# Bayesian Analysis of Student Test Scores


## Abstract

Bayesian modelling was performed to test for significant differences in Math, Reading, and Writing test scores across two primary variables: lunch status and gender. Further modelling was done to check whether test preparation has a significant interaction on student groups in these two variables. Math scores for students receiving standard lunches performed significantly better than their peers receiving free/reduced lunches. However, further modelling suggested that test preparation (in any subject) does not have a more significant impact on students receiving free/reduced lunches. Writing scores for female students at the school performed significantly better than their male peers. However, further modelling again suggested that test preparation does not have a more significant impact on male students in any subject.


## Introduction

This analysis set out to focus on the interaction of test preparation compared to other variables in the dataset. Being the sole variable that the school has some control over, it is of the most interest. Knowing which groups of students benefit the most from test preparation could help the school drive the most impact for students that need it the most. First, it is tested whether there is a statistically significant difference in score results among students receiving free/reduced lunches compared to their standard lunch peers. After confirming, this a hierarchical model is used to check if free/reduced lunch students receive statistically significant more benefits from test preparation than their standard lunch peers. The same process was repeated for gender, but no statistically significant difference was identified.

## Impact of Lunch Status and Test Preparation

This analysis aims to determine 1) whether there is a significant difference between test scores of students with free/reduced lunches compared with standard lunch students and 2) whether test preparation has a more significant impact on students with free/reduced lunches compared to those with standard lunches. The relationship between test preparation and student performance in math, reading, and writing is tested if it is influenced by the lunch status of the student.

### Data Overview

This analysis consists of two key categorical variables:

- **Lunch**: Indicates whether a student receives free/reduced lunches or standard lunches.
- **Test Preparation**: Indicates whether a student completed a test preparation course or not.

The dependent variables (outcomes) are:

- **Math Score**: Numeric test score in mathematics.
- **Reading Score**: Numeric test score in reading.
- **Writing Score**: Numeric test score in writing.

Bayesian modeling is performed for each test score variable to assess whether the interaction between lunch status and test preparation has a statistically significant impact on the scores.

**Bayesian Model**

We will use a hierarchical model structure to account for the potential differences in impact across the students. The model is:

$$Score_i \sim Normal(\mu_i, \sigma^2)$$
$$\mu_i = \alpha + \beta_1 \cdot Lunch_i + \beta_2 \cdot TestPrep_i + \beta_3 \cdot (Lunch_i \times TestPrep_i)$$

Where:

- $\alpha$ is the intercept (baseline score).
- $\beta_1$ captures the effect of lunch status.
- $B_2$ captures the effect of test preparation.
- $\beta_3$ captures the interaction between lunch status and test preparation, which is the main parameter of interest.

By using Bayesian inference, we aim to estimate the posterior distributions of these parameters, particularly focusing on $\beta_3$, to evaluate whether test preparation has a differential impact based on lunch status.

**Credible Intervals**

Bayesian credible intervals give us a range of values within which the true parameter is likely to lie, with a certain probability (e.g., 95%). In this context, we will assess the 95% credible interval for $\beta_3$. If this interval excludes zero, we can infer that there is a statistically significant interaction between lunch status and test preparation.

Here's an outline of the steps for modeling this data as performed in the R script:

1. **Data Preprocessing**: Convert categorical variables into numerical indicators (0/1 coding for lunch and test preparation).
2. **Prior Selection**: We will use weakly informative priors for the regression coefficients ($\alpha$,$\beta_1$,$\beta_2$,$\beta_3$) to avoid strong assumptions while allowing the data to inform the posterior distributions.
3. **Model Fitting**: We will use Markov chain Monte Carlos (MCMC) sampling to draw samples from the posterior distributions of the parameters.

4. **Model Diagnostics**: Check convergence using trace plots and effective sample sizes to ensure the MCMC chains are mixing well.
5. **Credible Intervals**: Calculate the 95% credible intervals for $\beta_3$.

**Visualizations**

95% credible intervals indicated that there is a significant difference in Math scores between students receiving free/reduced lunches and standard lunches (Figure 1). However, the model given the data suggests that there is no significant difference for Reading and Writing test results between the two lunch categories (Figures 2 & 3).

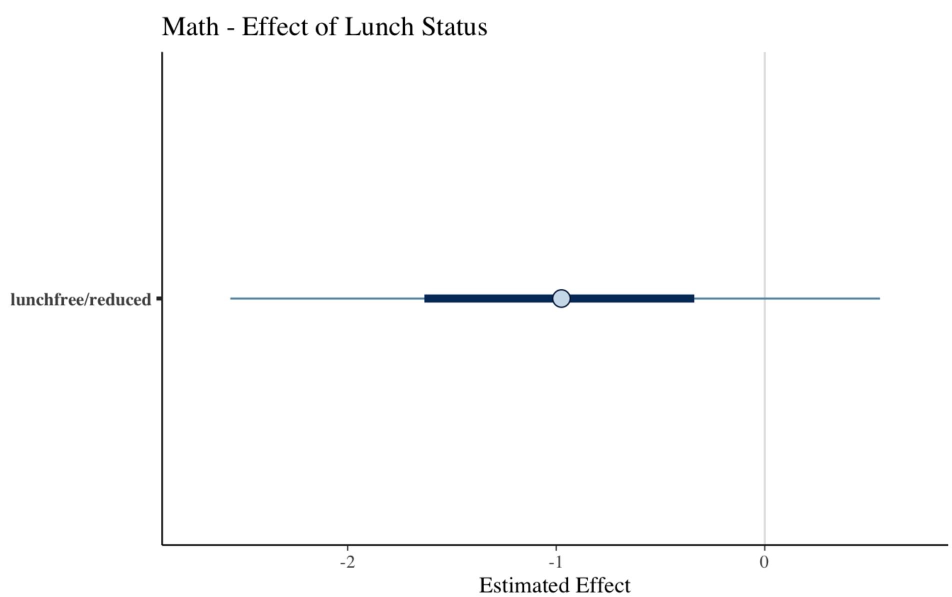


*Figure 1*

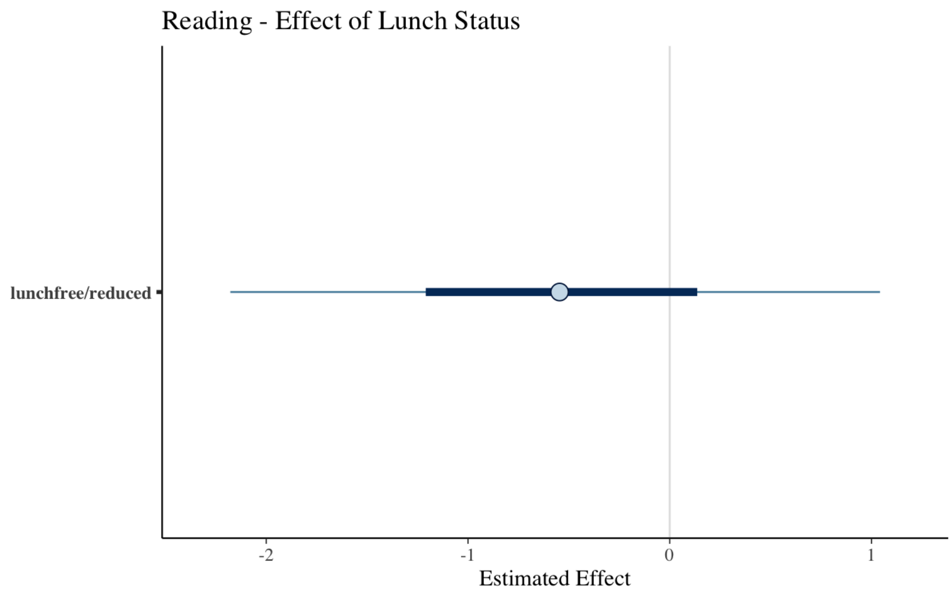


*Figure 2*

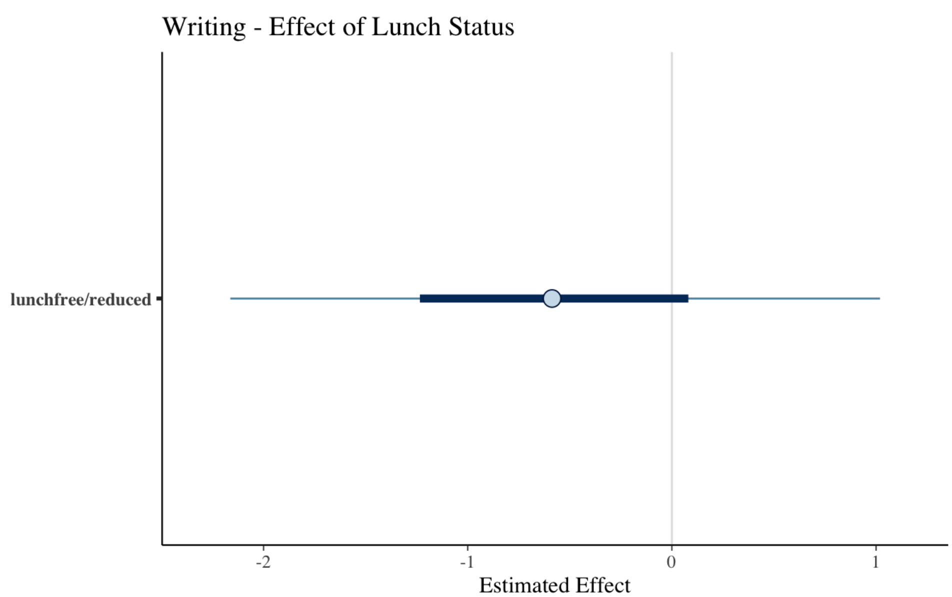


*Figure 3*

95% credible intervals for all three subjects suggest that there is no significant difference in test results for students that have completed test prep among free/reduced lunch students compared to standard lunch students (Figures 4, 5, & 6)

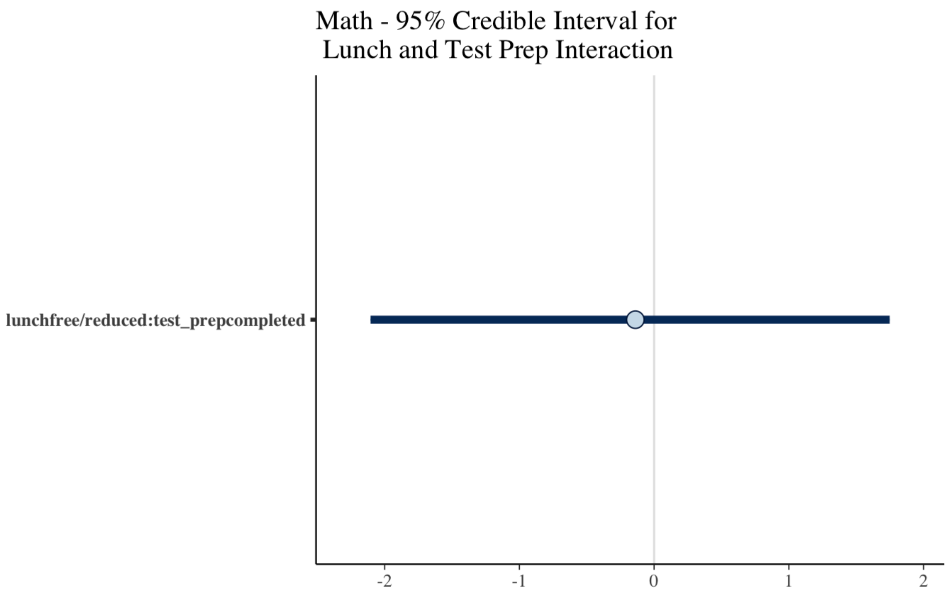


*Figure 4*

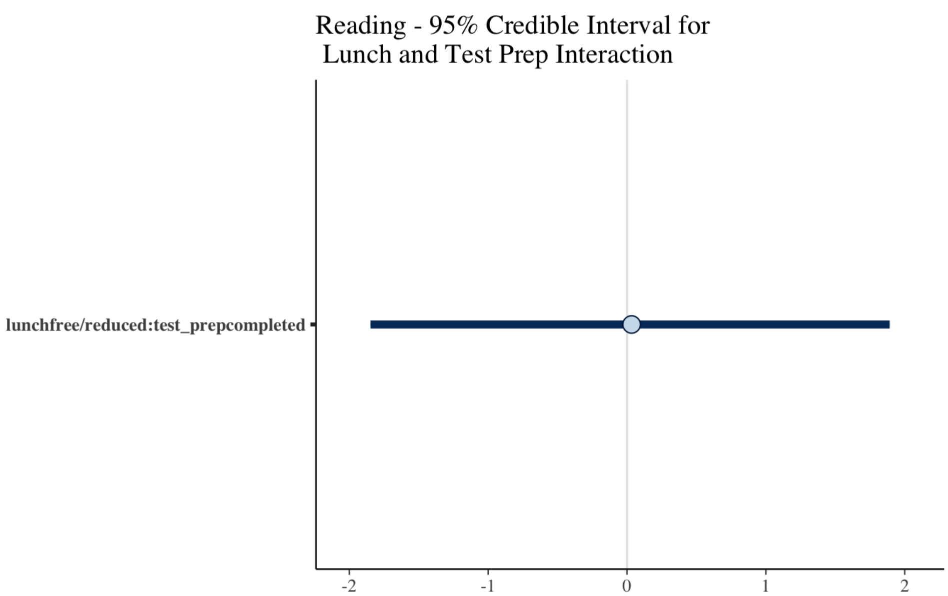


*Figure 5*

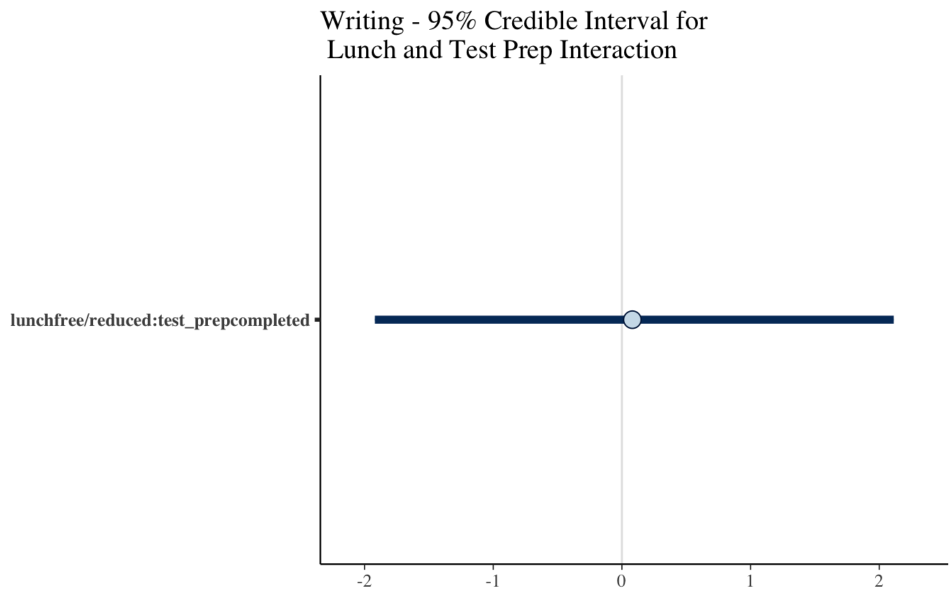


*Figure 6*

## Impact of Test Preparation on Students Based on Gender

This analysis aims to determine whether test preparation has a more significant impact on students based on gender, specifically whether the relationship between test preparation and student performance in math, reading, and writing is influenced by the student's gender.

**Data Overview**

This analysis consists of two key categorical variables:

- **Gender**: Indicates whether a student identifies as male or female.
- **Test Preparation**: Indicates whether a student completed a test preparation course or not.

The dependent variables (outcomes) are:

- **Math Score**: Numeric test score in mathematics.
- **Reading Score**: Numeric test score in reading.
- **Writing Score**: Numeric test score in writing.

Bayesian modeling is performed for each test score variable to assess whether the interaction between gender and test preparation has a statistically significant impact on the scores.

**Bayesian Model**

We will use a similar hierarchical model structure to account for the potential differences in the impact of test preparation across genders. The model is:

$$Score_i \sim Normal(\mu_i, \sigma^2)$$

$$\mu_i = \alpha + \beta_1 Gender_i + \beta_2 \cdot TestPrep_i + \beta_3 \cdot (Gender_i \times TestPrep_i)$$

Where:

- $\alpha$ is the intercept (baseline score).
- $\beta_1$ captures the effect of gender.
- $\beta_2$ captures the effect of test preparation.
- $\beta_3$ captures the interaction between gender and test preparation.

By using Bayesian inference, we aim to estimate the posterior distributions of these parameters, particularly focusing on $\beta_3$, to evaluate whether test preparation has a differential impact based on gender.

**Credible Intervals**

Bayesian credible intervals provide a range of values within which the true parameter is likely to lie, with a certain probability (e.g., 95%). In this context, we will assess the 95% credible interval for $\beta_3$. If this interval excludes zero, we can infer that there is a statistically significant interaction between gender and test preparation.

**Steps for Modeling in R**

Here's an outline of the steps followed for modeling this data:

1. **Data Preprocessing**: Convert categorical variables (gender and test preparation) into factors with appropriate levels.
2. **Prior Selection**: We will use weakly informative priors for the regression coefficients ($\alpha$, $\beta_1$, $\beta_2$, $\beta_3$) to avoid making strong assumptions and allow the data to inform the posterior distributions.
3. **Model Fitting**: We will use Markov Chain Monte Carlo (MCMC) sampling to draw samples from the posterior distributions of the parameters.
4. **Model Diagnostics**: Check convergence using trace plots and effective sample sizes to ensure that the MCMC chains are mixing well.
5. **Credible Intervals**: Calculate the 95% credible intervals for $\beta_3$.

**Visualizations**

95% credible intervals for math and reading suggest that there is no significant difference in test scores and gender (Figures 7 & 8). However, the model given the data shows there is a significant difference in writing test scores between male and female students, with females typically performing better (Figure 9).

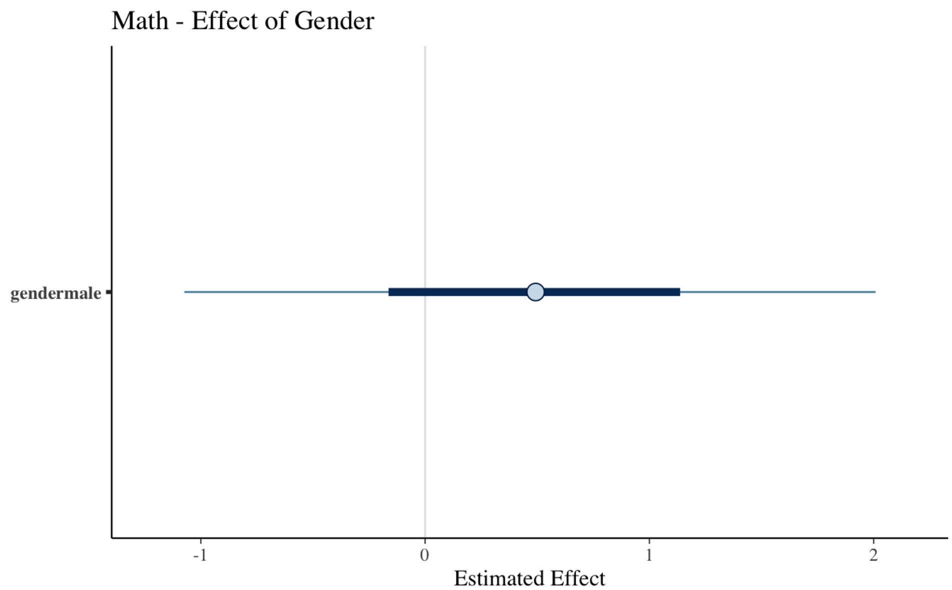


*Figure 7*

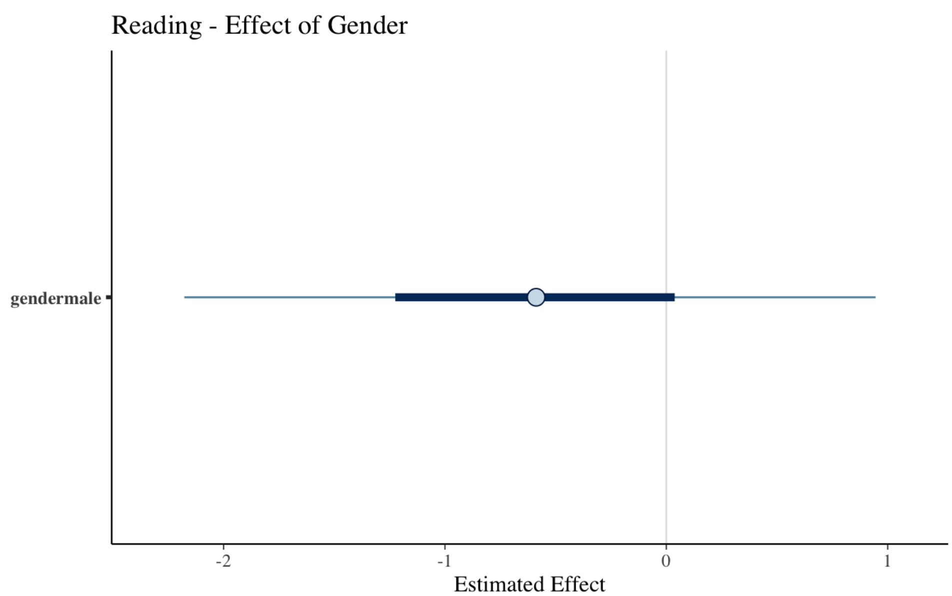


*Figure 8*

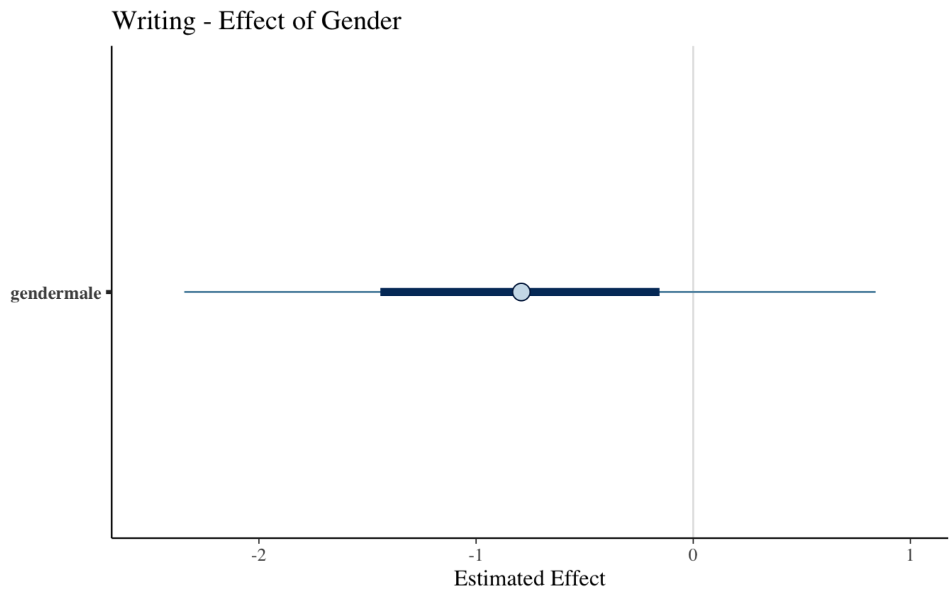


*Figure 9*

95% credible intervals for all three subjects suggest that there is no significant interaction in test results for male or female students that have completed test prep (Figures 10, 11, & 12).

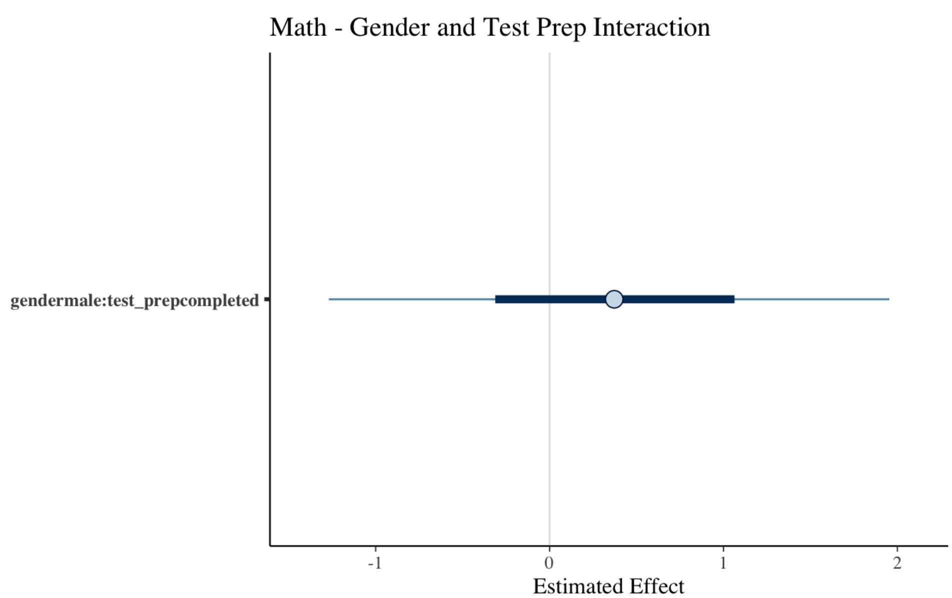


*Figure 10*

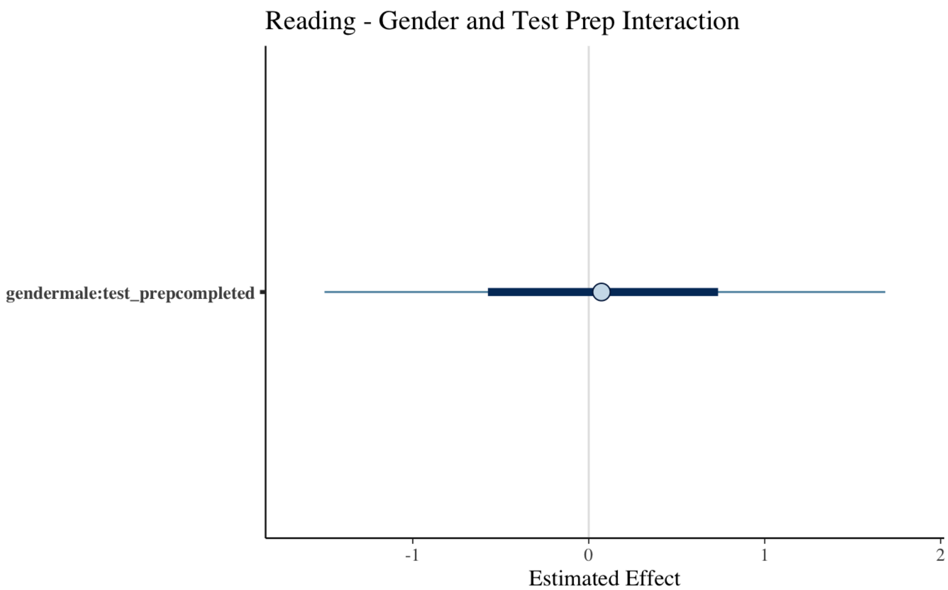


*Figure 11*

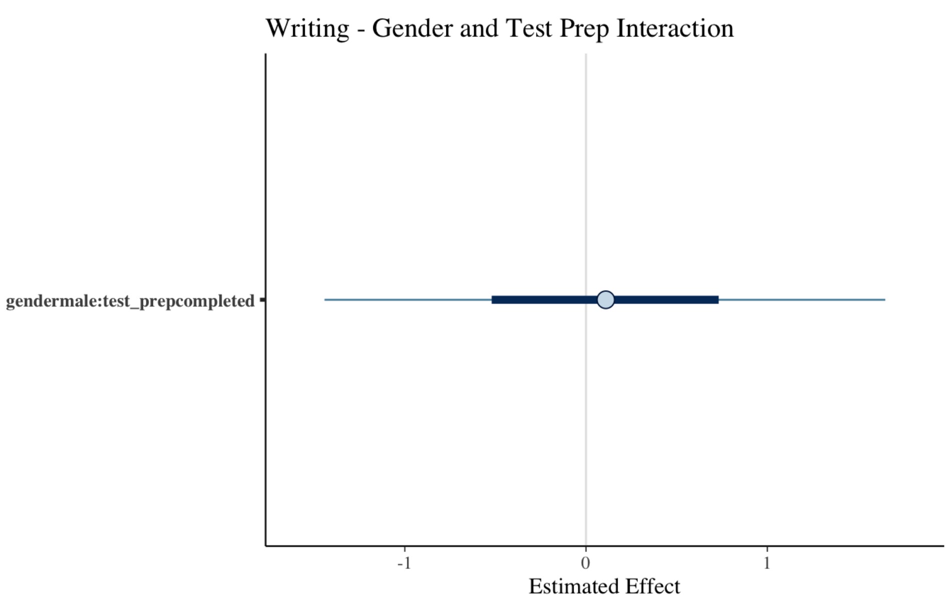


*Figure 12*

# Results and Discussion

## Test Prep and Lunch Status

The models show that there is a significant difference in math scores between the two lunch categories (with standard lunch students typically scoring better) and there is not a significant difference in reading and writing scores. The models testing the interaction between test preparation and lunch status did not show a significant interaction. Thus, contrary to likely assumptions, test preparation does not seem to have a more substantial impact on students with free/reduced lunches.

## Test Prep and Gender

When modelling test scores for gender, only writing test scores showed a significant difference. As for the interaction between gender and test preparation, each subject's model again produced posterior distributions where the interaction term's 95% credible interval includes zero, suggesting that there is not a significant interaction between test preparation and gender. Thus, one gender should not be prioritized over the other when planning test preparation courses.

# Conclusions

Bayesian hierarchical models are well-suited for this type of analysis because they can account for uncertainty in small sample sizes and provide credible intervals, which are easier to interpret than traditional p-values. However, while Bayesian models are more resilient to sparse data, the results from this analysis should be treated with caution due to the relatively small sample size. Further data collection would help confirm these findings and further analysis could perform similar tests for other variables available in the data including parent education levels and race.

Ultimately, females were found to perform better in writing. Students receiving standard lunches performed better in math. While test preparation is typically beneficial, the analysis did not show that it benefits any particular student group for gender or lunch status.