

# STAT 31900: Homework 3

## Econometric Methods

Robert Winter

### Table of Contents

<b>1 Dataset 1</b>	<b>1</b>
1.1 Question 1 . . . . .	1
1.2 Question 2 . . . . .	3
1.3 Question 3 . . . . .	6
<b>2 Dataset 2</b>	<b>9</b>
2.1 Question 4 . . . . .	9
<b>3 Dataset 3</b>	<b>11</b>
3.1 Question 5 . . . . .	11
3.2 Question 6 . . . . .	13
3.3 Question 7 . . . . .	15
3.4 Question 8 . . . . .	17

Throughout this document, **bolded** language signifies the problem set's instructions. My responses are provided in non-bold text.

This assignment has 8 questions plus two bonus questions.

## 1 Dataset 1

Please use CHDV 30102 webstar\_Winter2024\_rv.dta for Questions 1–3.

### 1.1 Question 1

Study 1. In the Tennessee class size study, students assigned at random to small classes in kindergarten (CLYPEK=1, which we denote with  $Z = 1$ ) were expected to stay in small classes while those assigned to regular classes in kindergarten (CLYPEK=2,3, which we denote with  $Z = 0$ ) were expected to remain in regular

classes in the subsequent years. According to the public documentation, there was no evidence of non-compliance during the kindergarten year. In Assignment 1, you found evidence of noncompliance in Grade 1 by comparing the first-grade treatment received  $CLTYPE1=1$ , which we denote with  $D = 1$ ;  $CLTYPE1=2,3$ , which we denote with  $D = 0$ ) with the kindergarten treatment assigned ( $Z$ ).

1a. When there is two-sided noncompliance (i.e., there are noncompliers in both the experimental group and the control group), Angrist, Imbens, and Rubin (1996) divide a population into four subpopulations: compliers, always-takers, never-takers, and defiers. Please define the meaning of each of these four subpopulations in the context of the current study.

In this study...

- Compliers consist of (1) students who were assigned to a small kindergarten classroom who remained in a small classroom in Grade 1, and (2) students who were assigned to a regular-sized kindergarten classroom who remained in a regular-sized kindergarten classroom in Grade 1.
- Always-Takers consist of (1) students who were assigned to a small kindergarten classroom who remained in a small classroom in Grade 1, and (2) students who were in a small classroom in Grade 1, even though they were assigned to a regular-sized classroom in kindergarten.
- Never-Takers consist of (1) students who were assigned to a regular-sized kindergarten classroom who remained in a regular-sized classroom in Grade 1, and (2) students who were in a regular-sized classroom in Grade 1, even though they were assigned to a small classroom in kindergarten.
- Defiers consist of (1) students who were in a small classroom in Grade 1, even though they were assigned to a regular-sized classroom in kindergarten, and (2) students who were in a regular-sized classroom in Grade 1, even though they were assigned to a small classroom in kindergarten.

1b. In this data set, what are the relative proportions of compliers, always-takers, and never-takers, under the assumed absence of defiers? Explain how you obtain these relative proportions.

Below, we tabulate the numbers of students in each kindergarten classroom-Grade 1 classroom pair:

	D=0	D=1
Z=0	2490	211
Z=1	106	1138

Let  $\pi_C$  denote the proportion of compliers,  $\pi_A$  the proportion of always-takers, and  $\pi_N$  the proportion of never-takers in the data.

Assuming there are no defiers in the population, all of the students who were assigned to a regular-sized kindergarten classroom but enrolled in a small Grade 1 class are always-takers. By the randomization of treatment assignment, the proportion of always-takers in the control group (students in regular-sized kindergarten classrooms) equals the proportion of always-takers in the entire population. So,  $\pi_A = \mathbb{P}(D = 1|Z = 0) = \frac{\mathbb{P}(D=1, Z=0)}{\mathbb{P}(Z=0)} = \frac{211}{2490+211} \approx 0.078$ .

Similarly, assuming there are no defiers in the population, all of the students who were assigned to a small kindergarten classroom but enrolled in a regular-sized Grade 1 class are never-takers. By the randomization of treatment assignment, the proportion of never-takers in the treated group (students in small kindergarten classrooms) equals the proportion of never-takers in the entire population. So,  $\pi_N = \mathbb{P}(D = 0|Z = 1) = \frac{\mathbb{P}(D=0, Z=1)}{\mathbb{P}(Z=1)} = \frac{106}{106+1138} \approx 0.085$ .

Finally, students who were assigned to small kindergarten classrooms and enrolled in small Grade 1 classrooms are a combination of compliers and always-takers. So,  $\mathbb{P}(D = 1|Z = 1) = \pi_C + \pi_A$ . By the randomization of treatment assignment, the proportion of always-takers in the treated group (students in small kindergarten classrooms) equals the proportion of always-takers in the control group (students in regular-sized kindergarten classrooms):  $\pi_A \approx 0.078$ . Thus,  $\pi_C = \mathbb{P}(D = 1|Z = 1) - \pi_A = \frac{\mathbb{P}(D=1, Z=1)}{\mathbb{P}(Z=1)} - \pi_A \approx \frac{1138}{106+1138} - 0.078 \approx 0.837$ .

We can “sanity check” our calculation of  $\pi_C$  as follows. Students who were assigned to regular-sized kindergarten classrooms and enrolled in regular-sized kindergarten classrooms are a combination of compliers and never-takers. So,  $\mathbb{P}(D = 0|Z = 0) = \pi_C + \pi_N$ . By the randomization of treatment assignment, the proportion of never-takers in the control group (students in regular-sized kindergarten classrooms) equals the proportion of never-takers in the treated group (students in small kindergarten classrooms):  $\pi_N \approx 0.085$ . Thus,  $\pi_C = \mathbb{P}(D = 0|Z = 0) - \pi_N = \frac{\mathbb{P}(D=0, Z=0)}{\mathbb{P}(Z=0)} - \pi_N \approx \frac{2490}{2490+211} - 0.085 \approx 0.837$ , as expected.

In summary, then,

$$\pi_C \approx 0.837$$

$$\pi_A \approx 0.078$$

$$\pi_N \approx 0.085$$

## 1.2 Question 2

In order to decide whether to use the initial treatment assignment ( $Z$ ) as an instrumental variable for  $D$  in this case for identifying the causal effect of class size reduction in Grade 1 on Grade 1 math achievement, you need to evaluate the key assumptions.

2a. In the table below, for each of the six assumptions, please insert “x” in the corresponding cells to indicate that the assumption is required for defining or for using the IV method to identify the “local average treatment effect” (LATE). Then explain in words what each assumption implies in the context of the current study.

Assumptions		The LATE
1.	Stable unit treatment value assumption (SUTVA)	X
2.	Exogeneity of $Z$	X
3.	Exclusion restriction	X
4.	Nonzero effect of $Z$ on $D$	X
5.	Monotonicity	X
6.	Constant treatment effects; or zero covariance between the effect of $Z$ on $D$ and the effect of $D$ on $Y$	

In this study...

1. SUTVA implies that each student’s Grade 1 classroom enrollment (regular-sized or small) was unaffected by the kindergarten classroom sizes of any other student. Moreover, SUTVA implies that each student’s Grade 1 math outcomes were unaffected by the kindergarten and Grade 1 classroom sizes of any other student.
2. The exogeneity of  $Z$ , sometimes referred to as the ignorability assumption, implies that  $D(z) \perp\!\!\!\perp Z$  for  $z \in \{0, 1\}$  and  $Y(z, d) \perp\!\!\!\perp Z$  for  $z, d \in \{0, 1\}$ . In words, a student’s potential Grade 1 classroom size if she is in a small kindergarten classroom and her potential Grade 1 classroom size if she is in a regular-sized kindergarten classroom have no bearing on the kindergarten classroom size that she is actually assigned to. Moreover, a student’s potential Grade 1 math scores under each possible combination of kindergarten and Grade 1 class sizes have no bearing on the kindergarten class size that he is actually assigned to.
3. The exclusion restriction assumption implies that  $Y(z, d) = Y(d)$  for  $z, d \in \{0, 1\}$ . In words, once a student’s Grade 1 classroom size has been taken into account, her kindergarten classroom size has no bearing on her Grade 1 math outcome.
4. The assumption of a nonzero effect of  $Z$  on  $D$  implies that for at least one student  $i$ ,  $\mathbb{E}[D_i(1)] \neq \mathbb{E}[D_i(0)]$ . In words, at least one student must have a different probability of enrolling in a small Grade 1 classroom if she were in a small kindergarten classroom than if she were in a regular-sized kindergarten classroom. Equivalently, at least one student in the study population must be a complier, meaning that if she is assigned to a small kindergarten classroom, she will go on to enroll in a small Grade 1 classroom, and if she is assigned to a regular-sized kindergarten classroom, she will go on to enroll in a regular-sized Grade 1 classroom.

5. The monotonicity assumption implies that no students are defiers. That is, the monotonicity assumption implies that no student would enroll in a small Grade 1 classroom if assigned to a regular-sized kindergarten classroom and enroll in a regular-sized Grade 1 classroom if assigned to a small kindergarten classroom.
6. The constant treatment effects assumption implies that each family's decision for their child's Grade 1 classroom size is uncorrelated with their expectations of their child's math achievement in either classroom size. In other words, a family's expectations about how well or poorly their child would learn in a regular-sized or small classroom are assumed to have no bearing on the class size in which they decide to actually enroll their child. Violation of this assumption would result in the causal estimand being biased, but still being identifiable.

**2b. For each of these assumptions, please explain whether the assumption can be empirically verified, why or why not, and if yes, how.**

1. SUTVA can be empirically supported in this case by examining the autocorrelation of math scores among students clustered within the same classrooms. If SUTVA is violated such that one student's presence in (or absence from) a classroom has an effect on her peers outcomes, we would expect scores to be highly correlated within classrooms. The absence of such autocorrelation would be suggestive that SUTVA is satisfied.
2. The assumption that  $Z$  is exogenous can be tested by running balance checks to evaluate whether there is balance in students' covariates and propensity scores of being treated across the two class types. If covariates and propensity scores are balanced, it is suggestive that treatment assignment cannot be predicted by an individual's potential outcomes, supporting the ignorability assumption.
3. The exclusion restriction cannot be empirically verified. The exclusion restriction requires that  $Z$  only affect  $Y$  through its intermediate effect on  $D$ . In other words, it must not be the case that  $Z$  affects  $Y$  through any additional unobserved variables  $\eta$  that are correlated with both  $Z$  and  $Y$ . But since any such variables are unobserved, it is impossible to determine whether they are correlated with our instrument or not. Even if we tried to deduce what those variables might be, there is no way to guarantee that the list we come up with is exhaustive, and it is always possible that there is still some unobserved variable that violates the exclusion restriction that we have not thought of.
4. The assumption of a nonzero average causal effect of  $Z$  on  $D$  can be easily empirically verified by simple computation. Mathematically, this assumption requires that  $\mathbb{E}[D_i(1) - D_i(0)] \neq 0$ . While we obviously cannot compare an individual's potential  $D$  outcomes under both  $Z = 1$  and  $Z = 0$  (since one of these is counterfactual), we *can* compute  $\mathbb{E}[D_i(1)|Z = 1] - \mathbb{E}[D_i(0)|Z = 0] = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$  using individuals observed values of  $D$ . Provided that the ignorability of treatment assignment assumption is satisfied, verifying that  $\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] \neq 0$  is equivalent to verifying that  $\mathbb{E}[D_i(1) - D_i(0)] \neq 0$ .
5. The monotonicity assumption cannot be empirically verified. This assumption requires that no students in the study are defiers. But to determine whether a student is a

defier, we would need to know which classroom type she would enroll in if assigned to a small classroom, *and* which classroom type she would enrollment in if assigned to a regular-sized classroom. In reality, of course, we only observe one of these two outcomes, so that we cannot verify the absence of defiers.

**2c. If an assumption cannot be empirically verified, please use theoretical reasoning to judge how plausible the assumption is in the context of this study.**

1. In this study, the exclusion restriction would require that a student's randomly assigned kindergarten class type only affect their Grade 1 math score through its effect on their Grade 1 class type. Equivalently, the exclusion restriction would imply that once we know a student's Grade 1 class type, knowing her kindergarten class type has no additional bearing on her Grade 1 math score. This is highly implausible. A first grader's quality of math education in kindergarten is obviously foundational to their math achievement in Grade 1 (and beyond). In particular, students who had minimal math achievement in kindergarten are likely to carry that struggle over to the first grade, while students who excelled in kindergarten math are likely to carry that excellence over to the first grade. So, if a student's math achievement in kindergarten is affected by their kindergarten class size, and their Grade 1 math achievement is affected by their kindergarten math achievement, then kindergarten class size undoubtedly has an impact on Grade 1 math achievement beyond that which is mediated by its effect on Grade 1 class size. In this case, the exclusion restriction is violated.
2. The monotonicity assumption, however, is highly plausible in this study. The monotonicity assumption would require that there is at least one student who, whatever her kindergarten classroom assignment, her Grade 1 classroom type is the opposite. It is difficult to imagine any child whose parents are so defiant that they make their child's enrollment decision solely on the basis of doing the opposite of what they are assigned in the Tennessee study. Indeed, if parents believe in the conventional wisdom that small classes are better for student learning, they would be unlikely to try to enroll their child in a regular-sized classroom if assigned the small treatment. And if parents believe in the conventional wisdom that being in larger classes is better for a child's socialization, it is difficult to believe that they would try to enroll their child in a small classroom if assigned the regular treatment. The implausibility of defiers in this study makes the monotonicity assumption highly plausible.

### 1.3 Question 3

**Study 2.** In each year, the participating teachers in Project STAR schools were assigned at random to teach either a small class or a regular class. Suppose that you are now interested in the average causal effect of kindergarten teacher degree on kindergartners' math achievement (`tmathssk`). Let  $G = 0$  if a kindergarten teacher had only a bachelor's degree (`hdegk = 2`); and let  $G = 1$  if a teacher had a master's degree or above (`hdegk = 3, 4, 5`). Among students with valid information on class type in kindergarten, restrict the sample to those with non-missing kindergarten math outcome (if `tmathssk != 999 & cltypek < 9`).

**3a. Were kindergartners in small classes and those in regular classes indeed equally likely to be taught by teachers with a master’s degree or above? Please analyze the data, show your evidence, and draw a conclusion.**

We restrict the data to the 5,078 children with valid information on class type in kindergarten with non-missing kindergarten math outcomes and whose teachers had non-missing educational attainment. Of the 3,526 children in regular-sized kindergarten classrooms, 1,307, or  $\frac{1307}{3526} \approx 37.1\%$ , had teachers with a master’s degree or above. On the other hand, of the 1,552 children in small kindergarten classrooms, 443, or  $\frac{443}{1552} \approx 28.5\%$ , had teachers with a master’s degree or above. As such, students in regular-sized kindergarten classrooms appear more likely to have had a teacher with a master’s degree or above. Using a  $\chi^2$  test to make this comparison rigorous, we recover a  $p$ -value of  $4.75 \times 10^{-9} < 0.001$ , confirming that the prevalence of teachers with master’s degrees or above was significantly different in small classrooms than regular-sized classrooms. In particular, students in regular-sized kindergarten classrooms were significantly more likely to have had a teacher with a master’s degree or above than students in small kindergarten classrooms. See the code output below:

```

      Z=0  Z=1
G=0 2219 1109
G=1 1307  443

```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```

data:  GZtab
X-squared = 34.289, df = 1, p-value = 4.75e-09

```

**3b. An analyst has proposed to use students’ randomized treatment assignment to a small or a regular class in kindergarten ( $Z$ ) as an instrument for identifying the average causal effect of kindergarten teacher degree ( $G$ ) on kindergarten math achievement in this case. Do you have concerns about this suggestion? Please explain why you think  $Z$  may (or may not) be a valid instrument for  $G$  in this study.**

While I appreciate the analyst’s creativity, I do have concerns about the reasonableness of kindergarten class size as an instrument for kindergarten teacher degree.

To the analyst’s credit, kindergarten class size *does* satisfy the “relevance” requirement for a good instrument. That is, as we showed in Part (a), students in regular-sized kindergarten classrooms are significantly more likely to have a teacher with a master’s degree or above than their counterparts in small classrooms, so  $\text{Cov}(G_i, Z_i) \neq 0$ . In particular,  $\text{Corr}(Z_i, G_i) \approx -0.083 \neq 0$ . Moreover, since kindergarten classroom assignment was randomized—and there was no evidence of non-compliance with this random assignment during the kindergarten year—the ignorability assumption is satisfied. After all, if a student’s assignment to a

classroom was random, then it could not have depended on the academic credentials of her potential teachers or on her potential Grade 1 math outcomes in either class size.

However, kindergarten class size most likely does *not* satisfy the exclusion restriction. To be a valid instrument for teacher degree, kindergarten class size would have to affect students' Grade 1 math outcomes *only* by means of its association with teacher degree. That is, once a student's kindergarten teacher's degree has been accounted for, that student's kindergarten class size must have no incremental impact on her Grade 1 math outcome. Conceptually, this is difficult to believe. Students in small kindergarten classrooms likely get more individualized support while learning their math fundamentals, which almost certainly carries over to better math outcomes by the end of Grade 1. Since kindergarten class size almost certainly has an effect on Grade 1 math outcomes even after taking kindergarten teacher degree into account, the former is almost certainly an invalid instrument for the latter.

**3c. Conduct the IV analysis as proposed by the analyst through two-stage least squares (you may use `ivregress 2sls` for IV analysis in Stata or the function `ivreg()` from the package `ivreg` in R). Report the analytic result. How does the result compare with simply regressing  $Y$  on  $G$ ? Please pay attention to both the point estimates and their standard errors.**

Our instrumental variables analysis yields a highly significant ( $p = 1.51 \times 10^{-4}$ ) coefficient estimate on  $G$  of  $-85.829$ . Intuitively, this implies that, on average, having a kindergarten teacher with a master's degree or above as opposed to just a bachelor's degree is associated with an 85.829-point *decrease* in Grade 1 math score. The standard error of this coefficient estimate is 22.628, which means that—assuming Grade 1 math score conditional on kindergarten teacher degree and kindergarten class size is normally distributed—we can be ~68% confident that the true effect of kindergarten teacher degree on Grade 1 math score is somewhere in the interval  $(-85.829 - 22.628, -85.829 + 22.628) = (-108.457, -63.201)$ .

A simple regression of Grade 1 math score on kindergarten teacher degree yields a very different result. Now, we recover a positive, much smaller coefficient estimate on  $G$  of roughly 1.010. However, this coefficient estimate is not statistically significant, as it bears a  $p$ -value of  $0.476 > 0.05$ . As such, under this analysis, we fail to reject the null hypothesis that kindergarten teacher degree has a nonzero effect on Grade 1 math outcomes. This can also be seen by looking at the standard error and confidence intervals associated with the coefficient estimate. In particular, the standard error is roughly 1.418, which means that—again assuming that Grade 1 math score conditional on kindergarten teacher degree and kindergarten class size is normally distributed—we can be ~68% confident that the true effect of kindergarten teacher degree on Grade 1 math outcome is somewhere in the interval  $(1.010 - 1.418, 1.010 + 1.418) = (-0.408, 2.428)$ , which includes 0.

Thus, our two models yield very distinct results: while the instrumental variables model (using a likely flawed instrument, as discussed in Part (b)) indicates that kindergarten teacher degree has a large, highly significant negative effect on Grade 1 math outcomes, the simple regression model indicates that the effect of kindergarten teacher degree on Grade 1 math outcomes is statistically indistinguishable from 0.



## 2 Dataset 2

Please use CHDV 30102 `webstar_gkclasssize.dta` for Question 4.

### 2.1 Question 4

**Study 3.** In Project STAR, there is information about the actual class size (`gkclasssize`, denoted by  $D$ , for each kindergartner. The actual size of a small class ranged from 12 to 17 students while that of a regular class (with or without an aide) ranged from 15 to 28 students. The research question is about the average causal effect of reducing class size on kindergarten math achievement (`gkmathss`, denoted by  $Y$ ).

4a. Statistician 1 suggests that the data be viewed as a fuzzy regression discontinuity design with a class size of 16 as the cutoff because the experimenters had initially designated kindergartners attending classes with fewer than 16 students to be members of the experimental group and the rest to be members of the control group. Statistician 1 therefore proposes a three-step analysis: (1) Fit a regression discontinuity model for estimating the effect of class type (small vs. regular, denoted by  $Z$ ) on  $Y$ ; (2) fit a regression discontinuity model for estimating the effect of class type on  $D$ ; (3) compute the ratio of the above two results. Do you agree with this suggestion? Why or why not?

I have to disagree with Statistician 1's suggestion. On the one hand, Statistician 1 *is* on the right track with his thinking: this situation involves an arbitrary cutoff of 16 students distinguishing small classes from regular-sized ones, which suggests that a regression discontinuity design (RDD) might be useful. However, his proposal essentially involves instrumenting class size with itself, just as a binary variable instead of an ordinal one. So if Statistician 1 is right that there is some endogeneity problem impeding our ability to determine the effect of class size on math achievement, instrumenting class size with a binary indicator of class size is certainly not going to alleviate this endogeneity problem.

Statistician 1's suggestion also betrays a misunderstanding of why arbitrary cutoffs are helpful for RDD analyses. Typically, in an RDD analysis, being just above or below a cutoff is associated with a decisive difference in treatment. For example, in Doyle et al., 2010<sup>1</sup>, researchers leveraged the fact that premature newborns just under an arbitrary cutoff weight receive extra care at many hospitals. In the present study, the 16-student cutoff *classifies* students' classrooms, but doesn't actually involve a substantive change in those students' treatments. This further undermines Statistician 1's suggestion.

4b. Statistician 2 suggests using the randomly assigned kindergarten class type (small vs. regular, denoted by  $Z$ ) as the instrumental variable for identifying the average causal effect of reducing the actual class size (`gkclasssize`, denoted

---

<sup>1</sup>Almond D, Doyle JJ, Kowalski AE, Williams H. ESTIMATING MARGINAL RETURNS TO MEDICAL CARE: EVIDENCE FROM AT-RISK NEWBORNS. *Q J Econ.* 2010 May 1;125(2):591-634. doi: 10.1162/qjec.2010.125.2.591. PMID: 20634927; PMCID: PMC2903901.

by *D*) on kindergarten math achievement. Do you agree with this suggestion? Why or why not?

While I appreciate Statistician 2's suggestion, I think it is somewhat unnecessary to perform the analysis she is interested in. Generally, instrumental variables are used when there is some selection or noncompliance problem, where individuals are self-selecting into a certain condition or ignoring their randomly assigned treatment. As a result of selection or noncompliance problems, the individuals who receive treatment are, on average, somehow fundamentally different from the individuals who don't receive treatment, making it difficult to discern the effect of the treatment on an outcome from the effects of the groups' fundamental differences.

In this case, however, we know that kindergartners were randomly assigned to different classroom types, and we know that there was no evidence of noncompliance with those assignments, at least during the kindergarten year. Since Statistician 2 is interested in the effect of kindergarten class size on kindergarten math achievement, and there is no evidence of noncompliance with students' randomly assigned treatment conditions, there is no need to use an instrumental variables analysis! That is, Statistician 2's suggestion attempts to solve a problem that doesn't exist in the first place, making it unnecessary.

**4c. If you follow Statistician 2's suggestion and conduct a two-stage least squares analysis, what is the analytic result? What is the estimated average effect of reducing class size by 10 students on kindergarten math achievement? What is the effect size?**

Call:

```
ivreg(formula = gkmathss ~ gkclasssize | Z, data = gkclasssize)
```

Residuals:

Min	1Q	Median	3Q	Max
-203.06	-32.78	-4.50	28.94	147.78

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	507.0953	3.7587	134.911	< 2e-16 ***
gkclasssize	-1.0693	0.1825	-5.859	4.92e-09 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	5869	16533.377	<2e-16 ***
Wu-Hausman	1	5868	2.864	0.0906 .
Sargan	0	NA	NA	NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.57 on 5869 degrees of freedom  
Multiple R-Squared: 0.00555, Adjusted R-squared: 0.005381  
Wald test: 34.32 on 1 and 5869 DF, p-value: 4.924e-09

Statistician 2's instrumental variables analysis yields a highly significant ( $p = 4.92 \times 10^{-9}$ ) coefficient estimate on  $D$  of approximately  $-1.069$ . Intuitively, this implies that, on average, a one-student increase in a kindergartner's class size causes a 1.069-point decrease in her kindergarten math score. The standard error of this coefficient estimate is approximately 0.183, which means that—assuming kindergarten math score conditional on kindergarten class size and assigned kindergarten class type is normally distributed—we can be ~68% confident that the true effect of increasing kindergarten class size by one student on kindergarten math outcome is somewhere in the interval  $(-1.069 - 0.183, -1.069 + 0.183) \approx (-1.252, -0.886)$ . Scaling up our coefficient estimate by a factor of 10, we estimate that a ten-student increase in a kindergartner's class size causes a 10.69-point decrease in her kindergarten math score.

We estimate the effect size using Glass's delta, calculating as the coefficient estimate on  $D$  divided by the standard deviation of kindergarten math scores among students in Statistician 2's proposed control group, students assigned to regular-sized classrooms. Here, Glass's delta equals approximately  $-0.023$ , which means that a one-student increase in a kindergartner's class size causes her math score to decrease by approximately 0.023 standard deviations with respect to the control group. A 95% confidence interval for this effect size, calculated using the Wald method, is approximately  $(-0.031, -0.015)$ .

### 3 Dataset 3

**Please use CHDV 30102 ECLSK98\_class\_size\_rv.dta for Questions 5–8.**

In the analyses that follow, in addition to filtering the data to Grade 1 students attending public schools (as requested), we also filter the data to exclude students with missing values for `C4R2MSCL` (Grade 1 math outcome, or  $Y$ ) or `A4CLSIZE` (Grade 1 class size, or  $D$ ), since these are treated as an outcome variable and an intermediate outcome variable, respectively. For simplicity, we also filter the data to exclude students with missing values for the instrument, `s4anumch` (public school enrollment, or  $Z$ ) (even though this invokes assumptions about the randomness of the missingness of such data). We also treat race/ethnicity as a six-factor variable, as in Assignment 2 (with categories for “White,” “Black,” “Hispanic,” “Asian,” “Indigenous or Native Americans,” and “Other”). For the X covariates `B4YRSTC` (teacher years of experience), `w1sesl` (socioeconomic status), and `C1R2MSCL` (math achievement at kindergarten entry) we create indicator variables for missingness and replace missing values with the means of the observed values.

#### 3.1 Question 5

**Study 4.** To address the research question about the average causal effect of class size reduction in Grade 1 on Grade 1 math achievement (`C4R2MSCL`), we

have previously analyzed the Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K) in Assignment 2. Even though the ECLS-K data were non-experimental, an analyst suggests that one may use public school enrollment (`s4anumch`, a continuous measure denoted as  $Z$ ) as an instrumental variable for Grade 1 class size (`A4CLSIZE`, a continuous measure denoted as  $D$ ) while making covariance adjustment for student and teacher characteristics ( $X$ ) including gender, race, SES, and Grade 1 teacher's teaching experience at the same time. The analysis is restricted to Grade 1 students attending public schools (`if s4pupri == 1`).

**5a. Do you have any concerns about this suggestion? Please explain why you think  $Z$  may (or may not) be a valid instrument for  $D$  in this study.**

While I think that this analyst's suggestion is on the right track, I ultimately think that  $Z$  would *not* be a valid instrument for  $D$  in this study. On the one hand,  $Z$  almost certainly satisfies the "relevance" condition for an instrument. That is, a good instrument should be correlated with the variable it is instrumenting. In this case, it is very likely that school enrollment is correlated with Grade 1 class size, with larger schools most likely having larger classes, and smaller schools most likely having smaller classes.

However, an instrumental variable should also satisfy the exclusion restriction, which requires that the instrument only affect the outcome by means of its impact on the variable it is instrumenting. In this case, the exclusion restriction would require that public school enrollment only affect a student's Grade 1 math achievement through its effect on that student's Grade 1 class size. This is highly implausible. For example, it is possible that, in addition to having larger classes on average, larger schools also have more resources to afford teaching aides or after-school tutoring programs. These potential perks of attending a larger school would mean that school enrollment could affect math achievement through channels above and beyond affecting class size, violating the exclusion restriction and invalidating the instrument.

**5b. If the schools that would increase Grade 1 class size in response to an increase in the enrollment are also the ones in which an increase in class size would be particularly harmful to Grade 1 students' math learning, do you anticipate that the IV analysis suggested by the analyst would likely produce a biased result? If so, would the bias be positive or negative? Please explain your reasoning.**

In this case, we would expect the IV analysis to produce a result with positive bias. Here, the students who are most likely to be in regular-sized classes (those at growing or large schools) are precisely the students who perform worse in regular-sized classes. So, instrumenting class size with school enrollment essentially introduces selection bias, where there is a fundamental difference between the students who receive the regular-sized class treatment and those who do not. In particular, since regular-sized classrooms are disproportionately full of students whose learning suffers in those very classrooms, regular-sized classrooms look *worse* for math achievement compared to small classrooms than they actually are. Equivalently, small classrooms look *better* for math achievement compared to regular classrooms than they

actually are. This means that the estimated effect of class size on Grade 1 math achievement would carry positive bias.

### 3.2 Question 6

In the first part of this analysis, regress  $Y$  on  $D$  with covariance adjustment for  $X$ . Record the coefficient estimate for the observed class size. In the second part of the analysis, regress  $D$  on  $Z$  with covariance adjustment for  $X$  and save the predicted class size  $\hat{D}$ ; then regress  $Y$  on the predicted class size  $\hat{D}$  with covariance adjustment for  $X$  and record the estimated coefficient for the predicted class size  $\hat{D}$ .

6a. Report the estimates obtained from the above two analyses.

First, we perform a “naive” regression analysis on the model  $Y = \beta \cdot D + X\alpha + \varepsilon$ . As shown below, we recover a non-significant ( $p = 0.509$ ) coefficient estimate on  $D$  of  $\hat{\beta} \approx 0.026$ , which would imply that, holding all  $X$  covariates fixed, changes to a student’s Grade 1 class size have no effect on her math achievement.

Call:

```
lm(formula = C4R2MSCL ~ A4CLSIZE + as.factor(GENDER) + as.factor(race6) +
    w1sesl_impt + w1sesl_miss + B4YRSTC_impt + B4YRSTC_miss,
    data = eclks)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-69.698	-8.973	-0.486	8.543	50.644

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	57.87789	0.84926	68.151	< 2e-16 ***
A4CLSIZE	0.02602	0.03937	0.661	0.509
as.factor(GENDER)2	-1.23055	0.28954	-4.250	2.16e-05 ***
as.factor(race6)Asian	-3.43350	0.61976	-5.540	3.10e-08 ***
as.factor(race6)Black	-7.50721	0.43747	-17.161	< 2e-16 ***
as.factor(race6)Hispanic	-4.65657	0.42207	-11.033	< 2e-16 ***
as.factor(race6)Indigenous	-6.55354	0.83257	-7.872	3.89e-15 ***
as.factor(race6)Other	-4.00685	0.91691	-4.370	1.26e-05 ***
w1sesl_impt	6.36594	0.20891	30.473	< 2e-16 ***
w1sesl_miss	-5.01447	0.55008	-9.116	< 2e-16 ***
B4YRSTC_impt	-0.01218	0.01422	-0.857	0.392
B4YRSTC_miss	5.09174	1.12512	4.526	6.10e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.12 on 9531 degrees of freedom  
Multiple R-squared: 0.1792, Adjusted R-squared: 0.1783  
F-statistic: 189.2 on 11 and 9531 DF, p-value: < 2.2e-16

Now, we perform an instrumental variables analysis broken out into two sequential steps. First, we estimate the model  $D = \gamma \cdot Z + X\alpha + v$ , and then use the resulting estimates  $\hat{D}$  to estimate the model  $Y = \beta^{IV} \cdot \hat{D} + X\alpha + \varepsilon$ . As shown below, we now recover a statistically significant ( $p = 0.010 \leq 0.01$ ) coefficient estimate on  $\hat{D}$  of  $\hat{\beta}^{IV} \approx 0.515$ . This implies that, holding all X covariates fixed, a one-student increase in a student's Grade 1 class size causes a 0.515-point increase in her end-of-year math score.

Call:

```
lm(formula = C4R2MSCL ~ Dhat + as.factor(GENDER) + as.factor(race6) +
    wlsesl_impt + wlsesl_miss + B4YRSTC_impt + B4YRSTC_miss,
    data = eclks)
```

Residuals:

Min	1Q	Median	3Q	Max
-69.555	-8.991	-0.489	8.548	50.151

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	48.12087	3.98172	12.085	< 2e-16 ***
Dhat	0.51529	0.19901	2.589	0.00963 **
as.factor(GENDER)2	-1.31338	0.29132	-4.508	6.61e-06 ***
as.factor(race6)Asian	-3.52906	0.62072	-5.685	1.34e-08 ***
as.factor(race6)Black	-7.79896	0.45253	-17.234	< 2e-16 ***
as.factor(race6)Hispanic	-4.82001	0.42694	-11.290	< 2e-16 ***
as.factor(race6)Indigenous	-6.25827	0.84058	-7.445	1.05e-13 ***
as.factor(race6)Other	-4.27756	0.92294	-4.635	3.62e-06 ***
wlsesl_impt	6.16698	0.22340	27.606	< 2e-16 ***
wlsesl_miss	-5.03660	0.54997	-9.158	< 2e-16 ***
B4YRSTC_impt	-0.02080	0.01462	-1.423	0.15486
B4YRSTC_miss	5.24035	1.12631	4.653	3.32e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.12 on 9531 degrees of freedom  
Multiple R-squared: 0.1798, Adjusted R-squared: 0.1788  
F-statistic: 189.9 on 11 and 9531 DF, p-value: < 2.2e-16

6b. Under what conditions would the result from the second analysis contain less bias than that from the first?

The first analysis, which uses a standard OLS approach, yields a biased estimate of  $\beta$  when  $\text{Cov}(D, \varepsilon) \neq 0$ , or intuitively, when unobserved variables that are correlated with Grade 1 math achievement are also correlated with Grade 1 class size. In the language of the potential outcomes framework, the OLS approach yields a biased estimate of  $\beta$  when  $D \not\perp\!\!\!\perp Y(d)$  for  $d \in \mathbb{N}$ , or intuitively, when a student's Grade 1 class size is somehow related to her potential math achievement in various clas

However, if  $Z$  is a “relevant” instrument for  $D$  in the sense that  $\text{Cov}(D, Z) \neq 0$ , and moreover,  $Z$  satisfies the exogeneity requirements  $\text{Cov}(Z, v) = 0$  and  $\text{Cov}(Z, \varepsilon) = 0$ , then the estimate of  $\beta^{IV}$  from the instrumental variables analysis is unbiased estimate of the effect of  $D$  on  $Y$ . Intuitively, if Grade 1 class size is correlated with public school enrollment, and public school enrollment only affects Grade 1 math outcomes by means of its effect on Grade 1 class sizes, then our coefficient estimate from the second analysis overcomes the problems in the first analysis and is unbiased.

Thus, our second analysis yields an estimate of the effect of Grade 1 class size on Grade 1 math achievement that is less biased than that of the first analysis if  $\text{Cov}(D, \varepsilon) \neq 0$ , but  $\text{Cov}(D, Z) \neq 0$ ,  $\text{Cov}(Z, v) = 0$ , and  $\text{Cov}(Z, \varepsilon) = 0$ . Notably, if  $\text{Cov}(D, \varepsilon) = 0$ , so that there is no endogeneity problem in the first place, then the OLS estimate of the effect of  $D$  on  $Y$  is unbiased, and both analyses above would be equally good in terms of unbiasedness.

### 3.3 Question 7

**Now proceed with a two-stage least squares analysis. Regress  $D$  on  $Z$  in the first stage controlling for  $X$ ; and simultaneously regress  $Y$  on  $D$  in the second stage controlling for  $X$ , where the coefficient for  $D$  is denoted as the IV estimand  $\beta_1^{IV}$ .**

**7a. Write down this pair of regression models. What is the sample estimate of  $\beta_1^{IV}$  along with its standard error,  $t$  statistic, and  $p$ -value?**

We estimate the following model using two-stage least squares:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1^{IV} \cdot \hat{D}_i + \beta_2 \cdot \text{Female}_i + \beta_3 \cdot \text{Asian}_i + \beta_4 \cdot \text{Black}_i + \beta_5 \cdot \text{Hispanic}_i + \beta_6 \cdot \text{Indigenous}_i \\ &\quad + \beta_7 \cdot \text{OtherRace}_i + \beta_8 \cdot \text{SES\_imputed}_i + \beta_9 \cdot \text{SES\_missing}_i + \beta_{10} \cdot \text{TeacherYoE\_imputed}_i \\ &\quad + \beta_{11} \cdot \text{TeacherYoE\_missing}_i + \varepsilon_i \\ D_i &= \alpha_0 + \alpha_1 \cdot Z_i + \alpha_2 \cdot \text{Female}_i + \alpha_3 \cdot \text{Asian}_i + \alpha_4 \cdot \text{Black}_i + \alpha_5 \cdot \text{Hispanic}_i + \alpha_6 \cdot \text{Indigenous}_i \\ &\quad + \alpha_7 \cdot \text{OtherRace}_i + \alpha_8 \cdot \text{SES\_imputed}_i + \alpha_9 \cdot \text{SES\_missing}_i + \alpha_{10} \cdot \text{TeacherYoE\_imputed}_i \\ &\quad + \alpha_{11} \cdot \text{TeacherYoE\_missing}_i + v_i \end{aligned}$$

where the  $\hat{D}_i$ 's in the first equation are the fitted values yielded by the estimation of the second equation.

As shown below, we estimate  $\hat{\beta}_1^{IV} \approx 0.515$ , which implies that a one-person increase in a student's Grade 1 class size causes a 0.515-point increase in her Grade 1 math score. This

coefficient estimate has a standard error of approximately 0.201, which means that, assuming Grade 1 math score conditional on all of our covariates is normally distributed, we can be ~68% confident that the true effect of Grade 1 class size on Grade 1 math score is somewhere in the interval  $(0.515 - 0.201, 0.515 + 0.201) \approx (0.315, 0.716)$ . The  $t$  statistic for this coefficient estimate is approximately 2.568, and the corresponding  $p$ -value is approximately 0.010, which means that our estimate of the effect of Grade 1 class size on Grade 1 math achievement is statistically significant at the  $\alpha = 0.05$  level.

Call:

```
ivreg(formula = C4R2MSCL ~ A4CLSIZE + as.factor(GENDER) + as.factor(race6) +
      w1sesl_impt + w1sesl_miss + B4YRSTC_impt + B4YRSTC_miss |
      s4anumch + as.factor(GENDER) + as.factor(race6) + w1sesl_impt +
      w1sesl_miss + B4YRSTC_impt + B4YRSTC_miss, data = eclks)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-68.8896	-9.1511	-0.4707	8.7251	51.0890

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	48.12087	4.01516	11.985	< 2e-16 ***
A4CLSIZE	0.51529	0.20068	2.568	0.0103 *
as.factor(GENDER)2	-1.31338	0.29377	-4.471	7.88e-06 ***
as.factor(race6)Asian	-3.52906	0.62594	-5.638	1.77e-08 ***
as.factor(race6)Black	-7.79896	0.45633	-17.091	< 2e-16 ***
as.factor(race6)Hispanic	-4.82001	0.43052	-11.196	< 2e-16 ***
as.factor(race6)Indigenous	-6.25827	0.84764	-7.383	1.67e-13 ***
as.factor(race6)Other	-4.27756	0.93070	-4.596	4.36e-06 ***
w1sesl_impt	6.16698	0.22527	27.376	< 2e-16 ***
w1sesl_miss	-5.03660	0.55459	-9.082	< 2e-16 ***
B4YRSTC_impt	-0.02080	0.01475	-1.411	0.1583
B4YRSTC_miss	5.24035	1.13577	4.614	4.00e-06 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	9531	388.01	<2e-16 ***
Wu-Hausman	1	9530	6.29	0.0122 *
Sargan	0	NA	NA	NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.24 on 9531 degrees of freedom

Multiple R-Squared: 0.1659, Adjusted R-squared: 0.165

Wald test: 186.8 on 11 and 9531 DF, p-value: < 2.2e-16



**7b.** Suppose that all the assumptions required for the IV method are met when conditioning on  $X$ . According to your analytic result in 6a, what is the estimated average effect of reducing class size by 10 students on first-grade math achievement in U.S. public schools. What is the effect size estimate?

Scaling up our coefficient estimate in Question 6a (which is identical to our coefficient estimate in Question 7a) by a factor of 10, we estimate that a ten-student increase in a U.S. public-schooled first-grader's class size causes a 5.15-point increase in her Grade 1 math score, on average. We estimate the effect size using Cohen's  $d$ , again calculated as  $\hat{\beta}_1^{IV}$  divided by the pooled standard deviation of Grade 1 math scores. Here, Cohen's  $d$  equals approximately 0.033, which means that a one-student increase in a first-grader's class size causes his math score to increase by approximately 0.033 standard deviations.

### 3.4 Question 8

As we did in the second assignment, we consider a Grade 1 class of no more than 18 students (i.e., less than or equal to 18) to be relatively “small” and a class of 19 or more students to be “regular.”

**8a.** Statistician 1 proposes a difference-in-differences (DID) analysis. The first difference is the average change in math achievement from the end of kindergarten (C2R2MSCL) to the end of Grade 1 (C4R2MSCL) for students attending small classes; the second is the corresponding change on average for those attending regular classes. Please write down the DID model, clarify the key assumptions, and explain why the ATT can be identified through this DID analysis when the identification assumptions hold. Report the analytic result.

We consider the following DID model

$$MathScore_{it} = \beta_0 + \beta_1 \cdot SmallClass_i + \beta_2 \cdot Post_t + \beta_3 \cdot SmallClass_i \times Post_t + \varepsilon_{it}$$

where:

- $SmallClass_i$  is an indicator equal to 1 if Student  $i$ 's Grade 1 classroom had 18 students or fewer, and equal to 0 if Student  $i$ 's Grade 1 classroom had 19 students or more, and
- $Post_t$  is an indicator equal to 1 if the corresponding math score is from the end of Grade 1, and equal to 0 if the corresponding math score is from the end of kindergarten.

$\beta_3$  identifies the ATT if the average change in math scores among students in small Grade 1 classes would have been the same as the average change in math scores among students in regular-sized Grade 1 classes if the students in small classes had been in regular-sized classes instead. That is, but for being in small classrooms, the trend in students' math scores would have been parallel to the math score trend of students who were not in small classrooms. If this assumption is satisfied, then the group of students in regular-sized classrooms can be

treated not just as a “comparison” group, but essentially as a “control” group, so that the trend in their math scores is precisely what the trend in the treated students’ math scores would have been but for the treatment.  $\beta_3$ , the difference between the end-of-Grade 1 score changes between students in the treated and control groups, therefore gives the causal effect of the treatment, thus identifying the ATT.

Assuming the parallel trends identification assumption holds, our estimate of the ATT is equal to the estimated coefficient on the interaction of  $SmallClass_i$  and  $Post_i$ :  $\hat{\beta}_3 \approx 0.161$ . If this estimate were statistically significant, it would imply that being in a small Grade 1 classroom causes a 0.161-point increase in a student’s Grade 1 math score compared to if she were in a regular-sized Grade 1 classroom. However,  $\hat{\beta}_3$  has a standard error of approximately 0.443 and a corresponding  $p$ -value of  $0.717 \gg 0.05$ . As such, our estimate of the ATT is *not* statistically significant, meaning that there is *not* statistical evidence that being in a small Grade 1 classroom has any effect on Grade 1 math achievement, at least under this analysis.

Call:

```
lm(formula = MathScore ~ q8Z * Post, data = did_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-63.378	-8.709	-1.100	7.612	66.301

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	30.9198	0.1660	186.220	<2e-16 ***
q8Z	-0.6912	0.3136	-2.204	0.0275 *
Post	23.4585	0.2344	100.063	<2e-16 ***
q8Z:Post	0.1607	0.4428	0.363	0.7166

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.72 on 19021 degrees of freedom

(61 observations deleted due to missingness)

Multiple R-squared: 0.4235, Adjusted R-squared: 0.4234

F-statistic: 4658 on 3 and 19021 DF, p-value: < 2.2e-16

**8b. Statistician 2 suggests that an improvement could be made to the above DID analysis by conditioning on pretreatment measures including child gender, race, SES, as well as math achievement at kindergarten entry (C1R2MSCL) through linear covariance adjustment. Please write down the modified DID model and clarify the modified identification assumptions. Report the analytic results.**

We now consider the following DID model with covariance adjustment:

$$\begin{aligned}
\text{MathScore}_{it} = & \beta_0 + \beta_1 \cdot \text{SmallClass}_i + \beta_2 \cdot \text{Post}_t + \beta_3 \cdot \text{SmallClass}_i \times \text{Post}_t + \beta_4 \cdot \text{Female}_i \\
& + \beta_5 \cdot \text{Asian}_i + \beta_6 \cdot \text{Black}_i + \beta_7 \cdot \text{Hispanic}_i + \beta_8 \cdot \text{Indigenous}_i + \beta_9 \cdot \text{OtherRace}_i \\
& + \beta_{10} \cdot \text{SES\_imputed}_{it} + \beta_{11} \cdot \text{SES\_missing}_{it} + \beta_{12} \cdot \text{MathKAutumn\_imputed}_i \\
& + \beta_{13} \cdot \text{MathKAutumn\_missing}_i + \varepsilon_{it}
\end{aligned}$$

Now,  $\beta_3$  identifies the ATT if the average change in math scores among students in small Grade 1 classes—*conditional* on students' gender, racial/ethnic heritage, socioeconomic status (as well as whether socioeconomic status was missing from the data), and math achievement at kindergarten entry (as well as whether such data was missing)—would have been the same as the average change in math scores among students in regular-sized Grade 1 classes—*conditional* on the same covariate values—if the students in small classes had been in regular-sized classes instead. This assumption essentially means that events that are contemporaneous with treatment (e.g., a change in national math achievement standards) would have the same affect on students in small and regular-sized classes, provided that those students have the same covariate values. Moreover, while gender, race/ethnicity, and math achievement at kindergarten entry are all time-invariant, it is plausible that socioeconomic status could change over time. For  $\beta_3$  to identify the ATT, we must now also assume that socioeconomic status changes in the same way for students in regular-sized and small kindergarten classrooms from the end of kindergarten to the end of Grade 1.

Assuming that these identification assumptions hold, our estimate of the ATT is now equal to  $\hat{\beta}_3 \approx 0.191$ . If this estimate were statistically significant, it would imply that being in a small Grade 1 classroom causes a 0.191-point increase in a student's Grade 1 math score compared to if she were in a regular-sized Grade 1 classroom. However,  $\hat{\beta}_3$  has a standard error of approximately 0.308 and a corresponding  $p$ -value of  $0.535 \gg 0.05$ . As such, our estimate of the ATT is still *not* statistically significant, meaning that there is still *not* statistical evidence that being in a small Grade 1 classroom has any effect on Grade 1 math achievement, even after adjustment for certain covariates.

Call:

```
lm(formula = MathScore ~ q8Z * Post + as.factor(GENDER) + as.factor(race6) +
    wisesl_impt + wisesl_miss + C1R2MSCL_impt + C1R2MSCL_miss,
    data = did_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-74.818	-5.674	-0.580	5.152	57.532

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	10.945548	0.250341	43.723	< 2e-16 ***
q8Z	0.027500	0.218915	0.126	0.900036

```

Post                23.461419    0.163152 143.801 < 2e-16 ***
as.factor(GENDER)2   -0.693354    0.138554  -5.004 5.66e-07 ***
as.factor(race6)Asian    1.342384    0.299379   4.484 7.37e-06 ***
as.factor(race6)Black   -3.365276    0.209818 -16.039 < 2e-16 ***
as.factor(race6)Hispanic -1.069751    0.203372  -5.260 1.46e-07 ***
as.factor(race6)Indigenous -1.463188    0.399987  -3.658 0.000255 ***
as.factor(race6)Other   -1.504263    0.439765  -3.421 0.000626 ***
wlsesl_impt          1.616710    0.105991  15.253 < 2e-16 ***
wlsesl_miss          -1.417642    0.264473  -5.360 8.41e-08 ***
C1R2MSCL_impt         1.021844    0.008816 115.907 < 2e-16 ***
C1R2MSCL_miss         0.185550    0.274710   0.675 0.499404
q8Z:Post              0.190952    0.308151   0.620 0.535482
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 9.545 on 19011 degrees of freedom
(61 observations deleted due to missingness)
Multiple R-squared:  0.721, Adjusted R-squared:  0.7208
F-statistic: 3778 on 13 and 19011 DF,  p-value: < 2.2e-16

```

**8c. In the above modified DID analysis, what would justify a decision to further include school enrollment (`s4anumch`) as an additional baseline covariate? Please provide your reasoning.**

School enrollment could be included as an additional covariate in our DID analysis under two conditions.

Firstly, as described in Part (b),  $\beta_3$  identifies the ATT under covariance adjustment if the average change in math scores among students in small Grade 1 classes would have the same as the average change in math scores among students in regular-sized Grade 1 classes if the students in small classes had been in regular-sized classes instead, *conditional* on common covariate values between students in the two classroom types. This naturally requires that students in each classroom type have counterparts in the other classroom type with shared covariate values. Thus, it would be appropriate to include `s4anumch` as an additional covariate if there is overlap in the school enrollments of students in small classrooms and regular-sized classrooms. In other words, it would be *inappropriate* to include `s4anumch` as a covariate if students in small Grade 1 classrooms attend schools whose total enrollments are systematically smaller than students in regular-sized Grade 1 classrooms.

Secondly, if `s4anumch` were to be treated as a time-varying covariate, it must be the case that growth in enrollment at students in small classrooms' schools from kindergarten to Grade 1 was the same as growth in enrollment at students in regular-sized classrooms' schools. In other words, it would be *inappropriate* to include `s4anumch` as a (time-varying) covariate if the schools of students in regular-sized classrooms grow at a faster (or slower) rate than the schools of students in small classrooms. In these data, it appears that `s4anumch` is treated as time-invariant, making it a student's school enrollment during a particular year. If `s4anumch`

is treated as time-invariant by the analyst, this second concern is irrelevant, but if not, it is important to bear in mind.