Empirical Project 2
Do Smaller Classes Improve Test Scores? Evidence from a Regression Discontinuity Design
Helen (Yingying) Huang

1. **Explain why a simple comparison of test scores in small classes versus large classes would not measure the causal effect of class size. Would this simple comparison likely be biased upwards or biased downwards relative to that true causal effect? Explain.**

A simplistic comparison of test scores from small classes versus large classes would not accurately measure the causal effect of class size on learning outcomes for several reasons. First, such a comparison could be confounded by factors such as teacher quality, where more effective educators might be assigned to smaller classes, thereby inflating the perceived benefit of class size alone. Additionally, socio-economic and demographic factors could skew the comparison; for instance, schools in affluent areas might have both smaller class sizes and higher-scoring students, not directly due to class size but because of other resources.

In terms of potential bias direction, this simple comparison is likely to be biased upwards. If smaller classes are indeed perceived as better and thus attract more skilled teachers or are implemented in schools with more resources, then the advantage in test scores would be overestimated when attributed solely to class size. This overestimation results from conflating the class size effect with the effects of other unmeasured or uncontrolled variables that correlate with both the class size and the outcome variable, i.e., test scores.

2. **(To answer this and the next question, read [Chetty et al. 2011]). How did the Tennessee STAR experiment overcome this problem? What did it find?**

The Tennessee STAR experiment addressed the issue of measuring the causal effect of class size by randomly assigning a cohort of 11,571 students and their teachers to different classrooms within their schools from kindergarten to third grade. This random assignment to small classes (15 students on average) or large classes (22 students on average) across 79 schools between 1985 and 1989 allowed for a controlled experiment that minimized selection bias and confounding variables, thus providing a more accurate measure of the causal impact of class size on educational outcomes (Chetty et al., 2011).

Chetty et al. (2011) evaluated the long-term impacts of the STAR experiment by linking experimental data to administrative records. They found that kindergarten test scores are highly correlated with later outcomes such as earnings at age 27, college attendance, home ownership, and retirement savings. The study documented experimental impacts showing that students assigned to small classes were more likely to complete high school and take the SAT or ACT college entrance exams. The research also indicated that these students were less likely to be arrested for a crime, contributing to the literature on the long-term impacts of class size in education (Chetty et al., 2011).

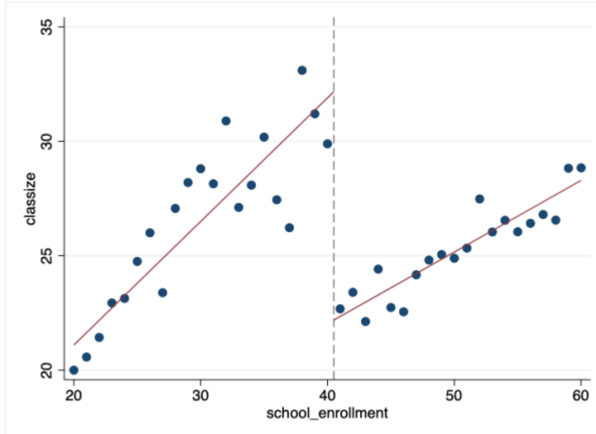
3. What is a binned scatter plot? Explain how it is constructed.

A binned scatter plot is created by dividing the data for an independent variable into equal-sized bins, calculating the mean of the dependent variable for each bin, and then plotting these means against the midpoints of the bins. It includes a line of best fit from a regression analysis to show the overall relationship, and the R^2 value indicates how well the independent variable explains the variance in the dependent variable (Chetty et al., 2011).

4. Graphical regression discontinuity analysis, focusing on the 40 student school enrollment threshold. See Table 2a and 2b for more guidance.

   a. Draw a binned scatter plot to visualize how class size changes at the 40 student school enrollment threshold. Display a linear or quadratic regression line based on what you see in the data.

Figure 1 Class Size at 40 Student Enrollment Threshold in the 5th Grade



b. Draw binned scatter plots to visualize how math and verbal test scores change at the 40 student school enrollment threshold. Display a linear or quadratic regression line based on what you see in the data.

Figure 2 Average Composite Math Test Scores at 40 Student Enrollment Threshold in the 5th Grade

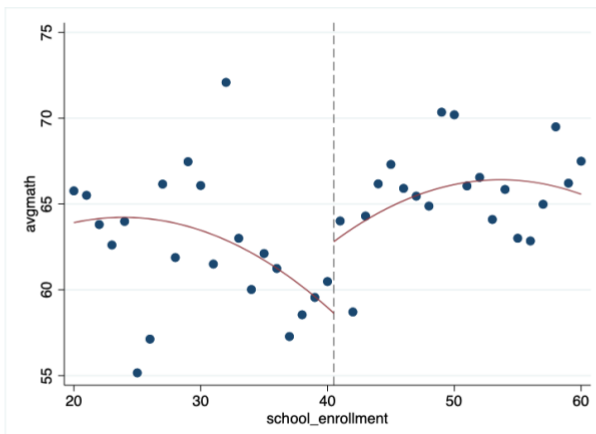
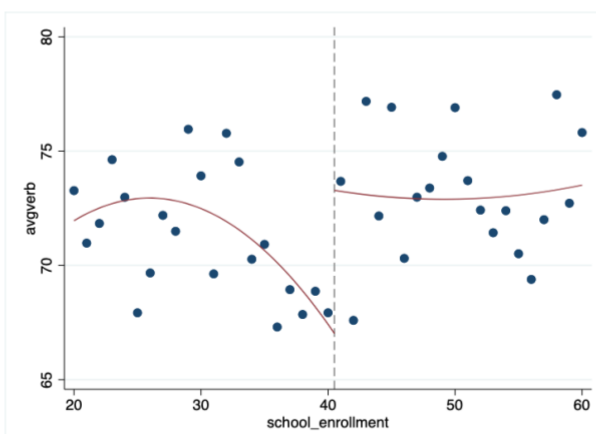


Figure 3 Average Composite Verbal Test Scores at 40 Student Enrollment Threshold in the 5th Grade



c. **Draw binned scatter plots to test whether (i) the percent of disadvantaged students, (ii) the fraction of religious schools, and (iii) the fraction of female students evolve smoothly across the 40 student school enrollment threshold. Display a linear or quadratic regression line based on what you see in the data.**

Figure 4 Proportion of Disadvantaged Students at 40 Student Enrollment Threshold in the 5th Grade

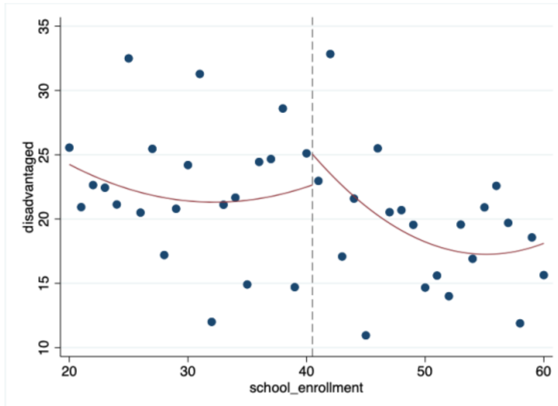


Figure 5 Prevalence of Religious Schools at 40 Student Enrollment Threshold in the 5th Grade

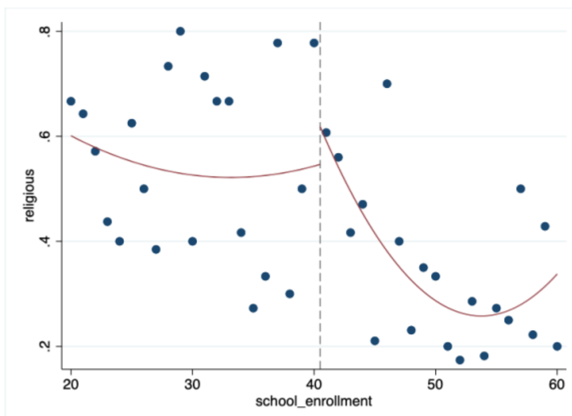
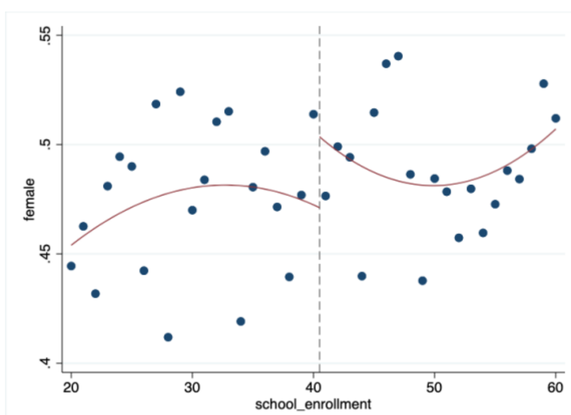
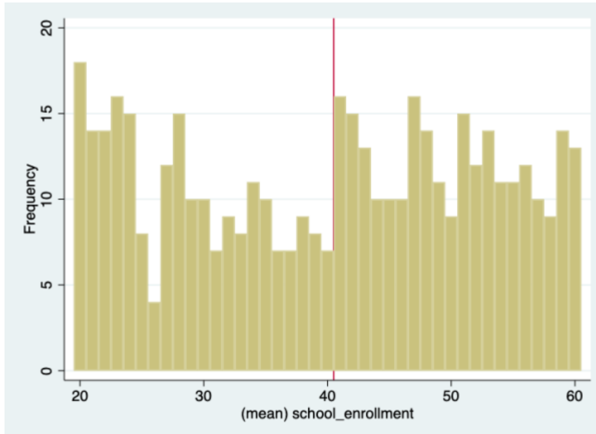


Figure 6 Fraction of Female Students at 40 Student Enrollment Threshold in the 5th Grade



d. **Produce a histogram of the number of schools by total school enrollment. Note that you must collapse the data by *school* to produce this graph.**

Figure 7 Distribution of Schools by the 5th Grade Enrollment Size



5. **Regression analysis. Run the regressions that correspond to your three graphs in 4a and 4b to quantify the discontinuities that you see in the data. In estimating these regressions, use all the observations with school enrollment less than 80. Report a 95% confidence interval for each of these estimates. See Table 2a and 2b for more guidance.**

The regression discontinuity design employs a cutoff of 40 students for 5th-grade enrollment to differentiate between schools with fewer than 40 students and those with more. This approach yields insightful findings, as analyzed in Stata. Specifically, schools with 5th-grade enrollments just above the 40-student cutoff experience a significant reduction in class size, indicated by an 'above40' coefficient of -11.00075 (95% CI: -13.32322, -8.67828). This suggests that schools slightly above the 40-student enrollment mark in the 5th grade typically have smaller classes. The pattern for average math scores is more complex; the scores exhibit a slight increase above the enrollment threshold (coefficient: 3.428896, 95% CI: 0.2639069, 6.593886). However, the confidence interval indicates some uncertainty regarding this estimate. Moreover, as the 5th-grade enrollment in schools nears 40, math scores tend to decrease but increase once the threshold is crossed, as shown by the 'x_above40' interaction term. Verbal scores demonstrate a similar pattern with a threshold-related increase, albeit with less statistical certainty (coefficient: 2.628131, 95% CI: -0.3109008, 5.567164). The analysis underscores the 40-student threshold as a critical juncture in the 5th grade, where class sizes decrease and the dynamics of academic outcomes shift, emphasizing the nuanced impact of class size on educational performance.

6. **Recall that any quasi experiment requires an identification assumption to make it as good as an experiment. What is the identification assumption for regression discontinuity design? Explain whether your graphs in 4c and 4d are consistent with that assumption.**

The identification assumption for a regression discontinuity design requires that potential outcomes vary smoothly across the cutoff, meaning any discontinuity in the observed outcomes can be attributed to the treatment. The graphs for non-treatment variables—percent of disadvantaged students, fraction of religious schools, and fraction of female students—show no discontinuities at the 40-student threshold, supporting this assumption. Additionally, the lack of bunching in the histogram of school enrollments suggests no manipulation of enrollment sizes around the cutoff, reinforcing the assumption that the observed discontinuities in outcomes are attributable to the treatment effect rather than other unobserved variables.

7. **(To answer this question, read [Angrist and Lavy (1999)]). If all schools followed the class size rule exactly as described in Angrist and Lavy (1999), how much would you expect class size to change at the 40 student enrollment threshold? Explain why the actual change in class size that you see in the data is less than this.**

In accordance with Maimonides' rule, as described by Angrist and Lavy (1999), a strict application would result in class sizes increasing in a one-to-one ratio with enrollment until reaching 40 students. At the point of enrolling the 41st student, the class would be split, leading to an average class size of approximately 20.5 students. This means that at the threshold of 40 students, a sharp decrease in class size would be expected—specifically, a reduction to about half when the 41st student is enrolled.

However, the actual change in class size at the 40-student threshold observed in the data is less than this theoretical expectation. This lesser change can be attributed to the fact that schools generally maintain average class sizes smaller than the maximum allowed by Maimonides' rule due to practical considerations, such as the availability of teachers, classroom space, or a preference for smaller classes to enhance educational outcomes, even before reaching the threshold.

8. **Suppose your school superintendent is considering a reform to reduce class sizes in your school from 40 to 35. Use your estimates above to predict the change in math and verbal test scores that would result from this reform.**

   ***Hint*: divide the RD estimate of the change in test scores by the change in number of students per class at the threshold.**

   Reducing class sizes from 40 to 35 students is anticipated to yield an increase in test scores, based on the above regression discontinuity estimates. Math scores are expected to increase by an average of 3.43 points, and verbal scores by 2.62 points at the 40-student threshold. The estimated increase per student is obtained by dividing these figures by the discontinuity in class size, which is 11 students.

   For math scores, the per-student increase is therefore 3.43 divided by 11, approximately 0.312 points. For verbal scores, it is 2.62 divided by 11, roughly 0.238 points. Multiplying these per-student increases by the reduction in class size of 5 students, the forecasted improvements in test scores are 1.56 points for math and 1.19 points for verbal. These calculations assume a consistent linear relationship between class size and test scores.

9. Suppose you are asked for advice by another school that is considering reducing class size from 20 to 15 students – a 5 unit reduction as above. Would you feel confident in making the same prediction as you did above about the impacts this change will have? Why or why not?

In advising a school considering reducing class size from 20 to 15 students, it would be imprudent to directly extrapolate predictions from the effects observed at the 40-student threshold due to potential non-linearity of class size effects. While Angrist and Lavy (1999) provide evidence of benefits at higher thresholds, educational research suggests that the impact of class size reductions may diminish as class sizes become smaller (Hanushek, 1999; Krueger, 2003). Therefore, without specific evidence pertaining to smaller class sizes, confidence in the prediction should be reserved, and recommendations should be grounded in context-specific empirical analysis.

10. Compare your estimates in 8 with the estimates from (i) the Tennessee STAR experiment (Chetty et al. 2011) and (ii) data from Sweden (Fredriksson et al. 2013) discussed in lecture. Give two reasons that your estimates might differ from those of these other studies.

In comparing the estimates from question 8 with those from the Tennessee STAR experiment (Chetty et al., 2011) and the findings from Sweden (Fredriksson et al., 2013), it is evident that results may differ due to variations in context and methodology. The STAR experiment was a randomized trial in the U.S. with significant reductions in class size, showing long-term academic and economic benefits (Chetty et al., 2011). Meanwhile, the Swedish study, which also observed long-term benefits of smaller classes on cognitive and noncognitive abilities and even on earnings, used a quasi-experimental design based on a maximum class size rule, revealing positive effects persisting into adulthood (Fredriksson et al., 2013).

11. **Chetty et al. (2011) show that being assigned to a smaller class in Kindergarten raises Kindergarten test scores, but has little impact in later grades. Does this "fade out" effect mean that class size doesn't really matter in the long run? Why or why not?**

The "fade out" effect of early test score improvements noted by Chetty et al. (2011) does not negate the long-term significance of class size. Although the direct impact on later test scores may wane, the study indicates that smaller class sizes in kindergarten are associated with substantial long-term benefits, such as higher earnings and better college attendance. These outcomes suggest that early educational interventions, while not always evident in test scores, can enhance non-cognitive skills and other factors that contribute to success in adulthood. Hence, class size in early education remains an influential factor with enduring implications beyond initial academic performance (Chetty et al., 2011).

12. **Given the evidence above, would you encourage your hometown school to reduce class size by hiring more teachers if the goal is to maximize students' long-term outcomes (e.g., college attendance rates, earnings)? Explain clearly what other data you would need to make a scientific recommendation and how you would use that data.**

To offer scientific advice to a local school in Baltimore on the impact of hiring more teachers to reduce class sizes on maximizing long-term student outcomes, it's crucial to analyze local data. Such data should encompass cost-benefit analyses, current class sizes, student-to-teacher ratios, teacher quality, and information on potential trade-offs with other educational initiatives. Additionally, it's important to take into account the community-specific educational goals, values, and resource limitations. Empirical evidence from studies by Chetty et al. (2011) and Fredriksson et al. (2013) indicates that smaller classes can yield long-term benefits. Nevertheless, these advantages must be weighed against the unique resource constraints and educational goals of the local community. A thorough analysis of these factors will lead to a recommendation designed to optimize long-term student outcomes.

References

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics, 114*(2), 533–575. https://doi.org/10.1162/003355399556061

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660. https://doi.org/10.1093/qje/qjr041

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics, 128*(1), 249-285. https://doi.org/10.1093/qje/qjs048

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143–163. https://doi.org/10.2307/1164297

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal, 113*(485), F34–F63. https://doi.org/10.1111/1468-0297.00098