

My title*

Yihang Cai

March 18, 2024

1 Introduction

Understanding the factors that influence academic performance has been a critical area of educational research, as it can inform strategies to enhance learning outcomes and identify areas where support is needed. The Section 2 section introduces the dataset obtained from kaggle. The Section 3 section justifies the choice of the model and then summaries the outcome of the model.

2 Data

The dataset hosted on Kaggle, titled “Student Study Performance,” provides a rich source of information to analyze various aspects that may affect a student’s academic scores. This dataset encompasses a range of variables including demographic details, parental level of education, test preparation courses, and students’ math scores.

Table 1: top 10 rows of the dataset

Gender	Level of Parent Education	Test Preparation	Math Score
female	bachelor’s degree	none	72
female	some college	completed	69
female	master’s degree	none	90
male	associate’s degree	none	47
male	some college	none	76
female	associate’s degree	none	71
female	some college	completed	88
male	some college	none	40

*Code and data are available at: <https://github.com/peachvegetable/STA302-mini-essay-10>

Gender	Level of Parent Education	Test Preparation	Math Score
male	high school	completed	64
female	high school	none	38

Table 1 shows the top 10 rows of the selected data, displaying students' math score, gender, parents' level of education and test preparation status.

3 Model

3.1 Model set-up

Define y_i as the math score. Then β_0 is the intercept and β_1 to β_3 is coefficient to the variables.

$$\begin{aligned}
y_i | \lambda_i, \theta &\sim \text{NegativeBinomial}(\mu_i, \theta) \\
\log(\mu_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{parental_level_of_education}_i + \beta_3 \times \text{test_preparation_course}_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5)
\end{aligned}$$

3.2 Model summary

Table 2 displays the summary of the model. Intercept: The model's intercept, 4.229, is the log count of the expected math_score when all other predictors are held at their reference levels (which typically means 'female' for gender, 'none' or a baseline category for parental level of education, and 'completed' for test preparation course).

gendermale: The coefficient for 'gendermale' is 0.082. This suggests that, holding all else constant, being male is associated with a log count increase in the math_score of about 0.082 as compared to being female.

parental_level_of_education: This set of coefficients compares different levels of parental education to the reference category. For example:

bachelor's degree: The coefficient is 0.021, meaning that having a parent with a bachelor's degree is associated with a log count increase in the math_score of about 0.021 as compared to the reference category. high school: The coefficient is -0.085, meaning that having a parent

Table 2: Negative binomial model summary

	Negative binomial
(Intercept)	4.229 (0.020)
gendermale	0.082 (0.015)
parental_level_of_educationbachelor's degree	0.021 (0.026)
parental_level_of_educationhigh school	−0.085 (0.023)
parental_level_of_educationmaster's degree	0.037 (0.035)
parental_level_of_educationsome college	−0.010 (0.023)
parental_level_of_educationsome high school	−0.074 (0.024)
test_preparation_coursenone	−0.082 (0.016)
Num.Obs.	1000
Log.Lik.	−4148.533
ELPD	−4156.0
ELPD s.e.	35.6
LOOIC	8312.0
LOOIC s.e.	71.2
WAIC	8312.0
RMSE	14.44

Table 3: Mean and variance of students' math score

Mean	Variance
66.1	229.9

with only a high school education is associated with a log count decrease in the math_score of about 0.085 as compared to the reference category. The other education levels can be interpreted similarly. A positive coefficient indicates an increase in log count score compared to the reference, and a negative coefficient indicates a decrease. test_preparation_course: The coefficient for 'test_preparation_coursenone' is -0.082. This indicates that not taking a test preparation course is associated with a log count decrease in the math_score of about 0.082 as compared to those who did complete a test preparation course.

3.3 Model justification

4 References