Empirical Project 2 | ECON 270

Group Members:

Shalin Luitel

Jahnavi Kalyan

Sweta Nanadakumar

Arnold Gyateng

Do Smaller Classes Improve Test Scores?

Evidence from a Regression Discontinuity Design

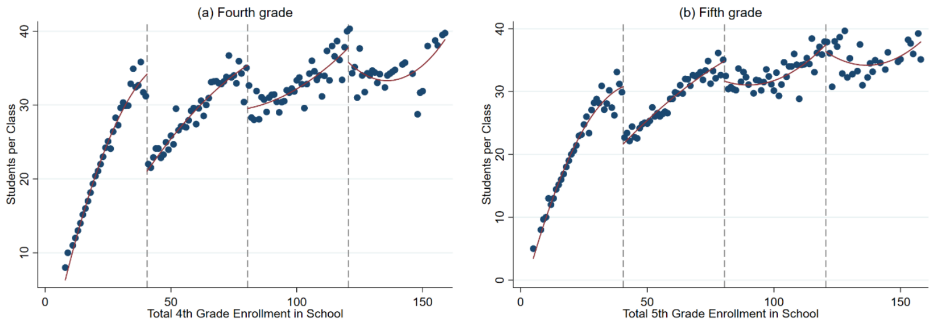
In this empirical project, you will use a regression discontinuity design to estimate the causal effect of class size on test scores. To answer some of the questions, you will need to refer to the following papers:

1. Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics* 126(4): 1593–1660.
2. Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics* 114(2): 533–575.

The Stata data file grade5.dta consists of test scores in fifth grade classes at public elementary schools in Israel. These data were originally used in Angrist and Lavy (1999). The graphs below were drawn using the same data.

Figure 1

Figure 1
Class Size as a Function of Total School Enrollment in Public Schools in Israel



Note: These figures plot class size as a function of total school enrollment for fourth grade and fifth grade classes in public schools in Israel in 1991.

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?

   A simple comparison of test scores in small versus large classes would not measure the causal effect of a class size since there are multiple other factors that contribute to test scores, including the quality of teachers, school and classroom resources, and the student's background as well.

   This simple comparison would be biased downwards, due to all these conditions and the partial observability of these individual circumstances.

2. How did Chetty et al. 2011 use data from the Tennessee STAR experiment to overcome this problem? What did they find?

   The Tennessee STAR experiment was conducted from 1985 to 1989 in Tennessee. It worked with around 12,000 K-3 grade level students from 79 mostly disadvantaged schools. The experiment randomly assigned students and teachers to different classrooms within these schools. Due to this arrangement, there were no systematic differences in background characteristics across the classrooms.

   The findings from the experiment confirmed that smaller class sizes were definitely favorable in regards to better test scores. Furthermore, evidence was found that kindergarten test scores were a strong predictor of later outcomes, including a student's yearly adult earnings.

3. Create a summary statistics table for the variables involved in the dataset grade5.dta (Similar to Table I from the Chetty et al. 2011 paper) Make sure to include the mean, standard deviation, minimum and maximum values of each variable.

   Sum scatter plots are good to explain different patterns and relationships between variables so outliers can be identified. To construct a sum scatter plot, you would have to use "scatter" and "twoway" in the command "scatter y x."

Variable	Mean	Std. Dev.	Min	Max
schlcode	39637.89	15266.09	11005	61365
school_enrollment	77.73551	38.82172	5	226
grade	5	0	5	5
classsize	29.92868	6.564	5	44
avgmath	67.29599	9. .596848	27.69	93.93
avgverb	74.37997	7.68254	34.8	93.89
disadvantaged	14.11639	13.50142	0	76
female	.4859675	.0982001	0	1
religious	.2461615	.4308804	0	1

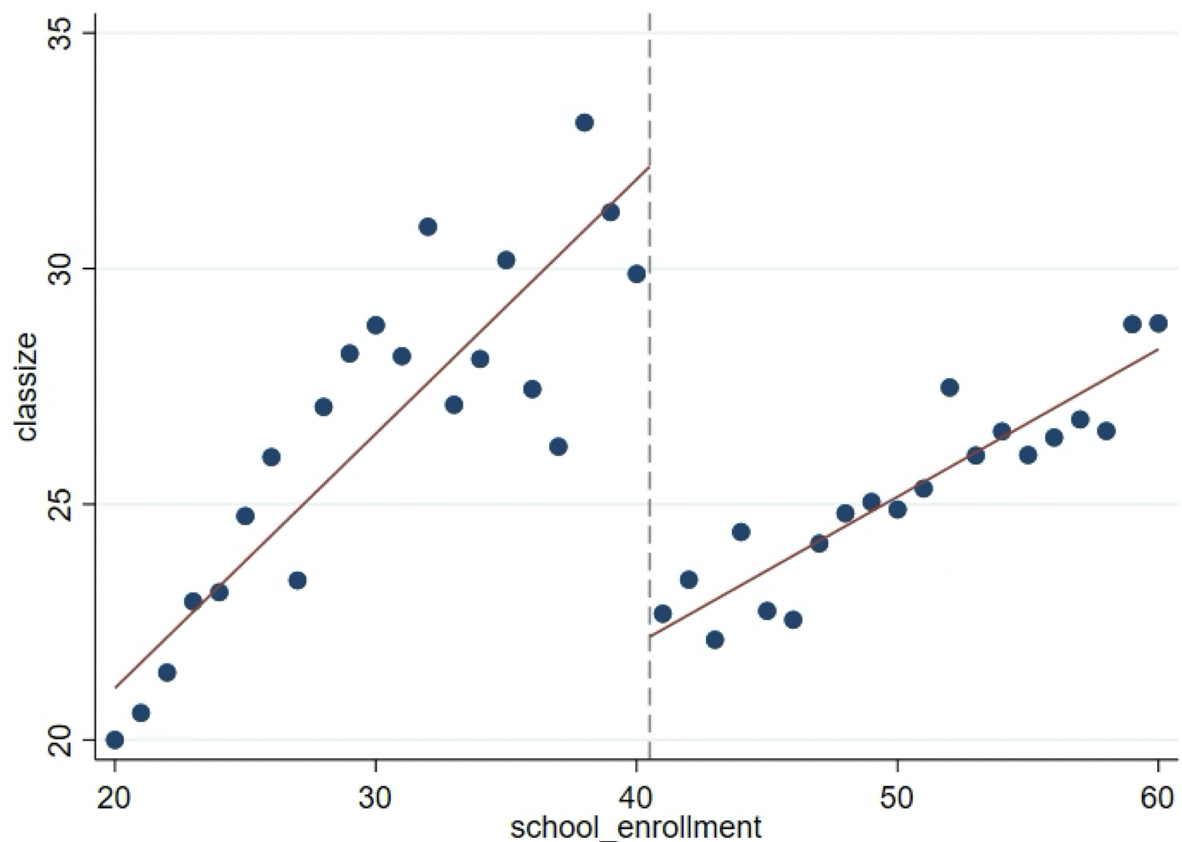
4. What is a binned scatter plot? Explain how it is constructed

A binned scatter plot is a good way of showing the relationship between two variables and is mainly used when dealing with very large data sets. To generate a binned scatterplot, the binscatter groups the x-axis variable into multiple same sized bins that generates the mean of the x and y-axis variables for each of the bins. A scatter plot is generated with these points.

STATA Command: binscatter avgmath school_enrollment [Example]

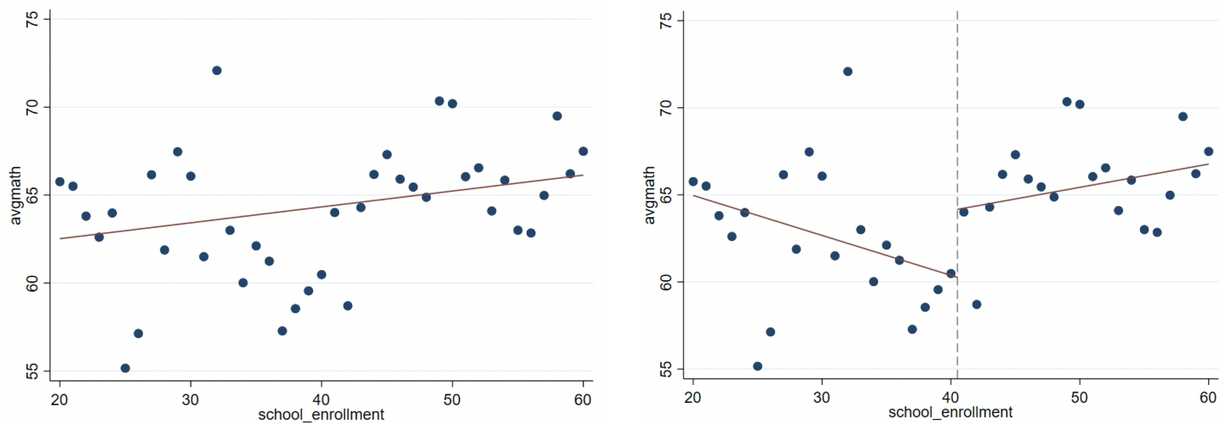
5. Graphical regression discontinuity analysis, focusing on the 40-student school enrollment threshold. Restrict the analysis to schools with student enrollment between 20 and 60 students. See Table 2 for more guidance.

a. Draw a binned scatter plot to visualize how class size changes at the 40-student school enrollment threshold. Display a linear regression line based on what you see in the data. What do you conclude from this visualization?



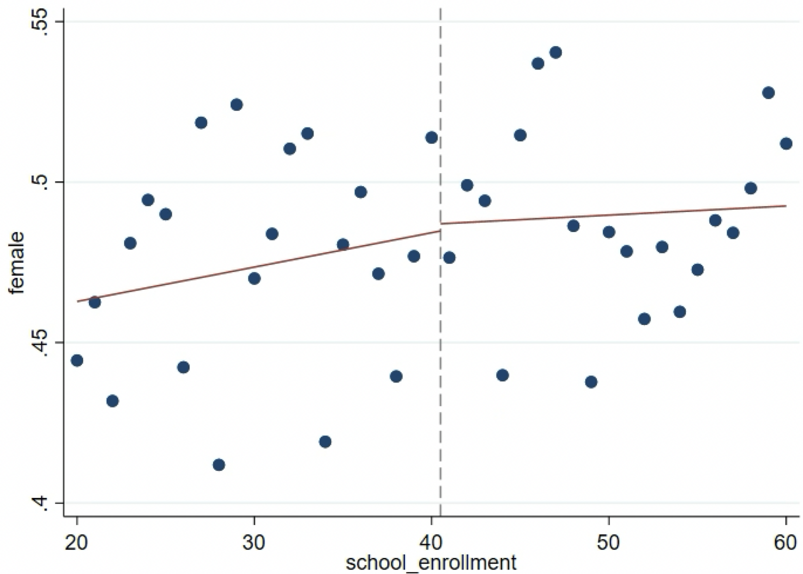
A line of regression is displayed in the graph above. There is a steeper line of regression for school enrollment of 20-40 students. The line becomes flatter once the enrollment increases from 40-60 students. The class size decreases as there is more school enrollment and the class size increases as there is less school enrollment.

b. Draw binned scatter plots to visualize how math scores change at the 40-student school enrollment threshold. Display a linear regression line based on what you see in the data. What do you conclude from this visualization?

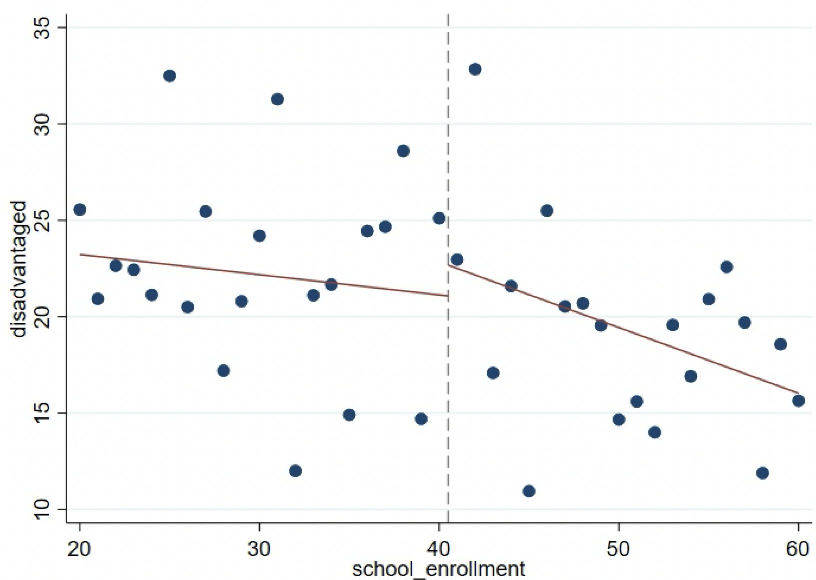


From this plot,without regression discontinuity, it is evident that the line has a positive slope which means that the average math score increases with school enrollment. However, with the second graph, when regression discontinuity is applied, it is obvious that the average math score only seems to go up alongside the school enrollment when the 40-student school enrollment is exceeded. This just shows that class size is not the main factor of class scores but perhaps there may be other underlying reasons.

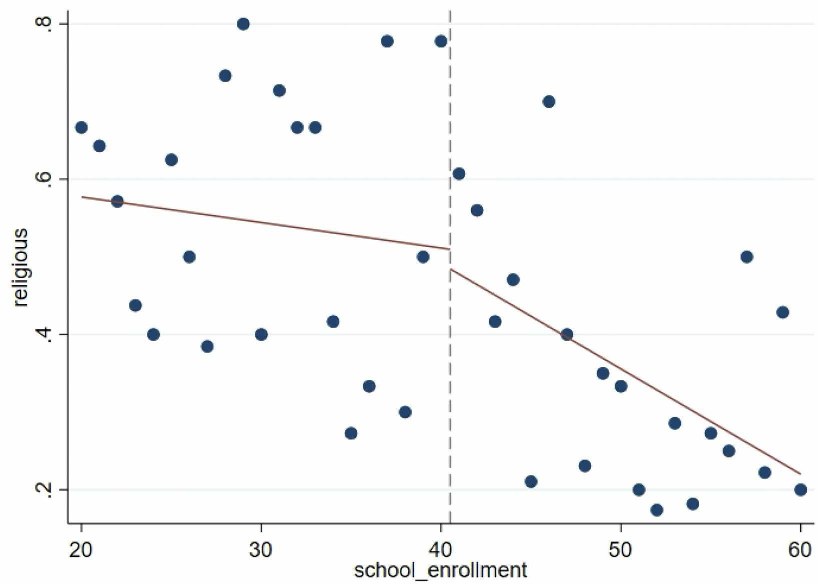
c. Draw binned scatter plots to visualize how the other variables (disadvantaged, female, religious) behave around the 40-student school enrollment threshold. What do you conclude from this visualization?



Function of Female: The line of regression slightly increases in terms of the percentage of females enrolled, however, the line of regression flattens (positive slope) in terms of the total school enrollment after the 40 threshold. The percentage of female students increases as the total school enrollment increases.



Function of Disadvantaged: The line of regression slightly increases in terms of the percentage of disadvantaged students at 40 students enrolled in the school, however, becomes steeper in a downward direction at the threshold.



Function of Religious: The line of regression is almost flat with school enrollment of 20-40 students, however, when there is a total enrollment of 40-60 students there is a steep line of regression that goes in a downward direction.

6. **Regression analysis.**

In estimating these regressions, use all the observations with school enrollment less than 80. You will need to first create a variable (call it "above40") as an indicator for the discontinuity. You also need to create a variable of normalized school enrollment (call it "x") that measures deviations of school enrollment from the 40-threshold. Look at Table 2 for more guidance.

a. Run a regression of classize on the above40 indicator. Are the estimated coefficients providing you with an estimate of the jump in class size at the 40- student threshold that you see in the binscatter plot? What's the problem?

```
. reg classize above40 if inrange(school_enrollment,0,80)

      Source |       SS           df       MS      Number of obs   =     1,180
-------------+----------------------------------   F(1, 1178)      =    144.63
       Model |  5598.95593         1  5598.95593   Prob > F        =    0.0000
    Residual |  45603.0305     1,178    38.71225   R-squared       =    0.1094
-------------+----------------------------------   Adj R-squared   =    0.1086
       Total |  51201.9864     1,179  43.4283176   Root MSE        =    6.2219

------------------------------------------------------------------------------
    classize |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     above40 |   5.030508   .4182949    12.03   0.000     4.209822    5.851195
       _cons |   23.63051    .362254    65.23   0.000     22.91977    24.34124
------------------------------------------------------------------------------
```

In this regression analysis table, we see a significant association of class size with enrollment above 40 (*above40*). The p-value is less than 0.01 and approx. 10% of the variation in class size is explained by enrollment above 40. We expect to see an increment of approx. 5 students as class size above 40 increases by 1. Moreover, we expect to see a class size of approx. 24 students with no enrollment above 40. However, as we cannot compare the difference in class size before and after the discontinuity threshold(40), we cannot make statistical inferences with above_40 and below_40 comparisons. The jump in class size is answered only after we add values of before_40 enrollment values and compare the values as multivariate regression.

b. Now run a regression of classize on the above40 indicator, x, and the above40 indicator interacted with x. What are the estimated coefficients now and are they consistent with the jump in class size at the 40-student threshold that you see in the binscatter plot? Explain.

```
reg classize above40 x x_above40 if inrange(school_enrollment,0,80)

      Source |       SS           df       MS      Number of obs   =     1,180
-------------+----------------------------------   F(3, 1176)      =    462.15
       Model |  27703.6782         3  9234.55939   Prob > F        =    0.0000
    Residual |  23498.3083     1,176  19.9815546   R-squared       =    0.5411
-------------+----------------------------------   Adj R-squared   =    0.5399
       Total |  51201.9864     1,179  43.4283176   Root MSE        =    4.4701

------------------------------------------------------------------------------
   classize |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     above40 |  -11.00075   .5950895   -18.49   0.000     -12.1683   -9.833193
           x |   .6725232   .0317162    21.20   0.000     .6102965    .7347498
   x_above40 |  -.3417826   .0342419    -9.98   0.000    -.4089647   -.2746005
       _cons |   33.04583   .5146785    64.21   0.000     32.03604    34.05562
------------------------------------------------------------------------------
```

In this regression analysis table, we see a significant association of class size with enrollment above 40 (*above40*). The p-value is less than 0.01 for all independent variables and approx. 54% of the variation in class size is explained by enrollment variables, greater than 10% with above40 as the only variable. We expect to see a decrease of approx. 11 students as class size above 40 enrollment increases by 1. Moreover, we expect to see a difference in cut-off of below40 and after40 enrollment discontinuity to be 0.67. This statistic shows that discontinuity in variables exists significantly to account for the change as also shown in the binscatter plot above. Likewise, we expect to see a decrease of approx. 0.34 students (*practically inapplicable*) as class size below 40 increases by 1. Finally, we expect to see a class size of approx. 33 students with no enrollment, with respect to the regression discontinuity.

c. Run a regression of avgmath on the above40 indicator. What do the estimated coefficients tell you?

```
reg avgmath above40 if inrange(school_enrollment,0,80)

      Source |       SS           df       MS      Number of obs   =     1,180
-------------+----------------------------------   F(1, 1178)      =     17.88
       Model |  1786.78784         1  1786.78784   Prob > F        =    0.0000
    Residual |  117709.284     1,178  99.9229916   R-squared       =    0.0150
-------------+----------------------------------   Adj R-squared   =    0.0141
       Total |  119496.072     1,179  101.353751   Root MSE        =    9.9961

------------------------------------------------------------------------------
     avgmath |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     above40 |   2.841809   .6720337     4.23   0.000     1.523292    4.160325
       _cons |   63.52526   .5819983   109.15   0.000     62.38339    64.66713
------------------------------------------------------------------------------
```

In this regression analysis table, we see a significant association of average math score with enrollment above 40 (*above40*). The p-value is less than 0.01 for the independent variables and approx. 1.5% of the variation in average math score is explained by enrollment above 40. We expect to see an increase of 2.84 in student's math score for a unit addition in enrollment above 40. We also expect to see an average math score of 65.5 with no enrollment above 40.

d. Given the above binscatter plot of the relationship between math scores and school enrollment, what are your concerns about the causal interpretation of your estimates of the effects of class size?
Hint: see Figure 2 and associated discussion in Angrist and Lavy (1999).

Given the outcome of the binscatter plot of the relationship between math scores and school enrollment, concerns regarding the causal interpretation of the estimates of the effects of class size include various factors such as location of the school and the resources it has, as well as the backgrounds of the students themselves.

e. Run a regression of avgmath on above40 indicator, x, and the above40 indicator interacted with x. What do the estimated coefficients tell you? How does this regression address your concerns about causality?

```
. reg avgmath above40 x x_above40 if inrange(school_enrollment,0,80)
```

Source	SS	df	MS
Model	3858.04708	3	1286.01569
Residual	115638.025	1,176	98.3316538
Total	119496.072	1,179	101.353751

Number of obs	=	1,180
F(3, 1176)	=	13.08
Prob > F	=	0.0000
R-squared	=	0.0323
Adj R-squared	=	0.0298
Root MSE	=	9.9162

avgmath	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
above40	3.428896	1.320123	2.60	0.010	.8388376	6.018955
x	-.192629	.070358	-2.74	0.006	-.3306702	-.0545878
x_above40	.298097	.0759609	3.92	0.000	.149063	.4471311
_cons	60.82845	1.141742	53.28	0.000	58.58837	63.06853

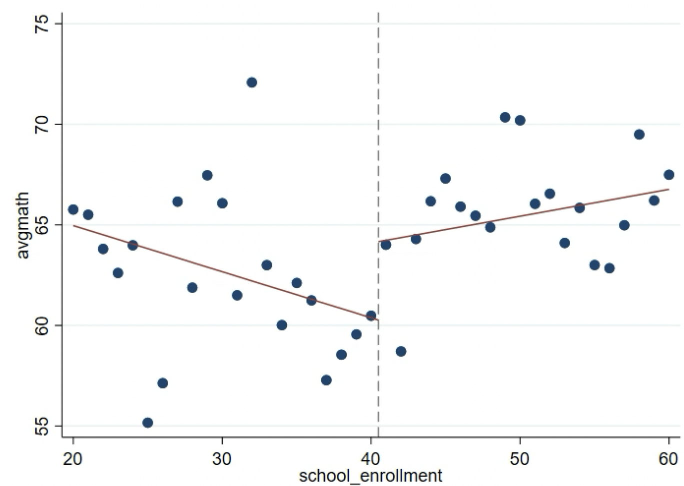
In this regression analysis table, we see a significant association of class size with enrollment above 40 (*above40*). The p-value is less than 0.05 for all independent variables and approx.

3.23% of the variation in average math test score is determined by enrollment variables (above and below 40 and discontinuity threshold). We expect to see an increase of 3.43 math score as class size above 40 increases by 1. Moreover, we expect to see a decrease in math test score by 0.19, explained by the cut-off of below40 and after40 enrollment discontinuity. This statistic shows that discontinuity in variables exists significantly to account for the change as also shown in the binscatter plot above. Likewise, we expect to see a decrease of 0.30 math test score as class size below 40 increases by 1. Finally, we expect to see a math score of 60.8 with no enrollment, with respect to the regression discontinuity.

7. 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 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.

5 * (-0.19)

0.96 point increase in math test score when the class size decreases by 5.



8. Now 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?

We can make reasonable predictions from 40 to 35 with 5 students decreasing but we don't know if we can stretch the prediction to half the numbers, and suddenly fall from 40 to 20. Regression discontinuity measures the marginal difference between the threshold value but extrapolating too far will not be reliable. Also, we cannot be confident about our findings as the R-squared value is low(0.03).

9. Compare your estimates in 7 with the estimates from the Tennessee STAR experiment (Chetty et al. 2011) discussed in lecture. Give two reasons that your estimates might differ from those of these other studies.

From the Tennessee Star Experiment, the estimated increase in test scores after a reduction in class size was 4.81%. Whereas our estimate was 0.96%. The reason for varied values may be because of the different sample sizes used in different experiments and also because unlike the STAR experiment, our values were not randomly sampled.

10. 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.

    The class size does affect factors such as income. Data regarding the student's income was determined through analyzing the W-2 forms that were filed. The earnings were capped to $100,000 to avoid any outliers; only less than 1% of the STAR samples made an income of < $100,000. The mean income of the STAR sample between 2005-2–7 was $15,912, which is lower than the same cohort in the general U.S population (Chetty et al. 2011). The college attendance rates could not be properly determined for several reasons. Title IV states that higher education institutions must be eligible for financial aid and the 1098-T must be filed in order to report tuition payments. Data was collected based on the 1098-T forms for STAR students; 26.4% students were enrolled in 2000 and overall 45.5% students enrolled between 1999 to 2007 (Chetty et al. 2011). The data was only based on the tuition collected per student and could not properly assess college completion. Chetty has used multiple samples and data points, which we did not have access to. Our data points were not sufficient enough to be compared to Chetty's experiments and they were also randomized. Increasing or decreasing a class size may not have the same impact as improving the quality of education.

Thank you!