity.