

ADSC1910 Case 1

Neha Malhan(T00757554), Thiwanka Abeyweera(T00754871)

2024-09-18

Case Study 1: Analyzing the Academic Performance of Grade 5 Students at Kamloops Elementary School

Introduction

An elementary school in Kamloops recently conducted a random survey of its Grade 5 students to understand their academic performance and preferences better. The data collected includes the students scores in Math, Science, and English, along with information on their favorite color and class. This case study will thoroughly analyze the quality and structure of the data, using descriptive statistical tools to evaluate performance trends, and a report on the overall data quality will be provided.

The following table summarizes the data collected:

Table 1: Collected Data from Grade 5 Students

Name	Math_Score	Science_Score	English_Score	Favourite_Colour	Favourite_Class
Amina	88	78	75	Red	Math
Hiroshi	77	75	64	Green	Science
Leila	55	68	47	Blue	Gym
Mateo	63	82	85	No Answer	Gym
Zainab	95	85	98	Blue	English
Rajesh	68	71	55	Red	Science
Anaya	75	88	79	Blue	Science
Omar	71	62	65	No Answer	Math
Xiang	86	76	55	Green	Math
Sofia	59	83	58	Green	Science
Alejandro	98	79	89	Orange	English
Nia	23	68	35	Blue	Gym
Amir	85	72	67	No Answer	Science
Yara	91	82	95	No Answer	English
Kofi	87	77	82	Red	Math

Methods

- Descriptive statistical measures, such as the mean, median, and standard deviation, are employed to assess the performance across various subject scores.
- Comparative boxplots are utilized to visually represent the distribution of scores across different subjects.
- Categorical data analysis is conducted to evaluate student preferences regarding classes and colors.

Findings/Results

Descriptive Analysis

To gain a clearer understanding of student performance across subjects, we calculated the mean, median, and standard deviation for Math, Science, and English scores. These statistics will assist in summarizing the data's central tendency and variability.

The following table shows the summary of the calculations:

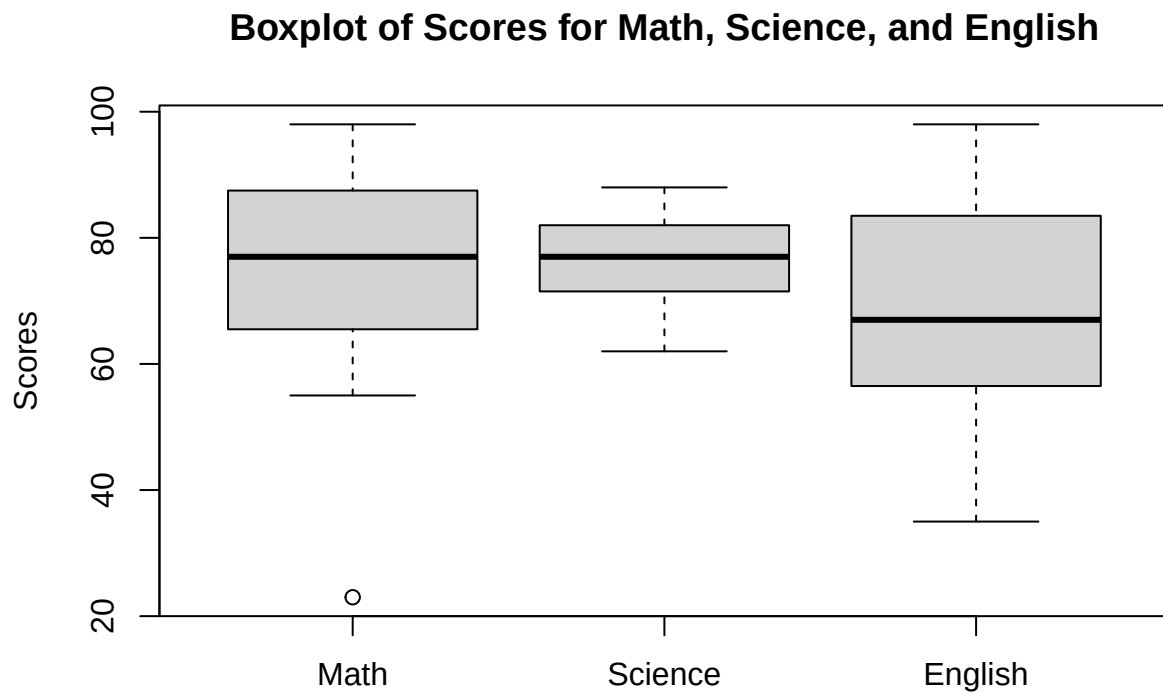
Table 2: Table for the statistical summary of the subject scores

	Mean	Median	Standard_Deviation
Math_Score	74.73333	77	19.440263
Science_Score	76.40000	77	7.189278
English_Score	69.93333	67	18.250114

In the analysis of the scores across subjects, Science has the highest mean score (76.40), followed closely by Math (74.73), while English shows the lowest mean score (69.93). Both Science and Math share the same median score of 77, whereas English has a lower median of 67. In terms of score variability, Science reports the lowest standard deviation (7.19). Math and English exhibit higher variability with standard deviations of 19.44 and 18.25.

Boxplot Analysis

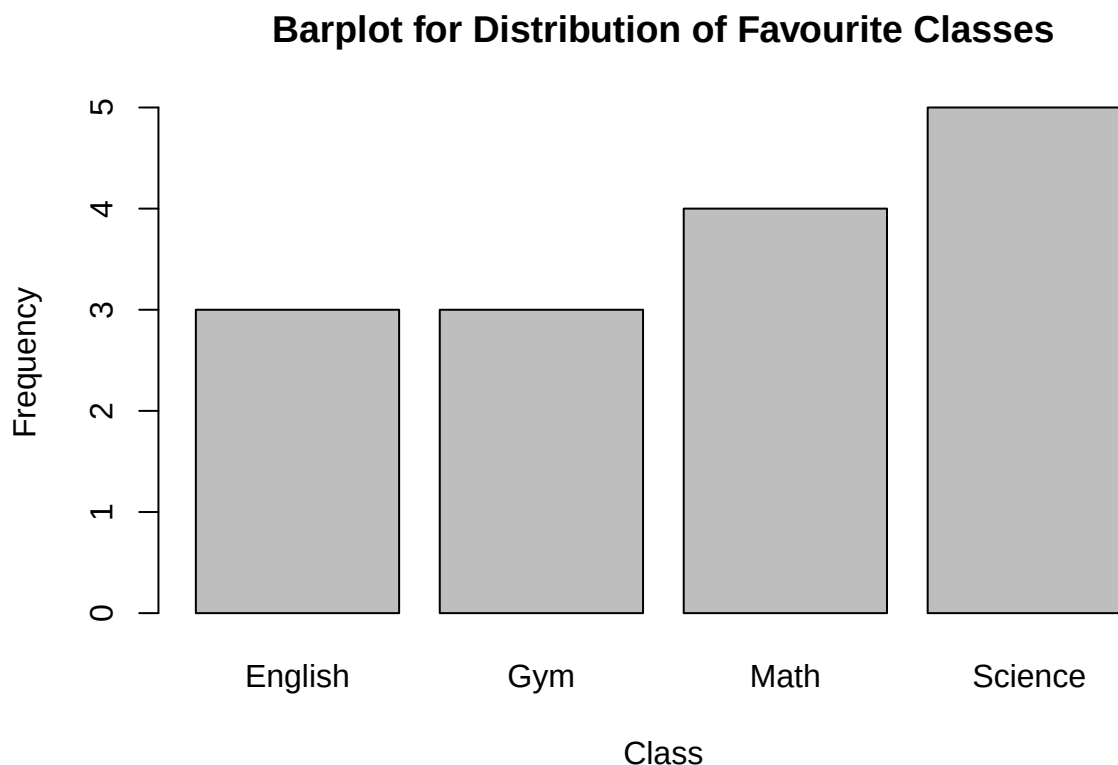
Boxplots provide a visual representation of the distribution of scores, showing the range, quartiles, and potential outliers. They assist in identifying the spread and central tendency of the data, making it easier to spot extreme values that may require further analysis.



The boxplot reveals the presence of an outlier in the Math scores, indicating a value that deviates significantly from the rest of the data distribution.

Categorical Data Analysis: Favourite Class

Barplot is used to examine the student preferences in the subject classes. Below is the visualization of the classes preferences.



The barplot indicates that Science has the highest frequency, suggesting it is the most popular class among students, followed by Math. In contrast, both English and Gym have the lowest frequencies, indicating that these subjects are equally the least preferred classes.

Discussion & Conclusions

Conclusion of Descriptive statistical and Boxplot analysis

- *Math:* The mean of Math score 74.7 and a median of 77 suggest that most students perform relatively well. However, the standard deviation of 19.4 indicates significant variability, with some students excelling while others face challenges. Nia's score of 23 stands out as a clear outlier, highlighting the need for additional support for lower-performing students. In contrast, students like Alejandro and Zainab are excelling. When applying a 10% trimmed mean to mitigate the impact of outliers, the adjusted mean increases to 76.92, closely aligning with the median of 77. This indicates that extreme values, such as Nia's score, skew the overall mean, leading to a misrepresentation of the group's overall performance, where the majority of students have performed well in Math.
- *Science:* The Science scores are closely clustered around the mean of 76.4, with a median of 77 and a low standard deviation of 7.18, indicating consistent performance among students. The smaller interquartile range and lack of extreme values suggest that most students are performing well. This uniformity implies that the Science curriculum is being effectively understood by the majority, with less disparity in performance compared to subjects like Math and English.

- *English:* The mean English score of 69.9 and the median score of 67 are notably lower compared to the other two subjects. The high standard deviation of 18.25 reflects considerable variability in student performance, with students like Zainab and Yara achieving very high marks, while others, such as Leila and Nia, are struggling. This variability highlights the differing levels of proficiency among students in English, indicating that while some students excel, others may require additional support.

Conclusion of Categorical Data Analysis: Favorite Class and Color

Comparing students' class preferences with their performance can reveal insights into the relationship between subject preferences and academic success. Based on the bar plot, Science emerged as the most preferred class among students, showing the highest frequency. This preference aligns with the subject's highest mean score of 76.4, suggesting that students performed well in the subject they liked the most. This positive performance could be attributed to factors such as their fondness for the teacher, engaging content in the science curriculum, or effective teaching methods. A similar pattern can be seen in Math and English. Math recorded the second-highest mean score, corresponding to the second-highest student preference. In contrast, English had the lowest mean score and was the least preferred subject among students. This trend indicates that student preference for a subject might be linked to their performance in it.

On the other hand, the other categorical variable in the dataset which is the "Favorite Color" of students, reflects their personal preferences, but lacks quantitative value and hierarchy, making it unsuitable for statistical analysis related to academic performance. Although it can offer some descriptive insights into students' preferences, it does not contribute directly to understanding their academic abilities. Therefore, the data of favorite class were not analysed in this report.

Recommendations

- Given the high variability in Math and English scores, it is recommended that the school implement additional support programs for struggling students. Personalized tutoring and differentiated instruction can help address performance gaps, while advanced materials can challenge high-achieving students. Additionally, curriculum adjustments may be necessary to better align with students' diverse learning needs and preferences.
- The consistency in Science scores suggests that current teaching methods are effective. To maintain this high level of performance, the school should continue utilizing these successful strategies. It is also advisable to apply similar effective teaching methods from Science to Math and English to enhance performance and engagement in these subjects.
- Students like Nia, who consistently perform lower across subjects, would benefit from individualized learning plans and additional support. Identifying and addressing the root causes of their difficulties is essential for improving their academic outcomes.
- Data such as "Favorite Color," which lacks quantitative value and does not correlate with academic performance, should be excluded from future analyses. Additionally, increasing the sample size in future data collection efforts will enhance the accuracy and reliability of conclusions, providing more representative and valid insights into the Grade 5 student population.

References

- Devore, J. L., Berk, K. N., & Carlton, M. A. (2012). Modern mathematical statistics with applications (Second Edition). New York: Springer.
- Crawley, M. J. (2007). The R Book (1st ed., p. 131). John Wiley & Sons.
- Teetor, P. (2011). R cookbook (1st ed., p. 172). O'Reilly Media.
- "cat" funtion

- Case Study Writing Information

Appendices

We, Neha Malhan(T00757554) and Thiwanka Abeyweera(T00754871), confirm that both members worked together equally on all parts of the case study. Every section was researched, written, and reviewed collaboratively, ensuring that both individuals contributed fully and shared responsibility for the entire project.

Code for analysis performed in this study

```
knitr::opts_chunk$set(echo = TRUE)
# Load knitr package

library(knitr)

# create a data set

data <- data.frame(
  Name = c('Amina', 'Hiroshi', 'Leila', 'Mateo', 'Zainab', 'Rajesh', 'Anaya', 'Omar',
           'Xiang', 'Sofia', 'Alejandro', 'Nia', 'Amir', 'Yara', 'Kofi'),
  Math_Score = c(88, 77, 55, 63, 95, 68, 75, 71, 86, 59, 98, 23, 85, 91, 87),
  Science_Score = c(78, 75, 68, 82, 85, 71, 88, 62, 76, 83, 79, 68, 72, 82, 77),
  English_Score = c(75, 64, 47, 85, 98, 55, 79, 65, 55, 58, 89, 35, 67, 95, 82),
  Favourite_Colour = c('Red', 'Green', 'Blue', 'No Answer', 'Blue', 'Red', 'Blue',
                       'No Answer', 'Green', 'Green', 'Orange', 'Blue', 'No Answer',
                       'No Answer', 'Red'),
  Favourite_Class = c('Math', 'Science', 'Gym', 'Gym', 'English', 'Science', 'Science',
                      'Math', 'Math', 'Science', 'English', 'Gym', 'Science', 'English',
                      'Math')
)

# create a table using kable()

knitr::kable(data, caption = "Collected Data from Grade 5 Students")
# Create a data frame with the scores
data <- data.frame(
  Name = c('Amina', 'Hiroshi', 'Leila', 'Mateo', 'Zainab', 'Rajesh', 'Anaya', 'Omar',
           'Xiang', 'Sofia', 'Alejandro', 'Nia', 'Amir', 'Yara', 'Kofi'),
  Math_Score = c(88, 77, 55, 63, 95, 68, 75, 71, 86, 59, 98, 23, 85, 91, 87),
  Science_Score = c(78, 75, 68, 82, 85, 71, 88, 62, 76, 83, 79, 68, 72, 82, 77),
  English_Score = c(75, 64, 47, 85, 98, 55, 79, 65, 55, 58, 89, 35, 67, 95, 82)
)

# Calculate mean, median, and standard deviation for each subject

# colMeans() calculates mean only for columns 2 to 4

mean_values <- colMeans(data[,2:4])

# apply() calculates median for col 2 to 4

median_values <- apply(data[,2:4], 2, median)

# standard deviation for col 2 to 4 , "2" stands for col if "1" it would be for row
sd_values <- apply(data[,2:4], 2, sd)
```

```

# Display summary
stat_summary <- data.frame(Mean = mean_values, Median = median_values,
                           Standard_Deviation = sd_values)
knitr::kable(stat_summary, cap="Table for the statistical summary of the subject scores")

# Boxplot for subject scores
boxplot(data$Math_Score, data$Science_Score, data$English_Score,
        names = c("Math", "Science", "English"),
        main = "Boxplot of Scores for Math, Science, and English",
        ylab = "Scores"
)

# Barplot for favourite class
data <- data.frame(
  Favourite_Class = c('Math', 'Science', 'Gym', 'Gym', 'English', 'Science', 'Science',
                     'Math', 'Math', 'Science', 'English', 'Gym', 'Science', 'English',
                     'Math')
)

df <- table(data$Favourite_Class)
barplot(df,
        xlab = "Class",
        ylab = "Frequency",
        main = "Barplot for Distribution of Favourite Classes"
)

# Calculate the trimmed mean for the math score (trim 10% from each end)

#trimmed_mean <- mean(data$Math_Score, trim = 0.1)
#print(trimmed_mean)

```