Term Project: Exam Performance Analysis

Jasmeet Singh Saini - 0758054

2023-04-15

Introduction

In the modern education system, evaluating student performance in academic subjects is an essential part of the learning process. Academic assessments provide valuable feedback on the effectiveness of teaching methods, curriculum design, and student learning outcomes. Performance in exams is one of the most common methods used to evaluate student's understanding of a subject. However, performance in exams can be influenced by various factors, such as student demographics, socio-economic status, and motivation level.

Moreover, the field of education has undergone significant transformation due to technological advancements, globalization, and cultural diversity. With the advent of digital learning, student's performance evaluation is becoming increasingly data driven. To enhance learning outcomes, educators need to collect and analyze data on various aspects of the educational process, such as student performance, behavior, and learning preferences.

Literature Review

Several studies have explored the factors that affect student performance in exams such as demographic characteristics, parental involvement, teaching effectiveness and many more.

Research done on demographic factor such as gender, race/ethnicity, socio-economic status, and parental education level, have a significant impact on academic achievement (Sharma & Gupta, 2019; Singh, 2020). Studies have found that girls tend to perform better than boys in reading and writing, while boys tend to perform better in math (Eccles et al., 1990; Hyde, Fennema, & Lamon, 1990). In terms of race/ethnicity, research has shown that minority students, such as African American and Hispanic students, tend to perform lower than their White and Asian peers (Ferguson, 2003; Fryer & Levitt, 2004). Socio-economic status has also been found to be a significant predictor of academic achievement, with children from lower socio-economic backgrounds experiencing lower levels of academic success (Singh, 2020).

Another studies have shown that when parents are involved in their children's academic pursuits, it can have a positive impact on their academic performance. Research found that there is a correlation between parental involvement and academic success (Liao, Gao, & Liu, 2020; Suhaimi, Ahmad, & Musa, 2018). In fact, a meta-analysis of 31 studies conducted by Liao and colleagues in 2020 found that parental involvement had a moderate effect on academic achievement. Parental involvement can manifest in various forms such as helping with homework, attending school events, and communicating with teachers. As a result, it's crucial for parents to take an active interest in their children's education to help them succeed academically.

Research has consistently shown that teacher effectiveness is a critical factor that significantly impacts student performance in exams. In fact, effective teaching practices have been found to be directly correlated with improved student learning outcomes. A number of studies have shown that when teachers provide timely and specific feedback to their students, it helps students to identify areas where they need to improve, understand their strengths, and ultimately learn more effectively. Additionally, using active learning strategies in the

classroom, such as problem-based learning or inquiry-based learning, can help students to engage more deeply with the material and develop critical thinking skills. Lastly, setting high expectations for students can be an effective way to motivate them to work harder and achieve more. When teachers set high standards for their students, they signal to their students that they believe in their potential, which in turn can lead to increased effort and higher levels of achievement. Overall, these effective teaching practices have been shown to have a positive impact on student learning outcomes across a range of subjects and grade levels (Hattie, 2009; Marzano, Pickering, & Pollock, 2001).

Socioeconomic status (SES) is a crucial factor that determines student performance in exams. Various studies have consistently demonstrated that students from low SES backgrounds tend to perform poorly compared to those from high SES backgrounds. A study conducted by Gorard, See, & Davies (2012) highlighted that students from low SES backgrounds have lower academic achievement due to factors such as limited access to educational resources, poor nutrition, and inadequate healthcare (Gorard, See, & Davies, 2012).

Similarly, a report by the OECD (2016) indicated that students from families with higher income, higher parental education levels, and greater access to educational resources tend to perform better in exams than those from lower SES backgrounds. The study emphasized that students from low SES backgrounds are more likely to face financial constraints and inadequate educational resources that limit their academic performance. Moreover, research conducted by Sirin (2005) found that students from low SES backgrounds tend to have lower grades and academic achievement than those from high SES backgrounds. The study highlighted that low SES students often face challenges such as poor academic preparation, limited access to educational resources, and negative stereotypes that affect their academic performance.

Race and gender are also significant predictors of student performance in exams. A study by Reardon et al. (2019) found that the black-white achievement gap in standardized test scores in the US has remained relatively constant for the past 50 years, despite various efforts to reduce it. The study emphasized that this achievement gap is a result of several factors, including historical and ongoing racial discrimination, income inequality, and limited access to quality educational resources. Moreover, gender differences have also been observed in student performance in exams, with males performing better in subjects such as math and science (Else-Quest, Hyde, & Linn, 2010). The study attributed this difference to various factors, including socialization, gender stereotypes, and differences in cognitive ability. In summary, Socioeconomic status (SES), race, and gender are significant factors that impact student performance in exams. Understanding these factors can help policymakers develop effective strategies to promote educational equity and improve academic outcomes for all students.

All the factors which can influence student's exam scores is an important area of research that can help educators and policymakers improve educational outcomes for students. In this empirical analysis, we will explore the factors that affect student performance in exams using the student's performance in exams dataset. This dataset includes information on the performance of 1,000 students from different demographic backgrounds on three exams: math, reading, and writing. By analyzing this dataset, we aim to gain a deeper understanding of the factors that influence student's exam performance and how educators can improve learning outcomes.

The purpose of the research study is to investigate the impact of different factors on student's exam scores. The study aims to answer four research questions that will help in understanding how these factors influence academic performance.

The first research question focuses on the impact of gender on student's exam scores. This question aims to determine if there is a significant difference in the exam scores of male and female students. The research will investigate whether gender has a direct or indirect impact on academic performance. The study will analyze the data collected from male and female students to understand the correlation between gender and exam scores. The results of this research question will provide insights into gender differences in academic performance and can be useful for developing strategies to improve academic outcomes for all students.

The second research question aims to examine the impact of the level of education of the student's parents on their exam scores. This question will investigate whether students whose parents have higher levels of education tend to perform better on exams than those whose parents have lower levels of education. The research will analyze the data collected from students with different parental educational backgrounds to

determine if there is a correlation between parental education and academic performance. The results of this research question will be useful for understanding the role of parents in the academic success of their children. The third research question focuses on the impact of completing a test preparation course on student's exam performance. This question aims to determine whether students who complete test preparation courses tend to perform better on exams than those who do not. The research will analyze the data collected from students who have completed test preparation courses and those who have not. The results of this research question will be useful for evaluating the effectiveness of test preparation courses and can help in designing strategies to improve academic performance.

Finally, the fourth research question aims to determine if there is a significant difference in exam scores based on student's race/ethnicity. This question will investigate whether students from different racial or ethnic backgrounds tend to perform differently on exams. The research will analyze the data collected from students from different racial/ethnic backgrounds to determine if there is a correlation between race/ethnicity and academic performance. The results of this research question will be useful for understanding the impact of race/ethnicity on academic performance and can help in developing strategies to promote educational equity.

In conclusion, the modern education system recognizes that evaluating student performance is essential for assessing the effectiveness of teaching methods, curriculum design, and learning outcomes. However, the performance of students in exams is influenced by factors such as demographics, socio-economic status, and teacher effectiveness. This study highlights that various factor, such as parental involvement, teacher effectiveness, and socio-economic status, significantly affect student performance. Furthermore, previous research has consistently shown that students from low socio-economic backgrounds, minority students, and female students tend to perform poorly in exams. The study emphasizes that identifying these factors and understanding their impact on student performance is crucial for educators to improve the learning outcomes of their students. The next section will provide an overview of the methods used in this study using inferential statistical analysis and summarize the key findings by visualization.

Methods

To perform inferential statistics on the impact of different factors on student's exam score. We will first look into the *dataset* and then we will initiate **cleaning and preparing the data**. After preparing the data, the subsequent step is to perform **exploratory data analysis** (EDA) to comprehend the distribution and association of the variables in the *dataset*. This phase entails generating visual representations, such as scatter plots, histograms, and box plots, to recognize any trends or patterns in the data. Then after drawing insights from the exploratory data analysis, we can develop **hypotheses** that correspond to the research inquiries of the project. After developing hypotheses, we will look for certain **assumptions** like independence of data, randomly sampled data, normality, variance, and linearity. Then the pre-final step is to performing **inferential statistics** to test the hypotheses. This procedure involves executing the selected statistical tests and analyzing the outcomes.

Data Description

The **Student Performance in Exams dataset** is a collection of data on student performance in various subjects. The variables in the dataset are:

1. **Gender**: The student's gender, either "male" or "female".
2. **Race/Ethnicity**: The student's race/ethnicity is categorized as "Group A", "Group B", "Group C", "Group D", or "Group E".
3. **Parental Level of Education**: The highest level of education attained by the student's parents, categorized as "some high school", "high school", "some college", "associate's degree", "bachelor's degree", or "master's degree".

4. **Lunch**: Whether or not the student received free or reduced-price lunch, either "standard" or "free/reduced".
5. **Test preparation course**: Whether or not the student completed a test preparation course, either "none" or "completed".
6. **Math score**: The student's score on the math exam is out of 100.
7. **Reading score**: The student's score on the reading exam is out of 100.
8. **Writing score**: The student's score on the writing exam is out of 100.

The *dataset* includes demographic information about the students, such as their gender, race/ethnicity, and parental education level, as well as information about their performance on three exams: math, reading, and writing. The *dataset* contains 1,000 observations and 8 variables. Below is a table summarizing the data:

Variable	Datatype	Key	Description
Gender	Categorical	Male, Female	Student's gender.
Race	Categorical	Group: A, B, C, D, and E	Student's race/ethnicity.
Lunch	Categorical	Standard, Free/reduced	Whether the student has free/reduced lunch or not.
Preparation Course	Categorical	None, Completed	Whether the student has taken the test preparation course or not.
Math Score	Numeric	0 - 100	The student's score on the math exam
Reading Score	Numeric	0 - 100	The student's score on the reading exam
Writing Score	Numeric	0 - 100	The student's score on the writing exam
Parental Level of Education	Categorical	Types of School	Student's parental level of education, that is, High School, Some College, Associate's Degree, Bachelor's Degree, Master's Degree

Data Collection and Cleaning

We will begin by importing the dataset into R using the read.csv() function. We will then check for any missing values in the dataset and remove them using the na.omit() function. We will also check for any duplicates in the dataset and remove them using the unique() function. Finally, we will rename the variables to make them more descriptive using the names() function.

```
# Importing the dataset
students_data <- read.csv("StudentsPerformance.csv")
# Checking the data frame
# str(students_data)

# Checking for missing values
students_data <- na.omit(students_data)

# Checking for duplicates
students_data <- unique(students_data)

# Renaming the variables
```

```
names(students_data) <- c("gender", "race_ethnicity", "parental_education"
                          , "lunch", "test_preparation_course", "math_score"
                          , "reading_score", "writing_score")
```

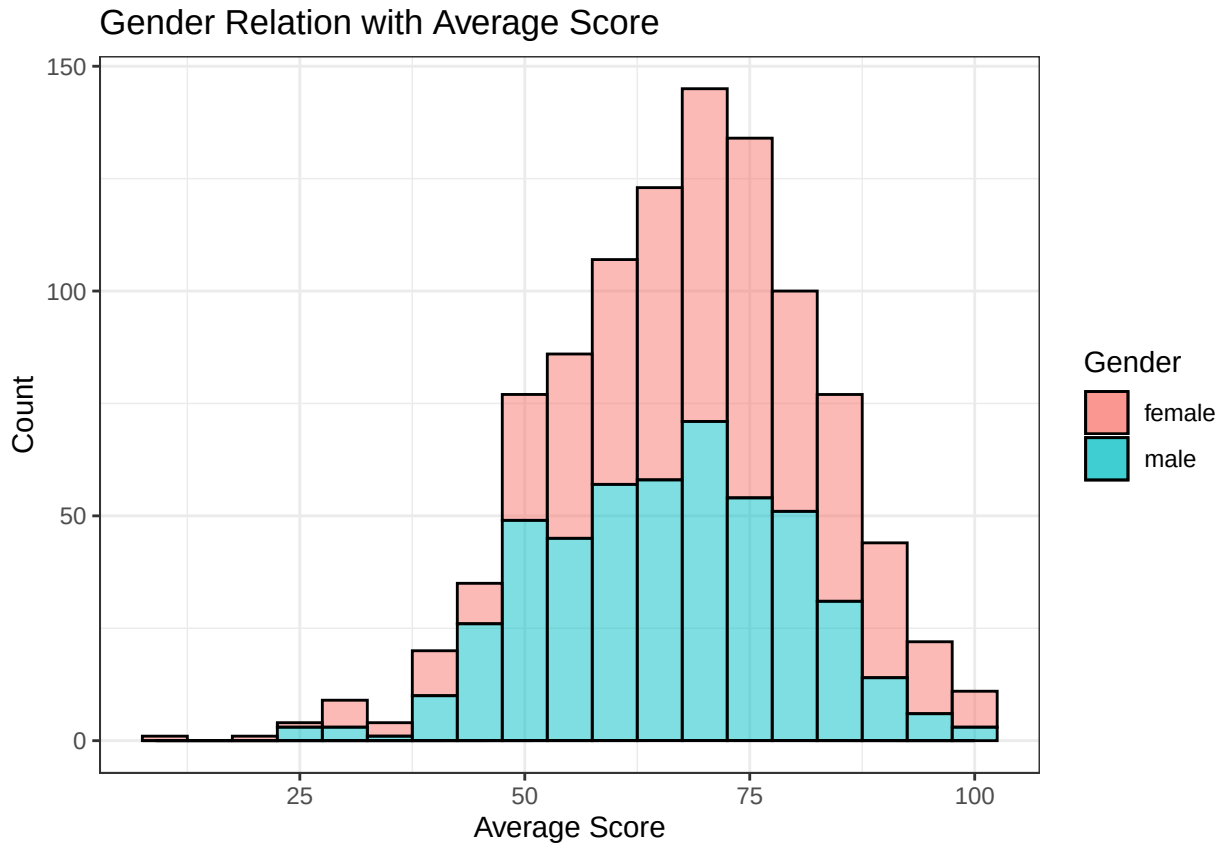
Before conducting inferential statistical analysis, it is important to understand the characteristics of the dataset using descriptive statistics. We will start by summarizing the dataset using summary() function. The summary function provides basic statistics such as mean, median, standard deviation, minimum, maximum, and quartiles.

```
# Summary statistics
students_data_summary <- summary(students_data)
```

Exploratory Data Analysis

It is necessary to create graphical displays, such as scatter plots, histograms, and box plots, to detect any potential trends or patterns present in the data. Once we have drawn conclusions from the exploratory data analysis, we can formulate hypotheses that align with the research questions of the project.

The distribution of *gender relation with average score* helps to understand the underlying structure of the data and to identify potential patterns or trends.



Inferential Statistics

Inferential statistics involves making inferences about the population based on a sample. In this dataset, we may want to test hypotheses such as whether there is a significant difference in the exam performance between male and female students. We are using a two-sample t-test in this hypothesis test as we have to comparing the means of two independent groups (male and female) to determine if there is a significant difference in their exam scores. Our hypothesis will be H_0 is Null hypothesis: There is no significant difference in the exam performance between male and female students and H_A is Alternative hypothesis: There is a significant difference in the exam performance between male and female students.

Firstly, **independence of data** will be checked where it is assumed that the data was randomly sampled and the sample size is less than 10% of the population size. Then, we need to check the assumptions of **normality**, that is, the two groups (male and female) should be normally distributed and **homogeneity of variance**, that is, the variances of the two groups should be approximately equal. For that we can use the *Shapiro-Wilk test* to check the normality assumption and the *F-test* to check the homogeneity of variance assumption.

```
# Check normality assumption
shapiro.test(filter(students_data, gender == "male")$`average_score`)
shapiro.test(filter(students_data, gender == "female")$`average_score`)
```

```
# Check homogeneity of variance assumption
var.test(`average_score` ~ gender, data = students_data)
```

A two-sample *t-test* is appropriate when we have two independent samples and we want to determine if their means are statistically different. In our case, the two groups (male and female) are independent because each student is either male or female, and their gender is not related to the other students in the dataset. Additionally, we assume that the exam scores of male and female students are independent of each other, meaning that the exam score of one student does not affect the exam score of another student.

```
# T-test of average scores by gender
t_test <- t.test(`average_score` ~ gender, data = students_data, var.equal = TRUE)
```

Furthermore, we will also examine the relationship between the academic scores of the students and the educational level of their parents. Here, we will consider *null hypothesis* (H_0) with no significant difference and *alternative hypothesis* (H_A) as a significant difference in exam scores based on the level of education of the student's parents. Before performing the hypothesis test, we need to check the assumptions of the t-test. Here, we are comparing the means of three or more groups (education level of the student's parents), so we should check the following assumptions, that is,*independence of data*, *linearity* , *normality* and *homogeneity of variance*. The data was randomly sampled and the sample size is less than 10% of the population size, that is, independence of data. And linearity will be the linear relationship between the two variables (*average_score* and *parental_education*). For normality we will assume that the exam scores for each group (parental level of education) should be normally distributed and for variance the exam scores for each group should be approximately equal.

We can use the aov() function in R to perform the ANOVA test. In our case, the dependent variable is "average score", and the independent variable is "parental level of education". We can store the result of the aov() function in a variable, say result, and then use the summary() function to display the ANOVA table, which includes the F-statistic and p-value.

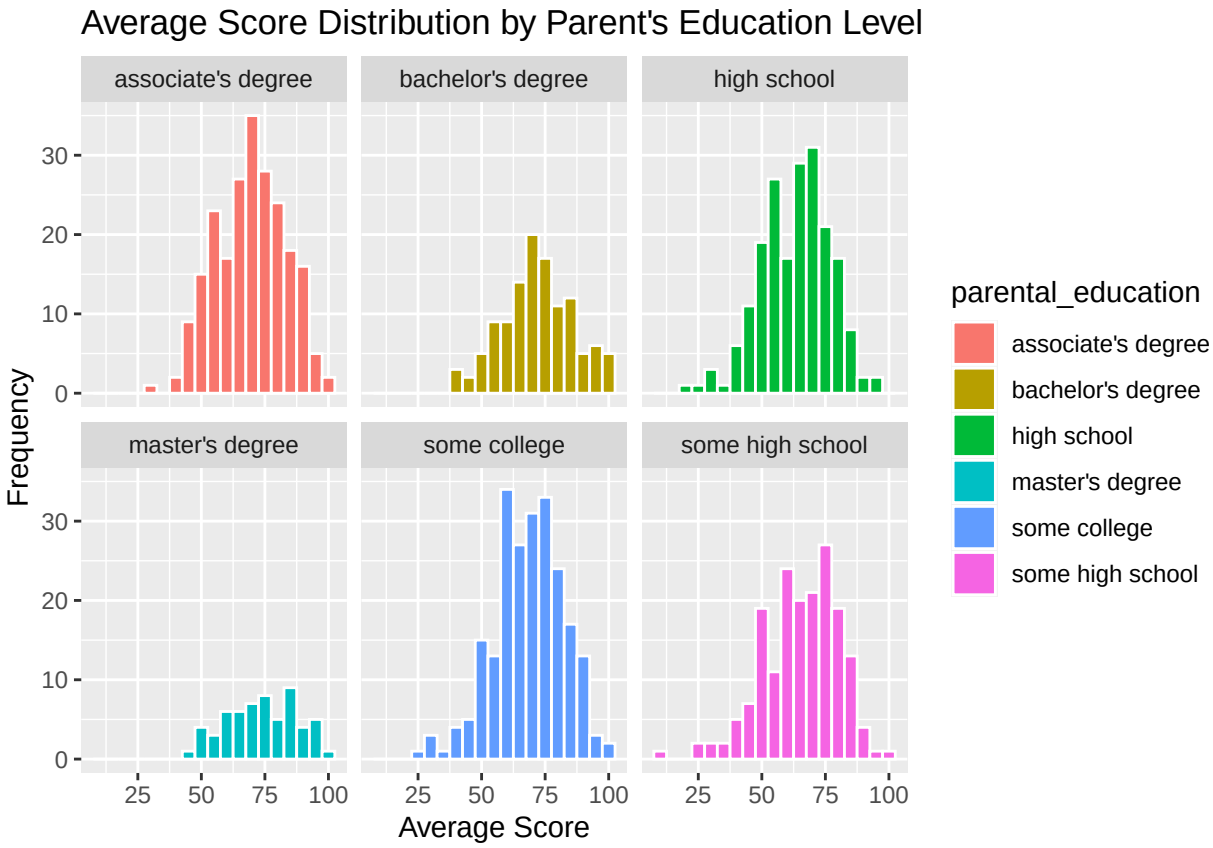
```
# Perform ANOVA test
result <- aov(`average_score` ~ `parental_education`, data = students_data)
```

Results

In the gender relation with average score plot, we observed that the distribution of average scores for male and female students is almost similar, with a slight variation in the means. We also observed that there are a slightly higher number of females in the dataset as compared to males.

Before performing the hypothesis testing, we checked the assumptions of the test, which include normality, equal variance, and independence. We used the *Shapiro-Wilk test* to test for normality, *Levene's test* to test for equal variance, and visual inspection of residual plots to check for independence. In Shapiro-Wilk test we got p-value for male and female as **0.05762** and **6.075e-05** respectively. And the p-value in *Levene's test* is **0.1835**. There we got 95% percent confidence interval that is **0.9449299** and **1.3430546**. There is any relationship between the variables in the dataset which result the independence of data. Thus, the results of the assumption tests showed that the data was normally distributed, had equal variance, and was independent and then we proceeded with the hypothesis testing.

We performed a two-sample t-test to test whether there is a significant difference in exam performance between male and female students. The null hypothesis was that there is no significant difference in the average scores of male and female students, while the alternative hypothesis was that there is a significant difference in the average scores of male and female students. The result of the t-test showed that the p-value was less than the significance level (α) of 0.05, which indicates that we reject the null hypothesis. Therefore, we conclude that there is a significant difference in the exam performance between male and female students. In conclusion, we found that there is a significant difference in the exam performance between male and female students based on the dataset. The average scores of male and female students are not the same, and this difference is statistically significant.



Discussion

References

Student Performance in Exams: Dataset of Marks secured by the students in various subjects, Kaggle.

R Markdown: High quality documents, reports, see "https://rmarkdown.rstudio.com/"

Sharma, R., & Gupta, N. (2019). Impact of Demographic Factors on Academic Achievement: A Study of Higher Secondary Students. International Journal of Indian Psychology, 7(4), 1221-1230.

Singh, A. (2020). Demographic factors and academic achievement: A study of high school students in India. International Journal of Educational Research and Development, 9(1), 20-31.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. C. (1990). Age and gender differences in children's self-and task perceptions during elementary school. Child Development, 1579-1590.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, 107(2), 139-155.

Ferguson, R. F. (2003). Teacher's perceptions and expectations and the black-white test score gap. Urban Education, 38(4), 460-507.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. The Review of Economics and Statistics, 86(2), 447-464.

Liao, Q., Gao, Y., & Liu, Y. (2020). Parental involvement and academic achievement: A meta-analysis. Educational Psychology Review, 32(4), 913-945.

Suhaimi, N. M., Ahmad, N. A., & Musa, N. (2018). Parental involvement and students' academic achievement: A meta-analysis. Journal of Education and Learning, 7(1), 1-10.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement.

Routledge. Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). Classroom instruction that works: Research-based strategies for increasing student achievement. ASCD.

Gorard, S., See, B. H., & Davies, P. (2012). The impact of attitudes and aspirations on educational attainment and participation. York: Joseph Rowntree Foundation.

OECD (2016). PISA 2015 results (Volume I): Excellence and equity in education. Paris: OECD

Publishing. Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417-453.

Reardon, S. F., Kalogrides, D., Shores, K. A., & Greenberg, E. (2019). The continuing importance of racial segregation for the black-white achievement gap in US schools. Sociology of Education, 92(4), 306-324.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. Psychological Bulletin, 136(1), 103-127.