

A Data Analysis of 2009-2010 CSAP Results of 9th Graders

Benjamin Javier (110787897)

4/29/2024

Ask the Question

The questions that the data will be used to answer are the following: “Was there truly an improvement in passing rates for the math exam between 2009 and 2010? If so, was there a decline in the performance on the Reading and Writing exams?” Essentially, we must compare the passing rates for the math exam between 2009 and 2010 and determine if it increases over time. Then we must also determine if there was a decrease in passing rates for the reading and writing exams IF the passing rates for math increased.

“Was there an association between a school’s passing rate on the math exam and its passing rates on the reading and writing exams? If so, how accurately could math passing rates have been predicted from the reading and writing rates?” In order to come up with an answer based on the data, the data must be wrangled, analyzed, and then after advising on the results, only then can an answer be made on the previous two questions. Therefore, the data set must be wrangled appropriately first.

Acquire the Data

By wrangling the data before sampling, we can ensure a clean and tidy dataset. This allows us to analyze the correct data, ensuring there are no observations that could affect the data analysis in a negative manner. To answer the questions stated in the first A: Ask the Question(s), we can narrow the previous data into the above variables, placing emphasis on the passing rates for each subject (proficient and advanced) and later observing the differences between 2009 and 2010. Doing this will leave us with 720 observations and 8 variables, containing the school No, subject, year, and the number of each proficiency level. This is seen in the below data structure:

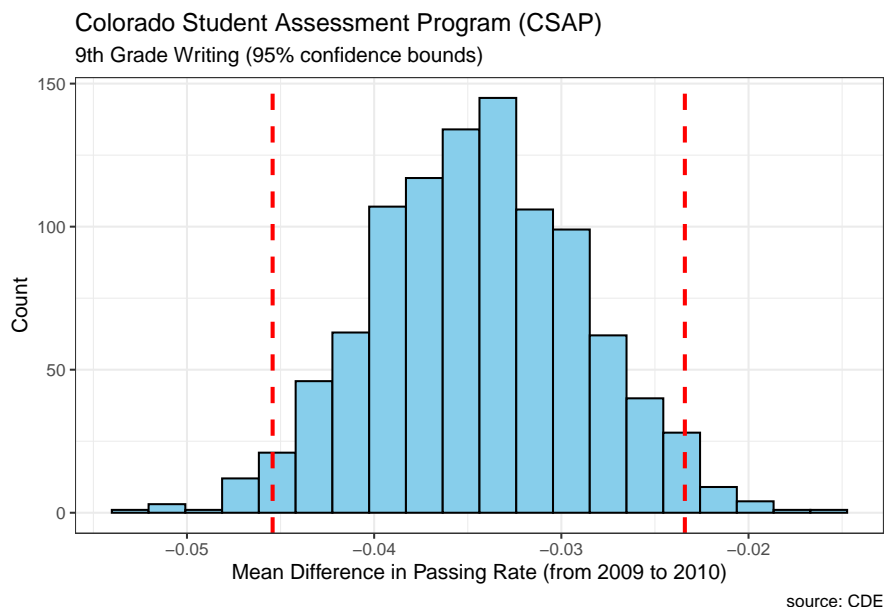
```
## Rows: 720
## Columns: 8
## $ schoolNo      <dbl> 187, 221, 309, 503, 210, 1752, 5043, 6060, 24, 7129, 77~
## $ subject1      <fct> READING, READING, READING, READING, READING, READING, R~
## $ year          <dbl> 2009, 2009, 2009, 2009, 2009, 2009, 2009, 2009, 2009, 2~
## $ none          <dbl> 0, 5, 3, 0, 6, 75, 3, 2, 17, 18, 0, 68, 4, 15, 7, 5, 9,~
## $ unsatisfactory <dbl> 13, 10, 11, 4, 3, 22, 20, 18, 76, 49, 4, 93, 26, 22, 14~
## $ partial       <dbl> 35, 34, 36, 16, 20, 91, 68, 129, 180, 121, 24, 258, 74,~
## $ proficient    <dbl> 48, 28, 28, 28, 13, 227, 403, 367, 148, 232, 63, 266, 7~
## $ advanced      <dbl> 1, 0, 0, 1, 0, 11, 21, 6, 1, 5, 1, 1, 0, 2, 1, 37, 21, ~
```

Analyze the Data

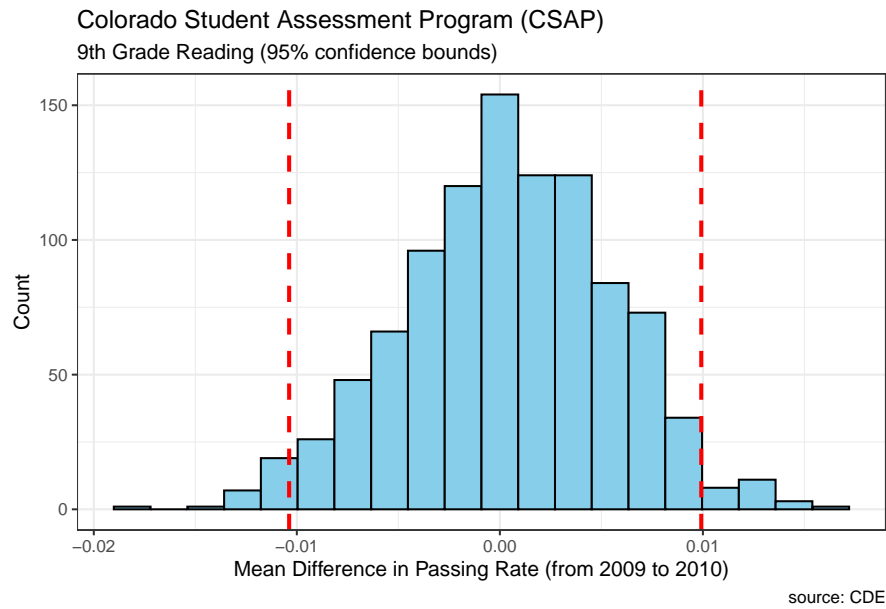
To properly analyze and find the difference in proportion for passing and non passing test results for each subject, we must create other data frames that will split off from the main one into the three respective subjects and filter for the passing and non passing results. In this case, what we are looking at is essentially an exploratory analysis, taking the test results from 2009 and comparing them to 2010 test results. We can do this by creating proportion

data for each subject from the wrangled data, taking the difference in passing rates between 2009 and 2010. The mean differences for each can be seen below:

Writing mean difference: -3.457956%, Math mean difference: 3.7769344%, Reading mean difference: 0.0473492%. The mean differences seem to suggest that the writing passing rates decreased by 3%, the math passing rates increased by 3% and the reading passing rates increased by 0.04%, or even no real change at all. However, all three of these values are point estimates, so we must create the respective confidence interval for each subject's passing rate. We will go in with a 95% confidence interval. We can then analyze and observe the changes through each respective plot.



Based on the plot for the writing pass rates, it seems that the passing rate for writing decreased between 2009 and 2010. Using the confidence interval bounds, we are 95% confident that the passing rates decreased between -4.5424496 and -2.340184 percentage points respectively between 2009 and 2010 for 9th grade test takers at the average Colorado school.

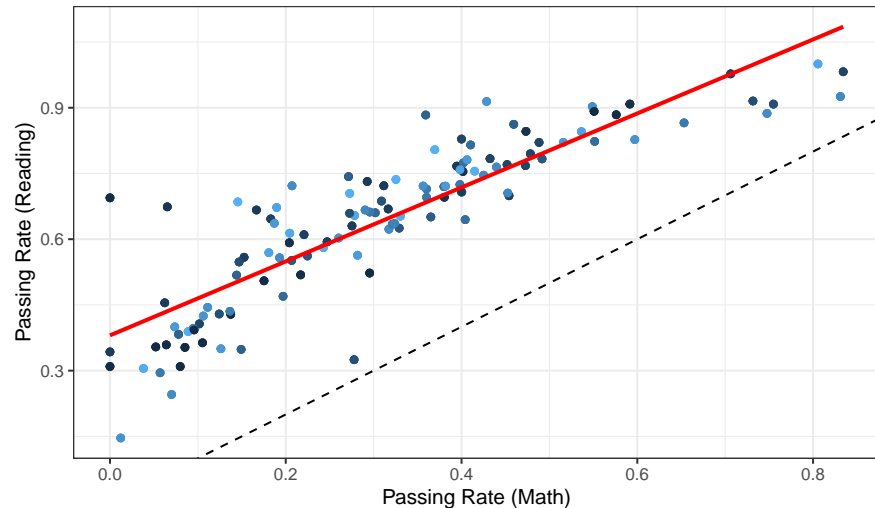


Based on the plot for the reading pass rates, it seems that there can be no observed difference in the passing rates between 2009 and 2010. This is because zero is included in the confidence interval, suggesting that the pass rates in reading between 2009 and 2010 underwent no real difference.

Based on the plot for the math pass rates, it seems that the passing rate for math increased between 2009 and 2010. Using the confidence interval bounds, we are 95% confident that the passing rates increased between 2.6968152 and 4.806337 percentage points respectively between 2009 and 2010 for 9th grade test takers at the average Colorado school.

Following that exploratory analysis on the data, we can also form some predictive analysis on the data set to predict results based on other subjects. For example, we can take the data from the math passing rates in 2009 and create a line of best fit equation to predict the reading and writing pass rates given a math pass rate. This is done in the below scatter plots, and its proceeding analyses:

Colorado Student Assessment Program(CSAP)
9th Grade Math vs Reading (2009)



Source: CDE

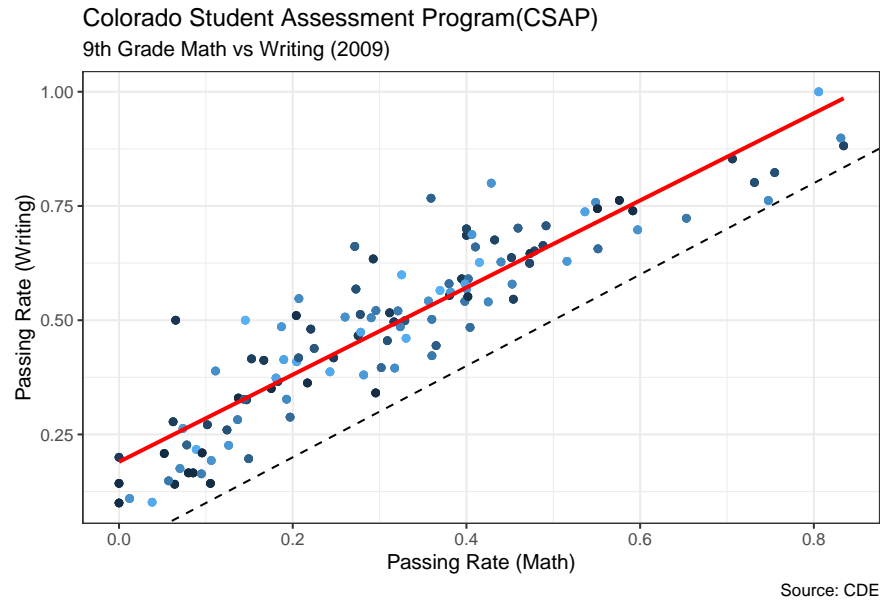
In the above graph we can see that the graph's intercept is above 0, so students tend to pass more on the reading test than on the math test. We can also see that the correlation can be described as linear, increasing and strong. This is further supported by the correlation coefficient: 0.8741303. The regression equation is seen below:

$$READING = 0.3806 + 0.8442 \cdot MATH$$

In the above equation, we observe that the intercept is greater than 0 and the slope is less than one. That means that students typically perform better in the reading exams than in math, and as the passing rates for math increase, the gap between reading and math pass rates will decrease. This makes sense logically as students will typically cover all subjects of a test to feel better prepared for the overall exam.

```
##          RMSE
## 1 0.100029
```

This value says that if we are given a math passing rate, we can predict the reading passing rate to be within 10.0029005 percentage points.



In the above graph, we observe results that are relatively similar to that of the previous graph, in that the correlation is strong, positive and linear, seen with 0.9118033. Again, as is the case with the Math vs Reading graph, we can observe that the intercept for the line of best fit is clearly above 0, therefore we can say that students tend to do better on the writing test than on the math test. This is seen in the below equation:

$$WRITING = 0.1901 + 0.9534 \cdot MATH$$

In the equation, we also observe that the slope is less than one, so we can say that as the pass rates for math increase, the gap between the writing and math pass rates will decrease. Next, we can find the RMSE value in order to analyze the accuracy of the above model. The RMSE is given below:

```
##          RMSE
## 1 0.1035294
```

This value above essentially means that given a math passing rate, we can predict a writing passing rate within 10.3529383 percentage points.

Advise on Results

In the first half of the previous section, we were able to observe a roughly 3% increase in math pass rates between 2009 and 2010 for 9th grade students taking the CSAP exam. We were also able to observe a roughly 3% decrease however in writing pass rates between 2009 and 2010 for the same data set. Lastly, we were unable to observe an increase nor a decrease in reading pass rates between 2009 and 2010 for the data set. This allows us to answer the first question based on the results, in that there was an observed increase in the pass rates for the math exam, however writing pass rates decreased while reading pass rates stayed relatively the same.

In the second half of the previous section we were able to create predictive equations that would let us predict the reading and writing pass rates given a math pass rate. We observed that students typically tend to do better in writing and reading than they do in math, however as the pass rates for math increase, the gap between math pass rates and writing/reading pass rates will decrease. Afterwards, we were able to determine an accuracy value for which the average difference between the actual and predicted value will be. This is roughly 10 percentage points for both predictive equations, but we can expect that when making a prediction, the difference, on average, will be within 10 percentage points.

Answer the Question

Using the results from the data, we can finally answer the questions asked by the administrators, which were established at the very beginning, before we even acquired the data. The first question asked wanted us to observe whether there was an improvement in passing rates for the math exam between 2009 and 2010, and if so, if there as a decline in performance on the reading and writing exams. Using the results from the exploratory analysis, in the confidence intervals for each subject, we observed a roughly 3% increase in math exam pass rates. To be more specific however, we are 95% confident that the math exam pass rates increased between 2.7%-4.8% between 2009 and 2010. This allows us to move onto the second part of the question, which wants us to observe if the writing or reading exam pass rates decreased between 2009 and 2010. For the reading exam pass rates, we observe no change between 2009 and 2010. This is because when taking the 95% confidence interval for the reading pass rates, we found 0 to be within the confidence interval. Therefore, we are 95% confident that the reading pass rates underwent no change between 2009 and 2010. However, for writing exam pass rates we observe a decrease in roughly 3% in the pass rates. More specifically, we are 95% confident that the writing exam pass rates decreased between 2.3%-4.5% between 2009 and 2010. This allows us to answer the question as a whole, that based on the observed confidence intervals, we observe that math exam pass rates increased between 2009 and 2010, while reading pass rates had no change and writing pass rates decreased between 2009 and 2010.

To answer the second question, which asked us to observe if there was an association between a school's passing rate on the math exam and the passing rates on the reading and writing exams. Following that, the administrators want us to determine the accuracy of this association,

given that we can predict the passing rates with a line of best fit. To find the answer to this question, we can apply methods of predictive analysis, seen in the second half of the analysis section. We created two scatter plots and two lines of best fit equations to create a predictive model on the passing rates in reading and writing given a math pass rate. In both graphs, we were able to observe a strong, linear, and positive correlation, a positive intercept in both, and a slope of less than one in both. This tells us that students tend to pass reading and writing more than math, and that as the math pass rates increase, the gap between reading/writing pass rates and math pass rates will decrease. Furthermore, we were able to give our predictive models an accuracy value by calculating an RMSE for both. In the RMSE we observed a value of roughly 0.10 for both, or 10 percentage points. That means that given a math pass rate, we can observe on average our model to predict the writing and reading pass rates to be accurate within 10 percentage points.