

# **Comparing Test Scores Between Boys and Girls**

Adam Bitner, Emily Dunham, Sammy Hilton

## **Introduction**

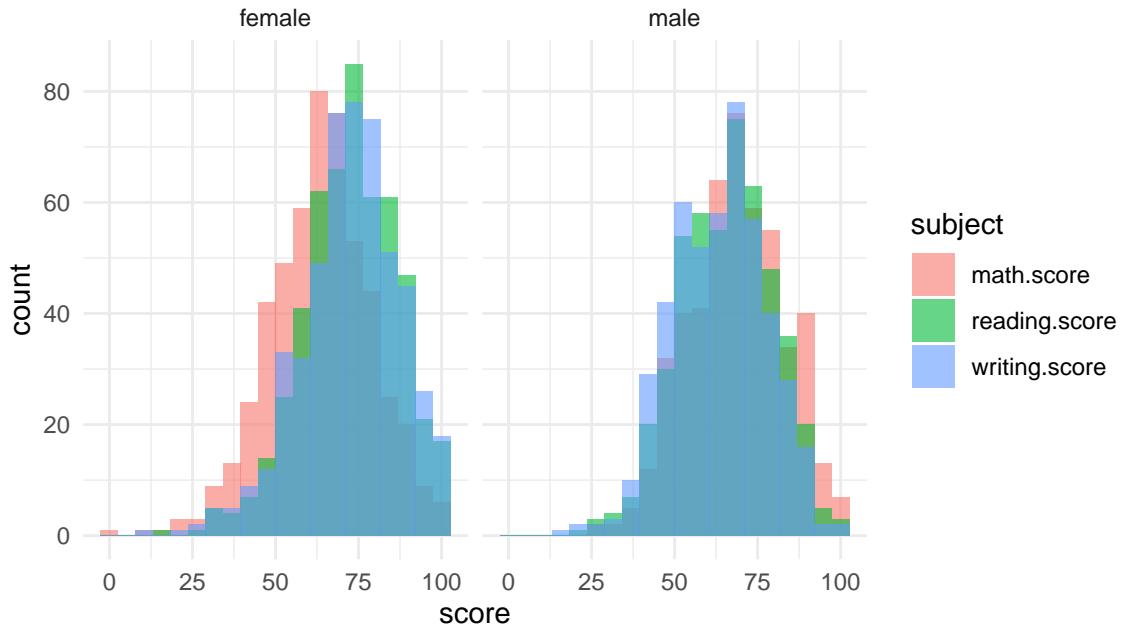
We are interested to see the difference mean score on some test scores between boys and girls. Do girls naturally do better in some subjects than men? Do mean do better in other subjects than girls? We aim to answer this question. This analysis is helpful because it can help teachers better understand who could be predisposed to struggle more in the class. Furthermore this can help universities and educators to understand effects of different teaching methods and give a better base line for expected performance. The American Psychology association has published an analysis claiming that girls do better than boys in school in essentially all subjects, we will be testing the validity of that claim.

## **Methods**

### **Data**

We got our data from Kaggle. The data contains test scores from high school students in the United States. It has test scores for both boys and girls for math, writing, and reading.  
<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

## Distribution of Exam Scores by Gender



We can see the mean math score overall is 66.089, the mean writing score is 68.054 and the mean reading score is 69.169.

### Getting our Likelihood and Priors

The relevant parameters for our likelihood that we chose would be from a normal distribution. Since we believe that grades on a test will behave in a distribution that is likely to look more normal where most students will do generally well and only a few will get really high scores and a few will get very low scores. This will be useful in answering our question because we know that for most students this will generally be the pattern. For our prior distribution we decided that our priors for both boys and girls will be the same. However we still need two priors, one for the mean and one for the standard deviation of our likelihood.

$$\text{Likelihood} = N \sim (\mu, \sigma^2)$$

For the mean we chose a normal prior with  $\mu = 80$  because we believe that most students will average around 80. We chose  $\sigma^2 = 5$  because we believe that our guess about the mean being 80 might be off by around 5 points.

$$\mu = N \sim (80, 5^2)$$

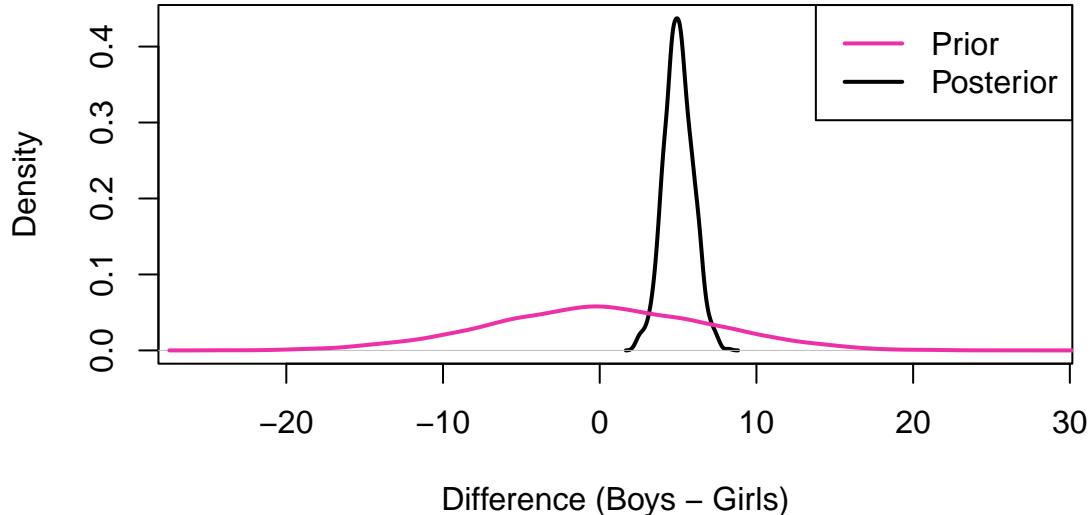
For the standard deviation we chose an inverse gamma prior. In order to find what our  $a$  and  $b$  values would be we chose what we expected the value of  $\sigma^2$  would be for our likelihood and what we thought the standard deviation of that would be. In our case we chose the mean and standard deviation to be 10 and 4 respectively. Then using method of moments we took those values and calculated  $a$  and  $b$ . From our results we got that  $a = 3.6$  and  $b = 256.3$ .

$$\sigma^2 = \text{IG} \sim (3.6, 256)$$

## Results

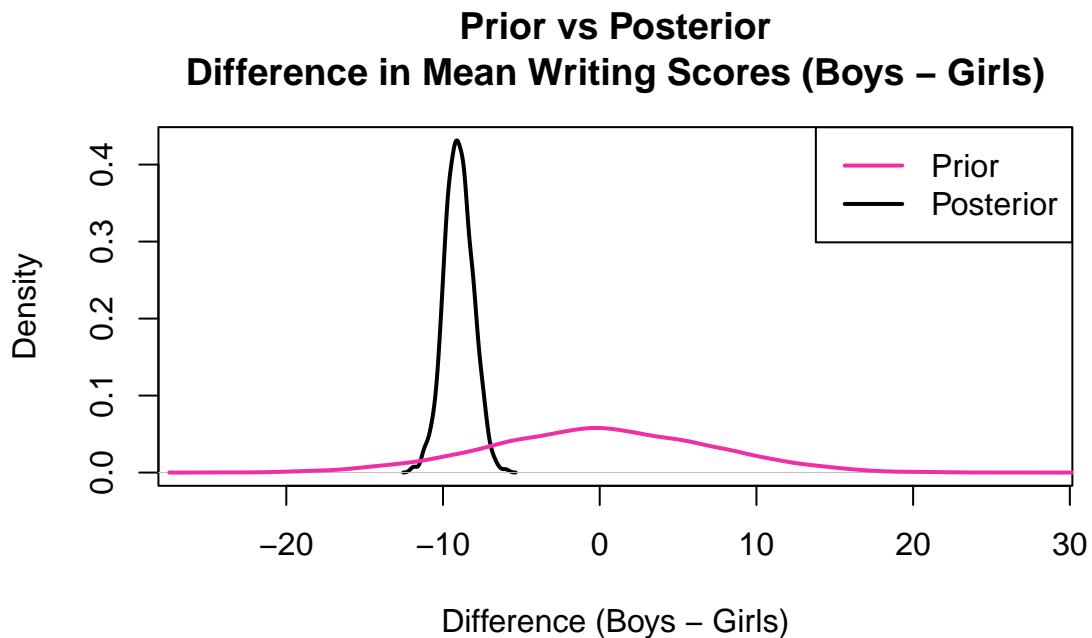
### Posterior for Math Test Scores

**Prior vs Posterior  
Difference in Mean Math Scores (Boys – Girls)**



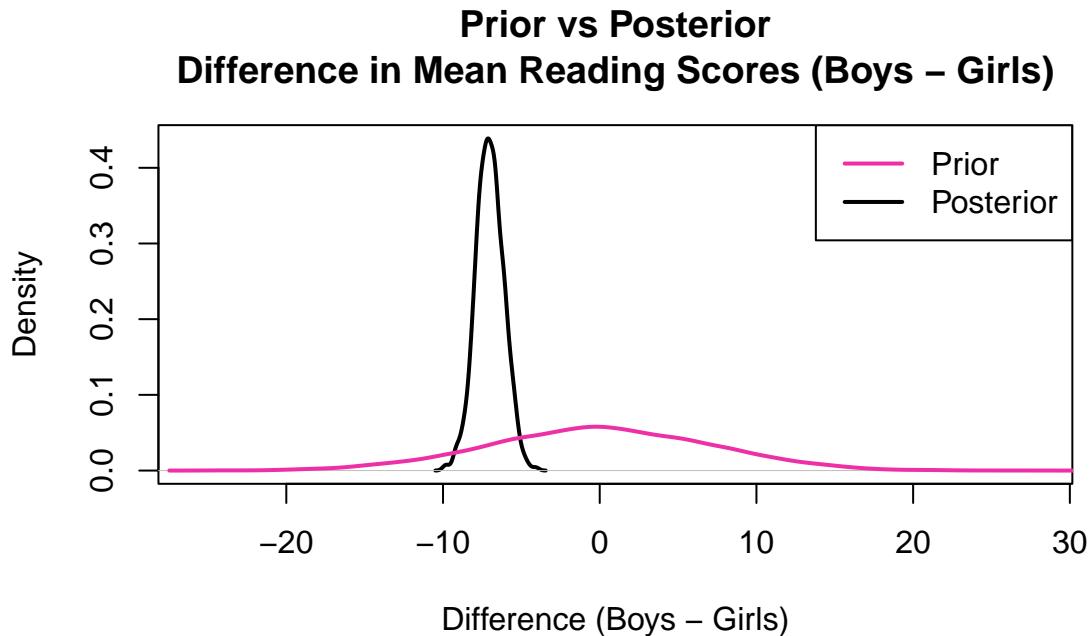
Based on our posterior we are 95% sure that the difference between boys and girls math scores is between 3.14 and 6.79 points. Since 0 is not in our interval we can conclude that boys are better at math than girls.

## Posterior for Writing Test Scores



Based on our posterior we are 95% sure that the difference between boys and girls writing scores is between -10.81 and -7.28 points. Since 0 is not in our interval we can conclude that girls are better at writing than boys.

## Posterior for Reading Test Scores



Based on our posterior we are 95% sure that the difference between boys and girls writing scores is between -8.77 and -5.32 points. Since 0 is not in our interval we can conclude that girls are better at reading than boys.

## Discussion

After our analysis, we found that boys and girls score better in different subjects. Boys score better than girls in the subject of math. We found that girls score better than boys in writing and reading. These differences may reflect different interests of boys vs. girls in school subjects. Understanding these differences can help teachers design targeted interventions to support students in subjects where they may struggle.

There are some limitations to our analysis. We are gonna use a normal posterior to model these, but technically we are assuming that these students could get more than 100 or less than 0, so that's a limitation. Another limitation is that we only have 3 subjects, so we could do another analysis with more school subjects. This is also only for students in the U.S., so we don't have data from students world wide.

Some follow up questions could be - Do boys vs. girls perform better in other subjects of interest? Do lower vs. upper class students perform better on math, reading, and writing?

## **Appendix**