Empirical Project 2

Steven Espinoza

3/07/2019

Measuring the Impact of Class Size on Student Outcomes

As Professor Chetty has discussed in his lectures, there are a variety of variables we can observe to understand their causal impact on student outcomes in the future. Specifically, we've spent a lot of time talking about teacher quality and class size, and the necessity for experimental and quasi-experimental methods when trying to understand their causal impacts on student outcomes in the long-run.

Experimental methods are needed to answer such questions because simply comparing outcomes (test scores, for example) in small and large classes fail to capture the *causal* effect of class size; they would instead just illustrate the correlation that might exist the two variables. In other words, simply comparing small classrooms to large classrooms would ignore the fact that students in schools with small classroom sizes will generally be from higher-income backgrounds and have other advantages, and would thus bias our estimates upwards relative to the true causal effect that class size has on student outcomes.

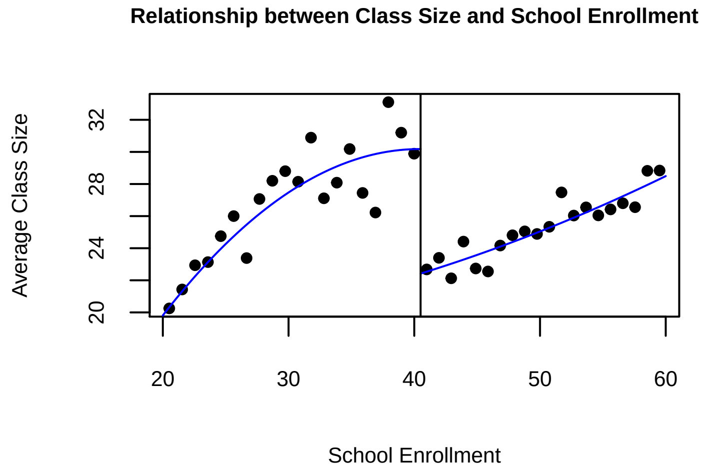
In order to minimize this bias, then, the best option would be to create a scenario where students and teachers from a variety of backgrounds are randomly assigned to classrooms of different sizes. This is what was done with the STAR Experiment, conducted across 79 Tennessee classrooms in the late 1980s. In this experiment, more than 11,000 students and teachers were randomly assigned to classrooms within their schools from kindergarten to the third grade. Under this random assignment, the study's main finding was that children in smaller classes did about 5 percentiles better on tests than children in larger classes–a statistically significant result that suggested a causal relationship between class size and test scores. Moreover, though it found no causal relationship between class size and future earnings, the study does reveal a significant relationship between kindergarten test scores and future earnings. The same was true for rates of college attendance.

Visualizing Causal Effects: Binned Scatter Plots and Regression Discontinuity

To visualize the results used in Chetty et al. (2011), we will used binned scatter plots as shown at the end of the paper. Binned scatter plots attempt to make scatterplots easier to read by first splitting the dataset into distinct "bins," taking the average value of the dependent variable, and then representing that average as a single dot. For example, in Figure I at the end of the Chetty et al. paper (2011), each dot shows average earnings for students in 5-percentile bins. This further helps explain why the R^2 value shown at the bottom-right corner is only 0.05, as opposed to a much higher R^2 value that a scatterplot like this should yield if each dot represented all the data in the dataset.

Class Size

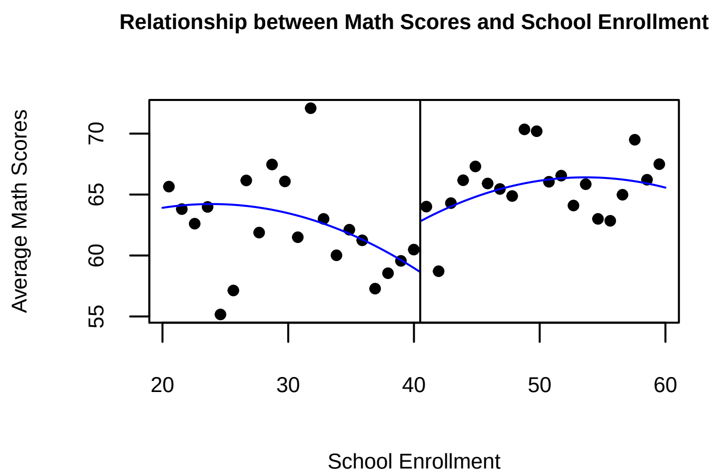
The binned scatterplot below shows how class size changes at the 40 student school enrollment threshold.

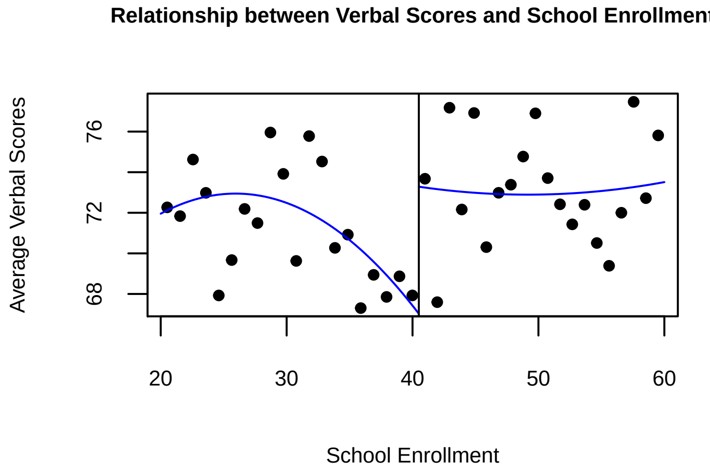


From the figure above, it is easy to see that class size increases with school enrollment. However, once Maimonedes' Rule kicks in at 40 students, the average class size suddenly decreases sharply, by roughly 10 students. The slope nonetheless stays positive.

Math and Verbal Test Scores

The following binned scatter plot shows how math and reading scores change at the 40 student school enrollment threshold.

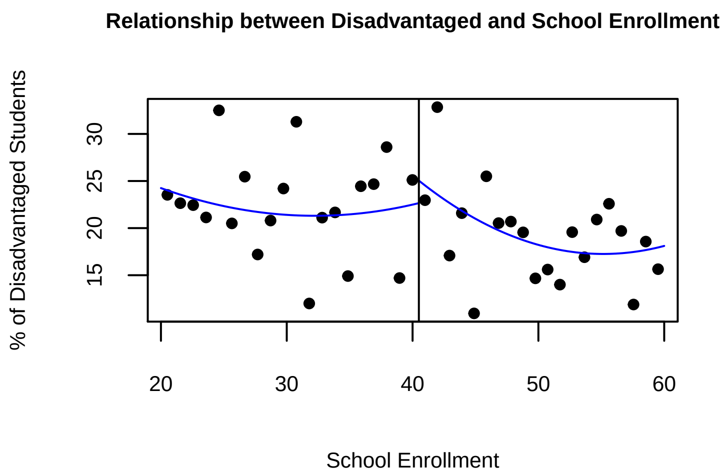




The two graphs above reveal a similar pattern: As school enrollment increases, there is steady decrease in average math and reading scores. However, once school enrollment reaches the 40 student mark, there is a small jump that suggests an increase in student test scores. Note that the jump seems to be significantly wider for the verbal test scores compared to the math test scores.

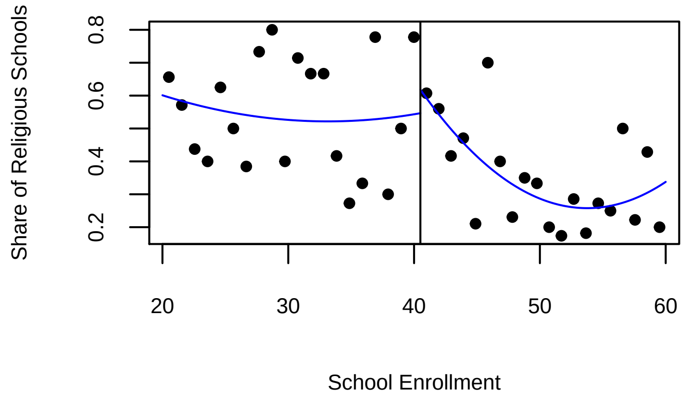
Demographic Variables: Looking at Disadvantaged Students, Religion, and Gender

The following binned scatter plots show the relationship between school enrollment and three different independent variables: the percent of disadvantaged students, the fraction of religious schools, and the fraction of female students.



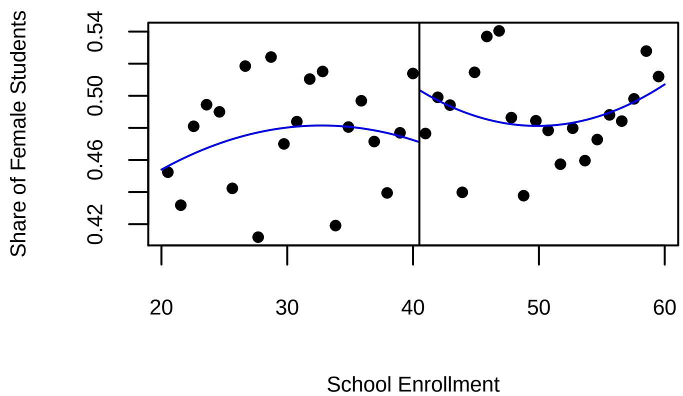
This graph seems to suggest that, in general, there are more disadvantaged students in schools with less than 40 students enrolled than in schools with more than 40 students enrolled. The exception is mostly for schools with around 40 to 45 students enrolled, where one might expect to have more disadvantaged students.

Relationship between Religious Schools and School Enrollment



The binned scatter plot above suggests that religious schools generally have much less students enrolled. Among schools with less than 40 students, about 55% of them are religious; however, among those with 45 students, only about 40% of them are religious. This share mostly decreases as the number of students enrolled increases.

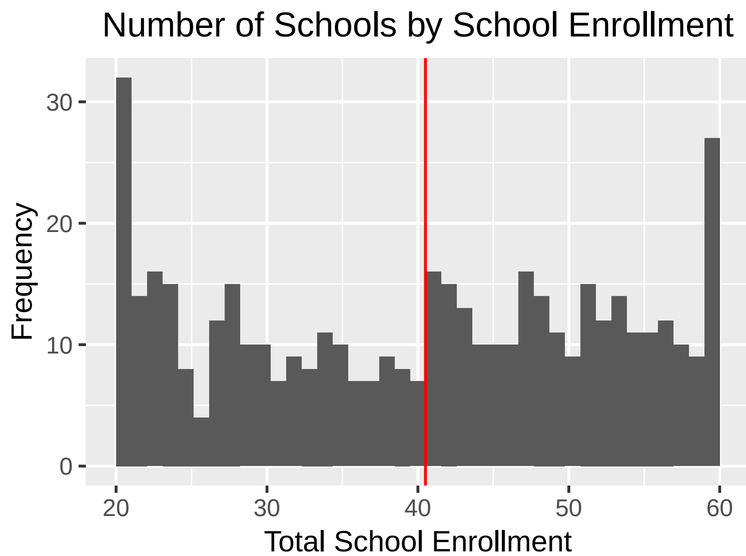
Relationship between Female Students and School Enrollment



A simple takeaway from the binned scatter plot above shows that the schools with more than 40 students are more likely to have a larger share of female students than male students. Among those schools with only 20 students enrolled, on average only 46% of the student population is female. However, among those schools with 60 students, on average about 50% of the student population is female. I'd question whether this difference is significant: Even at the discontinuity, the jump is only about 2 percentage points.

Frequency of Schools

The following histogram shows the number of schools by total school enrollment.



This histogram shows a sudden jump in the number of schools that enroll more than forty students. Moreover, it seems that there are more schools that enroll more than 40 students when compared to schooles below forty students.

Regression Analysis

In order to quantify the relationships we saw above, we will run regressions in order to quantify the jump at the discontinuity.

The following three regression outputs correspond to the first three binned scatter plots show earlier with school enrollment on the x-axis and class size, math test scores, and verbal test scores on the y-axis. The dependent variable for each of these regressions is shown in the column on the far-right labeled "outcome."

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	33.05	1.13	29.15	0.00	30.82	35.27	730.00	classize
2	above40	-11.00	1.18	-9.30	0.00	-13.32	-8.68	730.00	classize
3	x	0.67	0.06	11.80	0.00	0.56	0.78	730.00	classize
4	x_above	-0.34	0.06	-5.67	0.00	-0.46	-0.22	730.00	classize

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	60.83	1.37	44.54	0.00	58.15	63.51	730.00	avgmath
2	above40	3.43	1.61	2.13	0.03	0.26	6.59	730.00	avgmath
3	x	-0.19	0.08	-2.36	0.02	-0.35	-0.03	730.00	avgmath
4	x_above	0.30	0.09	3.31	0.00	0.12	0.47	730.00	avgmath

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	69.82	1.30	53.79	0.00	67.27	72.37	730.00	avgverb
2	above40	2.63	1.50	1.76	0.08	-0.31	5.57	730.00	avgverb
3	x	-0.16	0.08	-2.13	0.03	-0.32	-0.01	730.00	avgverb
4	x_above	0.23	0.08	2.72	0.01	0.06	0.39	730.00	avgverb

The first output above (classize) suggests that when the number of enrolled students reaches 40, there is a sudden decrease in class size by about 11 students on average. The 95% confidence interval suggests that, on average, this true decrease ranges anywhere between around 9 to 13 students. This change is interesting because, if all schools followed Maimonides' Rule exactly, we would expect class sizes to decrease by roughly half once the 40 student threshold is reached (a 19 student decrease). The reason why this decrease from the data is smaller might be because classes are splitting at levels below 40 students.

The second and third outputs above (avgmath and avgverb) suggests that when the number of enrolled students reaches 40, there is a sudden increase in test scores for math and reading at 3.43 and 2.63 points respectively. Though this jump is statistically significant for math scores (95% conf. interval: 0.26, 6.59), the same cannot be said at the 0.05 alpha level for verbal scores (95% conf. interval: -0.31, 5.57) because 0 is contained within the interval.

If these estimates are accurate at 40 students, we might expect that if a new rule where imposed where the cutoff was instead at 35 students, math scores would increase by about 3 points and verbal scores would increase by about 2.30 points. However, if a rule were imposed to reduce class size from 20 to 15 students, we might not see as large a difference because a class size of 20 is arguable already a pretty small size for a classroom.

We can conclude that these effects are causal based on our identification assumption, which in this case is the fact all other determinants of test scores are balanced on each side of the cutoff. Angrist and Lavy (1999) write that this might be assumed based on the fact that "parents do not selectively exploit Maimonides' rule so as to place their children in schools with small classes." We can evaluate the validity of this assumption by making sure that observable characteristics are similar on both sides of the cutoff. Whether this is actually true based on our graphs from 4(c) and 4(d) is questionable.

Conclusion

Our estimates on the change in math and verbal test scores (at 3.43 and 2.63 percentile changes) seem smaller than the STAR Experiment and the Swedish experiment we saw in class. The STAR experiment suggested an increase of about 4.81 percentiles, while the Swedish experiment saw an increase in test scores by about 8 percentiles. One potential reason for this might be because the Israeli students we've looked at in this dataset might already be scoring higher on tests than those in Tennessee or in Sweden, so a higher jump is less likely. From this logic, another reason for the smaller change in test scores might come from the supply-side: Teacher quality in Israel might already be much better than that of Tennessee or Sweden, so they might be better at teaching classes of, say, 30 students, while teachers in Sweden or Tennessee might not be able to handle as large a capacity.

According to Chetty et al. (2011), "the impacts of early childhood education fade out on test scores but re-emerge in adulthood." I would take this to mean that class size still matters in the long-run, but indirectly. As mentioned earlier, there is a relationship between class size and kindergarten test scores, as well as between test scores and factors such as college attendance and future earnings, even though the relationship between class size and these two factors is statistically insignificant. In other words, though class size may still matter in the long-run, other variables such as the quality of your teacher in high school could still change these expected outcomes of students.

As discussed in the lectures, both class size and teacher quality are important in predicting student outcomes. While aiming for smaller classes alone might be a good policy proposal, a better one would aim for both smaller classes *and* improving teacher quality. A good experiment that might quantify these effects would be one that analyzes the heterogeneity in smaller classrooms among teacher quality, and seeing how this heterogeneity plays a predictive role in student outcomes.