An Examination of Secondary School Success Indicators

Luke Denton*

April 27, 2021

Abstract

In this project, I seek to answer a difficult question many experts and organizations have been asking for years: What drives academic performance for students in secondary school? To answer this question, I have computed linear regressions to better understand the impact of different variables on student grade performance. With currently washy results, the implications of this analysis are still in need of additional research to find causal relationships between student demographics and performance.

*Department of Economics, University of Oklahoma. email address: lukedenton@ou.edu

1 Introduction

Education is one of the biggest components of adolescent life. School is where many children spend thousands of hours learning and interacting with teachers, administrators, and their peers. Heralded as a wonderful opportunity to surpass generational poverty, public education provides schooling for millions of students around the world. Despite the universal existence of public education, a single strategy or set of strategies that help students succeed better in school have largely gone undefined. This drains administrative resources and forces school districts as well as parents to continue to seek the best strategy to foster success for school age children. Without a clear strategy for how parents or teachers can best support children in school, children in schools that are poorly funded are susceptible to a lower quality education.

2 Literature Review

Much work has been done to better understand the factors that go into student achievement at school. Scholars have studied numerous inputs and scenarios to understand what makes a student succeed in their grades. K. Vijayalakshmi Suggests that emotional intelligence, a fairly difficult trait to quantify, serves as a useful indicator of how well students do in secondary school. Parcel and Dufur (2001) Suggests that higher amounts of capital in both schools and at home produce positive effects for student academic achievement, suggesting that income may not be as large of a definitive indicator of kids' grades. Boyd et al. (2008) Takes an alternative approach and attempts to assess student performance from the teacher side of the classroom. Teacher preparation is something that varies from school to school and may be an overlooked indicator for how well students do in school. Similarly focusing on teachers, Dee (2004) focuses on the effects of racial pairings of students and teachers on school performance.

3 Data

The primary data source for this research is the student performance database from UCI Machine Learning Data Sets Cortez and Silva (2008). The data tracks secondary school students at two different schools in Portugal. The data measure the first, second, and final period grades of students in mathematics and Portuguese classes, as well as many other demographics concerning home life and school participation. There are 649 total observations in this data set. For an explained list of all the variable codes, see 2.

4 Empirical Methods

While I humbly await my office hours with you to help guide me through getting some of the ML stuff sorted out, the main tactic I have engaged to better understand these indicators is a multiple linear regression model. The basic equation for the multiple lm is shown as follows:

$$Y = \beta X + u \tag{1}$$

In this analysis, the dependent variable (Y) that I am choosing to focus on (and hopefully be able to predict) is G3, a student's final grade recorded in their class. The independent variables (X's) vary depending on the model, but are included to help provide an explanation for the variation in final grades (Y) among students.

- Model 1: Multiple Linear Regression Model

  This initial model attempts to explain variation in final grades based off of numerical variables that are typically seen as indicative of a student's performance.

$$final = \beta_0 + \beta_1 Medu + \beta_2 Fedu + \beta_3 studytime + u \tag{2}$$

- Model 2: Expanded Multiple Linear Regression Model

  This model contains the previous independent variables but also includes new ones in hopes of increasing the amount of variation explained.

$$final = \beta_0 + \beta_1 Medu + \beta_2 Fedu + \beta_3 studytime + \beta_4 address \beta_5 famsize + \beta_6 Pstatus + \beta_7 activities + u \tag{3}$$

- Model 3:

  This is where I want to dive into ML and attempt to craft a model that penalizes for complexity but does well at predicting out of sample. Need some help from you to make this happen though.

5 Research Findings

The results for models 1 and 2 can be reviewed in 1. From the first model, we learn the substantial importance of mother education level. It's important to keep in mind that the final grades in this data set are on a scale of 20 points; so, an increase of 0.794 per increase in education level of a student's mother is a very significant level of change. Father's education level and also amount of time spent studying each week also have positive effects on students' final grades. These are all rather intuitive relationships that help portray some aspects of students' educational ability that are difficult to quantify (ex: well educated parents are likely to be well-abled and have well-abled children).

From the second model, there are a few interesting relationships within the data. First, there is a significant positive effect in families with size less than or equal to 3. This indicates that children with no siblings perform better in school. Additionally, the relationship between having parents living together and school performance is negative. There is also a negative relationship between being involved in extracurricular activities and final grade performance.

These results are not very strong. The R squared of model 1 was a mere 0.055, and model 2 was only 0.070. There is a significant portion of the variation in final grades that remains unexplained from these models. The variance within each variable's estimated effect is also very high.

6 Conclusion

The arduous and seemingly mystical task of understanding what precisely causes a student's performance in school is a subject of careful study and mixed reviews. In an environment where factors from home, school, teachers, and peers all interact, predicting education performance as well as finding causal relationships between indicators is difficult. Beyond intuitive indicators that weren't even statistically significant, there is much more that needs to be done in the scope of this project.

References

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Teacher Preparation and Student Achievement." .

Cortez, Paulo and Alice Silva. 2008. "USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE." *Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008)* :8–14.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86 (1):195–210.

K. Vijayalakshmi, M. Selvarani. ???? "WHAT MAKES A STUDENT BRIGHT? HOW EMOTIONAL INTELLIGENCE MEDIATES THE RELATIONSHIP BETWEEN STUDENT'S SELF-EFFICACY AND ACADEMIC ACHIEVEMENT." *Journal of critical reviews* 7 (12).

Parcel, Toby L. and Mikaela J. Dufur. 2001. "Capital at Home and at School: Effects on Student Achievement." *Social Forces* 79 (3):881–911.

7 Figures and Tables

	Model 1	Model 2
(Intercept)	6.907	6.249
	(0.833)	(1.205)
Medu	0.794	0.742
	(0.264)	(0.267)
Fedu	0.147	0.180
	(0.265)	(0.265)
studytime	0.468	0.526
	(0.269)	(0.271)
addressU		0.794
		(0.547)
famsizeLE3		0.918
		(0.505)
PstatusT		-0.253
		(0.756)
activitiesyes		-0.103
		(0.458)
Walc		-0.082
		(0.183)
Num.Obs.	395	395
R2	0.055	0.070
R2 Adj.	0.048	0.053
AIC	2310.0	2311.6
BIC	2329.9	2347.4
Log.Lik.	-1150.017	-1146.787
F	7.576	4.178

Table 1: Model Summary Output

Table 2 is an explanation of the variables used in this analysis.

variable	description
school	student's school
sex	student's sex
age	student's age
address	student's home address type
famsize	family size
Pstatus	parent's cohabitation status
Medu	mother's education
Fedu	father's education
Mjob	mother's job
Fjob	father's job
reason	reason to choose this school
guardian	student's guardian
traveltime	home to school travel time
studytime	weekly study time
failures	number of past class failures
schoolsup	extra educational support
famsup	family educational support
paid	extra paid classes within the course subject
activities	extra-curricular activities
nursery	attended nursery school
higher	wants to take higher education
internet	Internet access at home
romantic	with a romantic relationship
famrel	quality of family relationships
freetime	free time after school
goout	going out with friends
Dalc	workday alcohol consumption
Walc	weekend alcohol consumption
health	current health status
absences	number of school absences
final	final grade

Table 2: variable codes and explanations