Small Classroom Size as an Indicator of Better Academic Performance in Kindergarten Schools

Aditya Raman

20 November 2022

1 Introduction

This project investigates the following research question: What is the relationship between class size and kindergarten academic performance. In specific, we look to investigate whether smaller classroom sizes are correlated with better kindergarten math SAT scores, and understand why such a correlation may or may not exist. There have been studies which investigate whether small classroom sizes are indicative of better academic performance. Some studies argue that smaller classroom sizes reduce teachers' workload, allow for less chaotic and more conducive environments, as well as enable better individual engagement between teachers and students (Kamuche 2011). Such arguments seem very plausible, and therefore this projects aims to empirically validate whether such a correlation is present in historical school data as well.

2 Methodology

The *star.dta* file contains 5710 observations and 15 columns (one per student), which include variables such as the student's ID, kindergarten SAT scores (for math, reading, listening, and word skill), as well as information on the class size. The two variables which pertain to class size are *class_size* which represents the numerical size of the student's kindergarten classroom, as well as a binary variable *small*, which indicates whether the classroom size is considered as small or not. In order to determine the relationship between class size and academic performance, we will use two methods. Firstly, the metric we use for academic performance is the student's kindergarten Math SAT score. The first method involves comparing the mean Math SAT score for students in large classrooms ($small = 0$), and the mean Math SAT score for students in small classrooms ($small = 1$). This descriptive statistic enables us to gain a broad picture of the difference in average Math SAT scores depending on whether a child is in a small classroom or not.

The second method involves using an OLS (Ordinary Least Squares) regression of Math SAT score against class size. Here, the independent variable is

the class size and the dependent variable is the Math SAT score. Running the regression enables us to calculate the p-value, R-squared coefficient, as well as the coefficient on the classroom size, which have different interpretations. The coefficient on the classroom size tells us, given an increase in classroom size by 1 student, how much does the Math SAT score typically change. On the other hand, the R-squared coefficient tells us how much of the variability in the Math SAT score is explained by classroom size, which indicates how well our linear model fits the data.

3 Results

For the first method, we calculate the mean Math Kindergarten score for students in large classrooms to be 483.4 and the mean Math Kindergarten score for students in small classrooms to be 491.4. This descriptive statistic shows us that, on average, children enrolled in smaller classroom sizes tend to have average Math SAT scores. However, this descriptive statistic is limited in the sense that we cannot use it to make predictions and specific inferences, only to gain a general understand of the distribution of Math SAT scores depending on the classroom size.

In order to better understand the relationship between classroom size and Math SAT scores, we regress Math SAT score on classroom size. The results are shown below:

Source	SS	df	MS	Number of obs	=	5,710
				F(1, 5708)	=	31.13
Model	70596.4761	1	70596.4761	Prob > F	=	0.0000
Residual	12943926.8	5,708	2267.68164	R-squared	=	0.0054
				Adj R-squared	=	0.0053
Total	13014523.3	5,709	2279.65025	Root MSE	=	47.62

math	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-.8867015	.1589194	-5.58	0.000	-1.198244	-.5751591
_cons	503.7592	3.284227	153.39	0.000	497.3208	510.1975

Figure 1: Regression of Kindergarten Math SAT Score on Classroom Size

As seen above, the results are significant since $p < 0.001$. In this example, the constant term is not meaningful, since it represents the predicted Math SAT score when the classroom size is zero, but we cannot extrapolate our data since it is not possible to have a student in a class of size zero. The coefficient on the class size is -0.887. The negative sign on the coefficient empirically validates that classroom size and kindergarten Math SAT score are negatively correlated. Moreover, the magnitude suggests that, an increase in class size by 1 typically leads to a decrease in kindergarten Math SAT score by -0.887 points. Furthermore, the R-squared coefficient is 0.0054 which implies that class size can only explain 0.54% of the variability in kindergarten Math SAT score. This

value is extremely low but may be expected, since there are several confounding variables which affect a student's performance in class. For example, the school teacher's experience is another significant variable that influences students' academic performances and math SAT scores in specific. For example, when we regress kindergarten Math SAT score on classroom size, conditional on the teacher's years of experience being 8 years, we get the following results:

Source	SS	df	MS	Number of obs	=	485
				F(1, 483)	=	15.82
Model	31473.02	1	31473.02	Prob > F	=	0.0001
Residual	960763.029	483	1989.15741	R-squared	=	0.0317
				Adj R-squared	=	0.0297
Total	992236.049	484	2050.07448	Root MSE	=	44.6

math	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-1.991863	.5007546	-3.98	0.000	-2.97579	-1.007937
_cons	519.1707	10.12775	51.26	0.000	499.2708	539.0706

Figure 2: Regression of Kindergarten Math SAT Score on Classroom Size, conditional on teachers' years of experience = 8

Here, we can see that controlling for the teacher's level of experience enables for much more significant results, since $p < 0.001$, the coefficient on the classroom size is larger, i.e. equal to -1.992, and the R-squared coefficient is larger as well at 0.0317, meaning that this 3.17% of the variability in Math SAT score is explained by classroom size, when comparing across classrooms which have teachers of the same level of experience.

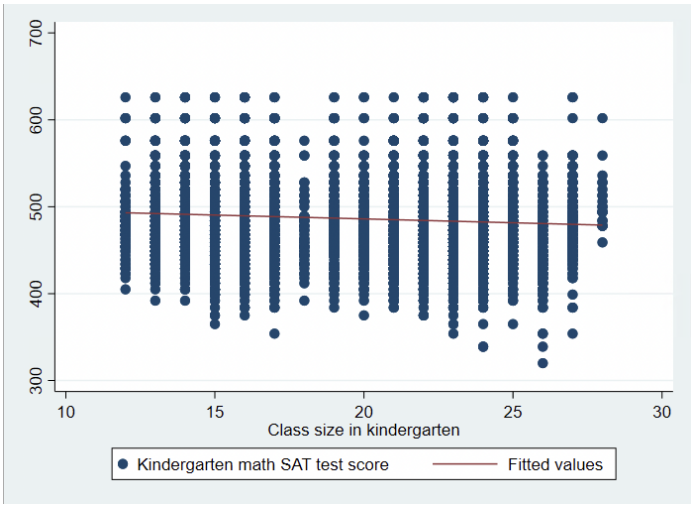

Figure 3: Scatter plot of Kindergarten Math SAT Score against Classroom Size

4 Conclusion

In conclusion, the data shows us that classroom size is negatively correlated with academic performance which agrees with the literature cited in the introduction. However, in this data set, classroom size is not a strong indicator of kindergarten math SAT scores, possibly due to various confounding variables, such as the teacher's level of experience. The model also has limitations, since we implicitly assume linearity between classroom size and math SAT score, which may not be the case, e.g. it may be the case that classroom size influences math SAT scores up to a certain threshold, past which linearity fails to hold. Moreover, the regression analysis is not resistant to outliers in data, which may create a less confident model. In our case, we see that, ceteris paribus, small classrooms tend to be correlated with higher academic scores, and therefore a policy implementation may be to divide up public schooling class sizes into smaller divisions, as long as each teacher in the smaller division has roughly the same level of experience. Such a policy may be effective in increasing kindergarten academic performances. For a future research possibility, it would be interesting to consider whether there are some other factors which are stronger indicators of academic performance in kindergarten schools. For example, we could consider other independent variables such as a teacher's level of experience, as well as different metrics / combinations of academic performance, e.g. cumulative SAT scores.

Through this project, I have had the opportunity to effectively devise, investigate, and analyse a real life research question with meaningful implications, with Stata.

5 References

Kamuche, Felix. (2011). Does Smaller Class Size Affect Student's Performance In Basic Statistics Class? An Empirical Study. International Business Economics Research Journal (IBER). 5. 10.19030/iber.v5i8.3496.