An Examination of Secondary School Success Indicators

Luke Denton

5/4/2021

Introduction

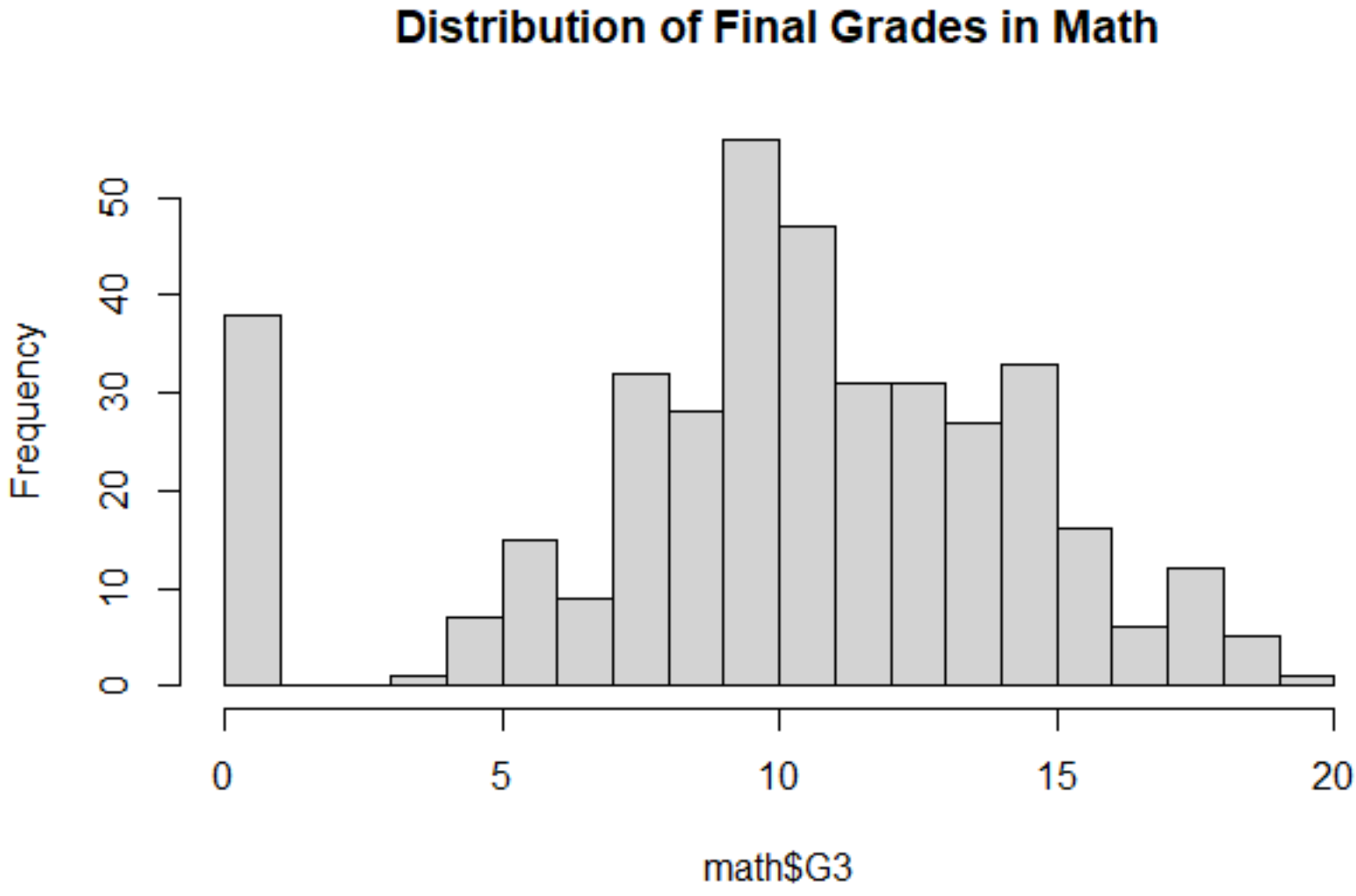
- Public education is heralded as an opportunity to overcome generational poverty
- Difficult to quantify causes of student success
- Insufficient resources
- Best method for student success is unclear
- Inferior educational outcomes for students

Data

UCI Machine Learning - Student Performance Database

- 649 observations, students ages 15-22
- 1,2,3 period grades in math & portuguese
- student demographics & school activities
- parent education, family size, # failed classes, # absences, urban or rural housing, study time, extracurricular activities, internet access, romantic relationships, pursuing higher ed, alcohol consumption

Data


Distribution of Final Grades in Math
Frequency
0
10
20
30
40
50
0
5
10
15
20
math$G3

Methods

Multiple Linear Regression to explain causality, in progress using ML to predict outcomes

$$\text{final} = \beta_0 + \beta_1 G1 + \beta_2 G2 + u$$

$$\text{final} = \beta_0 + \beta_1 G1 + \beta_2 G2 + \beta_3 Medu + \beta_4 Fedu + \beta_5 studytime + u$$

$$\text{final} = \beta_0 + \beta_1 G1 + \beta_2 G2 + \beta_3 Medu + \beta_4 Fedu + \beta_5 studytime + \beta_6 address + \beta_7 famsize + \beta_8 Pstatus + \beta_9 activities + u$$

Findings

	Model 1
(Intercept)	-1.830
	(0.335)
G1	0.153
	(0.056)
G2	0.987
	(0.050)
Num.Obs.	395
R2	0.822
R2 Adj.	0.821
AIC	1648.2
BIC	1664.1
Log.Lik.	-820.115
F	906.134

Findings

- 1st and 2nd period grades have a significant causal effect on final grades (Rsq $= 0.822$)
  - 2nd period grade had a much stronger effect (beta $= 0.987$)
- No other variables in other models were statistically significant

Conclusion

- Many of the factors attributed to student success did not have a causal effect in this model
- ML model would prove useful for prediction, since statistically significant causality is difficult to establish
- Continued study: using US Dept of Ed database on student performance