Modeling the association between low-income and minority students on graduation rate in California high school districts

Stat 230 | November 19th, 2022
Helen Moses, Ben Griesel, Elena Ea

Introduction

In this study, we will develop a multiple linear regression model to determine how race and different socioeconomic factors influence graduation rate. The New York Times Article "For the Poor the Graduation Gap is Even Wider Than the Enrollment Gap" by Susan Dynarski discusses the 2002 Educational Longitudinal Study conducted by the National Center for Education Statistics (NCES)[1]. The study showed that students whose parents made the lowest incomes and had the lowest level of education were more likely to work in unskilled jobs, while those whose parents were highly educated and had a high income were more likely to work as professionals or managers. Furthermore, the study reported that 58% of students from lower-income families aspired to earn at least a bachelor's degree. However, among those participants, only 14% actually earned a bachelor's degree after thirteen years. Conversely, 87% of students from high-income families aspired to earn at least a bachelor's degree and 60% of those students actually earned a bachelor's degree after thirteen years. Following the results of this study, we are interested in exploring relationships between socioeconomic status and graduation for students at a high school level. More specifically, we are curious if the graduation rate for high school districts in California is associated with the percentage of minority students and the percentage of low-income students in the districts.

Data

Each case in the dataset is a public school district enrolling more than 15,000 students in California in 2018[2]. Selected statistics on enrollment, teachers, dropouts, and graduates were compiled for each case by the NCES. Each case was collected independently of each other as each school district is a separate case. For our study, we fit a multiple linear regression model with graduation rate (Graduation) as our response variable and percent Black students (Black) and percent of students on free and reduced lunch (Free) as our primary explanatory variables of interest. However, because this is an observational study, we wanted to control for potential confounding variables. The variables we elected to control for were the size of the school (small or large) represented by the variable Size, and the ratio of students to teacher (Ratio). We included Black as one of our primary explanatory variables to investigate an association between graduation rate and the percentage of minority students because Black students are the largest percentage minority race in our data set. We elected to use the Free variable as our other primary variable of interest because analyzing the percentage of students on free and reduced lunch helps identify the economic diversity of that school district[3]. For example, a district with a larger percentage of students on free and reduced lunch suggests there are more low-income families in that district than a district with a smaller percentage of students on free and reduced lunch. It is important to note that while we did control for the effects of potential confounding variables, this is still just an observational study, so we will not be able to assert causation in our findings.

Results

A multiple linear regression model was fitted to the data. The fitted model equation is:

$$\hat{y} = 89.84171 - 0.34250Black - 0.07992Free - 0.18834Size + 0.25794Ratio$$

*$\hat{y}$ *denotes the predicted graduation rate*

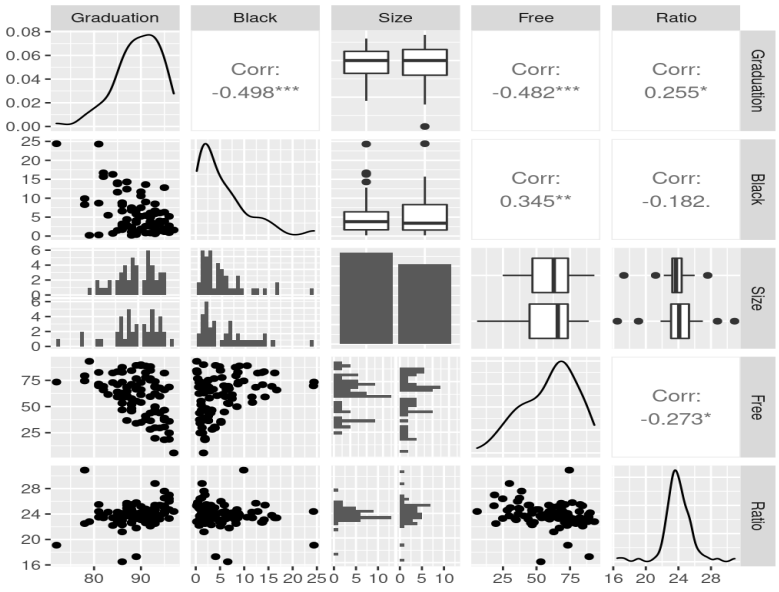


Figure 1: Scatterplot Matrix

Figure 1 shows a scatterplot matrix comparing each of the explanatory variables to Graduation rate. We see from the scatterplots in the left-most row that the quantitative variables, percentage of black students and percentage of students on free/reduced lunch, have a negative association with graduation rate. Ratio of students to teachers does not appear to have a clear relationship with graduation rate. Rather, the majority of such ratios are between 0.22 and 0.26 and there are varying graduation rates throughout districts with those ratios. Out of the 4 explanatory variables, percentage of Black students and percentage of students on free/reduced lunch had significant slope coefficients and therefore were associated with graduation rate as shown in

Table 1. Conversely, the size of the school district and ratio of students to teachers had insignificant slopes. We found the percentage of Black students in a district had a negative association with graduation rate, holding all other variables constant. The graduation rate decreases at a rate of approximately 0.34 per percentage increase in Black students for a school district (Table 1). In addition, we are 95% confident that a one percent increase in percentage of Black students is associated with the graduation rate decreasing between 0.515 to 0.169. Likewise, the percentage of students on free/reduced lunch in a district had a negative association with graduation rate. As the percentage of students on free/reduced lunch increases, the graduation rate decreases at a rate of approximately 0.08. We are 95% confident that a one percent increase in percentage of students eligible for free/reduced lunch in a district is associated with the graduation rate decreasing between 0.125 to 0.034.

Variable	Slope Coefficient	Standard Error	p-value
Black - Percentage of Black students	-0.34250	0.08850	0.000216
Free - Percentage of students on free/reduced lunch	-0.07992	0.02321	0.000904
Size - Large or small school	-0.18834	0.85874	0.826938
Ratio - Student:teacher ratio	0.25794	0.23551	0.276591

Table 1: Table of summary statistics for the multiple regression model

Discussion

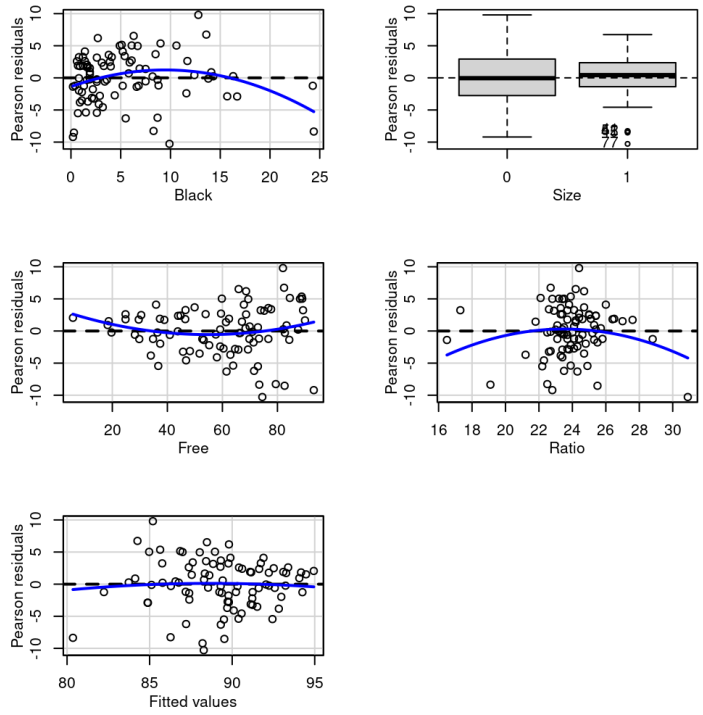
Our surrounding research on the topic of socioeconomic circumstance and high school graduation rate, such as Dynarski's article and the 2002 study by the NCES, conclude a negative association between family income and a student's future success in terms of college graduation. In addition to variables pertaining to socioeconomic status, we also were interested in investigating the relationship between the proportion of minority students and school success in terms of high school graduation. In our study we hypothesized that a greater percentage of students on free/reduced lunch is associated with a lower graduation rate and our findings confirmed this hypothesis. As the percentage of students on free or reduced lunch increased, the graduation rate decreased. However, while checking model assumptions we found that there were signs of slight heteroscedasticity with the free and reduced lunch variable that did not improve as we tried fitting a variety of different models to the data (See Appendix 1). This means that the residual variance is not constant among observations for the variable percentage of students on free/reduced lunch. If we had more time to complete this study, we would have investigated a wider variety of remedies to see if any new models would be a better fit for our data. Unfortunately, with our time constraints this was the best model we could find to fit onto the data. Moreover, there were no significant outliers that would contribute to the observed heteroskedasticity. Because there were only concerns with the free and reduced lunch variable, We can be confident in our findings for the association between an increase in minority students and a decrease in the graduation rate. The data we collected was from 2018, and since then a new law has passed in California that all students are eligible to receive free and reduced lunch, regardless of their family income[4]. Because of this, in future research it would be interesting to

revisit the association between percentage of students using free and reduced lunch and high school graduation rates. Furthermore, variables such as how a student gets to school (mode of transportation) and percentage of students who hold a job (could indicate less time for school work) could also have an association with graduation rate and it would be interesting to do analysis on such relationships in future research.

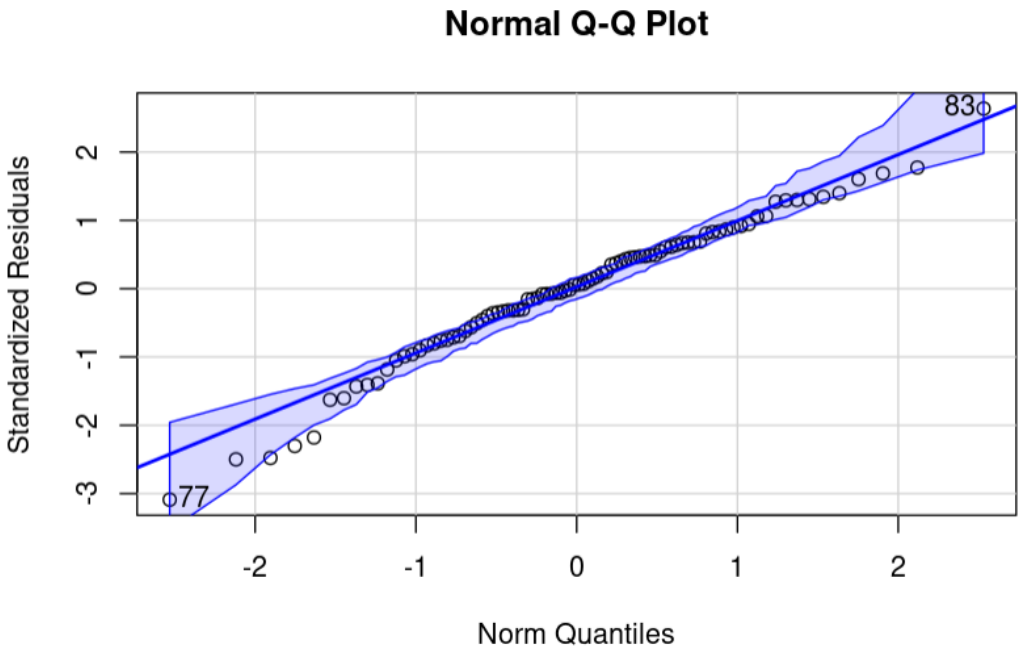
References

1. Dynarski, Susan. “For the Poor, the Graduation Gap Is Even Wider than the Enrollment Gap.” *The New York Times*, The New York Times, 2 June 2015, https://www.nytimes.com/2015/06/02/upshot/for-the-poor-the-graduation-gap-is-even-wider-than-the-enrollment-gap.html?smid=nytnow-share&smprod=nytnow&_r=1&abt=0002&abg=1.
2. “Digest of Education Statistics, 2020.” *National Center for Education Statistics (NCES) Home Page, a Part of the U.S. Department of Education*, 2020, https://nces.ed.gov/programs/digest/d20/tables/dt20_215.10.asp.
3. “Income Eligibility Scales for School Year 2018–19.” *Income Eligibility Scales for School Year 2018–19 - Rates, Eligibility Scales, & Funding (CA Dept of Education)*, https://www.cde.ca.gov/ls/nu/rs/scales1819.asp.
4. “California Universal Meals.” *California Universal Meals - School Nutrition (CA Dept of Education)*, https://www.cde.ca.gov/ls/nu/sn/cauniversalmeals.asp.

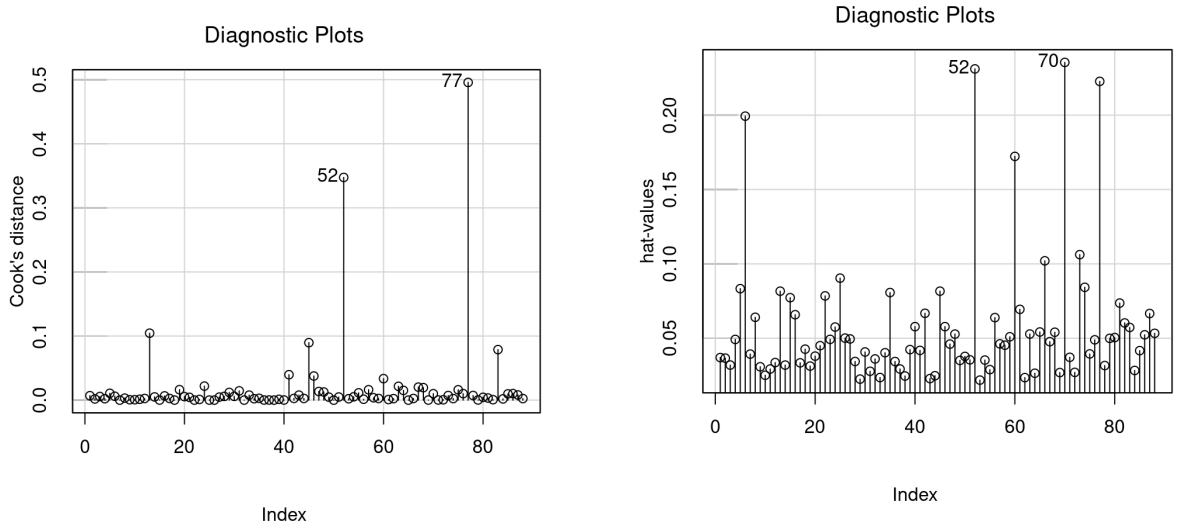
Appendix



Appendix 1: Residual Plots



Appendix 2: Normal Q-Q Plot



Appendix 3: Diagnostic plots to identify potential outliers

Variable	VIF Test Value
Black	1.161305
Free	1.206477
Size	1.027984
Ratio	1.101654

Appendix 4: VIF Test results