Impact of Class Size on Test Scores
By Adam Smith

I.   Cover Sheet

	Name	**Description**
Name of your dataset	*caschools.csv*	*Data on school characteristics and test performance for 420 school districts in California from 1998-99.*
Dependent variable	*testscr*	*Average reading and writing score*
Independent variable	*str*	*Student-teacher ratio*
Control variables	*high_comp_stu**	*Binary variable that takes value 1 if computers per student are above the median, and 0 otherwise*
	meal_pct	*Percent of students qualifying for reduced-price lunch*

*I will create this variable using *comp_stu (computers per student).*

II. Introduction

In this project, I will study how class sizes affect the performance of elementary school students. This is an important question as it informs policymakers of the value of mandating maximum class sizes. Since substantial costs are associated with reducing class sizes across schools, quantifying its benefits can help policymakers make more informed decisions. This question is also significant as it pertains to equity in education. Larger class sizes may disproportionately impact schools in under-resourced communities, indicating that if reducing class sizes proves effective in enhancing quality, such measures could play a crucial role in equalizing educational opportunities.

III. Data and Variables Description

To answer this question, I use data from 420 school districts in California for the years 1998 and 1999. This dataset is from the California Standardized Testing and Reporting (STAR) program and was originally obtained from the California Department of Education. To measure performance, I use the average test score on reading and math (*testscr*), which is my dependent variable. The primary independent variable is the student-teacher ratio (*str*).

In addition, I will also consider how the availability of computers and the percentage of students qualifying for school lunch (*meal_pct*) interact with test performance and the student-teacher ratio. To measure the availability of computers, I construct a binary variable, *high_comp_stu,* that takes the value 1 if computers per student are above the median and 0 otherwise.

To make meaningful comparisons, I limit the sample to only K-8 schools. The final data contains 359 observations.

IV. Summary Statistics

Table 1 presents the summary statistics for all the variables of interest. From this table, we can see that the mean test score for reading and math across schools is approximately 653. The lowest test score in the data is 605.5, while the highest score is 706.75. Furthermore, the median test score is 653.55, which indicates that the distribution of test scores is not skewed.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
testscr	359	652.90	18.68	605.55	653.55	706.75
str	359	19.71	1.88	14.00	19.78	25.80
high_comp_stu	359	0.50	0.50	0	1	1
meal_pct	359	45.93	26.78	0.00	44.68	100.00

On average, schools have 19.71 students per teacher, with a range of student-teacher ratios varying from 14 to 25.80. This suggests that there is a notable difference in class sizes across schools. The average of the indicator for high computer availability is 0.5, showing that the number of computers is above the median for 50% of the schools. We can also see that, on average, 45.93% of students in a school qualify for school lunches, with a range of 0-100.

V.   Relationship between Test Score and Student-Teacher Ratio

I start by exploring the relationship between test scores and the student-teacher ratio. Figure 1 presents a scatterplot with test scores on the y-axis and student-teacher ratio on the x-axis. From the figure, I can see that schools with a higher student-teacher ratio tend to have lower test scores and vice versa. I confirm this pattern by computing the correlation between the two variables, which is calculated to be -0.20.

Figure 1: Scatterplot for Test Score and Student-Teacher Ratio
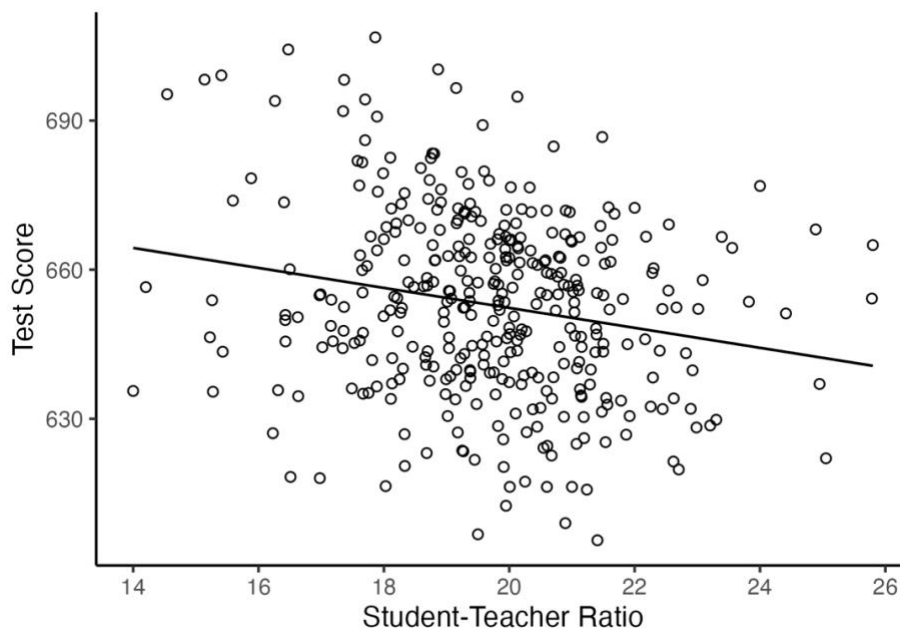
VI.   Exploring Other Factors

After examining the relationship between test scores and student-teacher ratio, I now explore how other school and student characteristics influence this relationship. Specifically, I investigate the impact of two variables: the availability of computers and the percentage of students eligible for school lunch.

Table 2 displays the correlation between these two additional variables and test scores and the student-teacher ratio. As expected, I observe a positive correlation between the

availability of computers and test scores. Conversely, the correlation between computers and the student-teacher ratio is negative, suggesting that schools with higher student-teacher ratios often have fewer computers. Hence, failing to account for the number of computers may lead us to overestimate the negative impact of higher student-teacher ratios on test scores.

Table 2: Correlation Matrix

```
=================================================
               testscr  str   meal_pct high_comp_stu
-------------------------------------------------
testscr           1     -0.20  -0.86       0.23
str             -0.20    1      0.11      -0.26
meal_pct        -0.86   0.11    1         -0.19
high_comp_stu   0.23    -0.26  -0.19       1
-------------------------------------------------
```

This suggests that schools with higher student-teacher ratios may be situated in districts with limited resources, which could potentially lead to lower test scores. Consequently, neglecting to account for the percentage of students qualifying for school lunches or other markers of socioeconomic factors may result in an overestimation of the detrimental effect of student-teacher ratios on test scores.