EasyChair Preprint

№ 15946

Analysis of Student Scores in Math, Reading, and Writing: Patterns and Predictive Relationships

Khaled M.M. Alrantisi and Gulnaz Gimaletdinova

March 27, 2025

Analysis of Student Scores in Math, Reading, and Writing: Patterns and Predictive Relationships

Authors:

Khaled M.M. Alrantisi *Student, Department of Computer Science and Engineering, Ala-Too International University* *Email: khaled.alrantisi@alatoo.edu.kg*	**Gulnaz Gimaletdinova** *Teacher, Department of Computer Science and Engineering, Ala-Too International University* *Email: gulnaz.gimaletdinova@alatoo.edu.kg*

Abstract

This research investigates the relationship between students' scores in three academic subjects: math, reading, and writing. The study aims to identify patterns in student performance across these subjects and explore the predictive nature of math scores for reading scores. We used descriptive statistics to summarize the data, histograms to illustrate score distributions, and box plots to compare performance across genders. A correlation matrix was constructed to examine the relationships between subjects, and linear regression was applied to predict reading scores based on math scores. The Shapiro-Wilk test was employed to assess the normality of math scores. Results suggest a strong correlation between math and reading scores, with math scores explaining 66.7% of the variation in reading scores. These findings provide valuable insights into the interconnectedness of student performance in these subjects.

Keywords

- Student Scores
- Math
- Reading
- Writing
- Descriptive Statistics
- Correlation Analysis
- Linear Regression
- Educational Data Analysis

Introduction

Understanding the relationships between students' performance in various subjects is crucial for improving educational outcomes. This study explores how student scores in math, reading, and writing are interrelated and how these relationships can inform teaching practices. By analyzing the scores, we aim to uncover patterns in performance and gain insights into how well students perform across subjects. This study also investigates if math scores can be used to predict reading performance, which could aid educators in identifying students who may need additional support in specific areas.

Methods

1. **Data Collection**:
   Data on student scores in math, reading, and writing were collected from a set of students. The scores were recorded on a scale from 0 to 100 for each subject.
2. **Descriptive Statistics**:
   The mean, standard deviation, and range were calculated for the scores in math, reading, and writing to summarize the data.
3. **Histograms**:
   Histograms were created to visualize the distribution of scores for each subject.
4. **Box Plots**:
   Box plots were used to compare the score distributions across genders (boys vs. girls) for each subject.
5. **Correlation Matrix**:
   A correlation matrix was constructed to examine the strength of relationships between math, reading, and writing scores.
6. **Linear Regression**:
   A linear regression model was applied to predict reading scores based on math scores. The equation for the regression line was derived, and the R-squared value was calculated to assess the fit of the model.
7. **Shapiro-Wilk Test**:
   The Shapiro-Wilk test was conducted to determine if the math scores follow a normal distribution.

Results

Descriptive Statistics

- We calculated the average (mean), spread (standard deviation), and the highest and lowest scores (range) for math, reading, and writing.
  Math: The average score is 66.1, with scores ranging from 0 to 100.
  Reading: The average score is 69.2, with scores ranging from 17 to 100.
  Writing: The average score is 68.1, with scores ranging from 10 to 100.
  These numbers show that students generally have similar scores across all three subjects.
-

```
# Descriptive statistics code
data.describe()
```

Histograms

Histograms show how the scores are spread out.
Math scores: Most students scored lower, with fewer students getting high scores.
Reading and Writing scores: These scores are more evenly spread across different ranges.

```
# Histograms code
data[['math_score', 'reading_score', 'writing_score']].hist(bins=15, figsize=(10, 8))
plt.suptitle('Score Distribution Histograms')
plt.show()
```

Box Plot

A box plot shows how scores are distributed by gender (boys vs. girls).
Boys: Boys generally scored higher in total scores, with their scores spread out more.
Girls: Girls have a more even spread of scores.
The box plot helps us understand how boys and girls perform in total scores.

```
# Box plot code
plt.figure(figsize=(10, 6))
sns.boxplot(x='gender', y='total_score', data=data)
plt.title('Box Plot of Total Score by Gender')
plt.show()
```

Correlation Matrix

A correlation matrix shows how scores in different subjects are related. For example:
Math and Reading: These two subjects are strongly connected. The higher a student's math score, the higher their reading score tends to be.
Reading and Writing: These scores are very closely related, too.
This helps us understand how well students do in one subject and predict their scores in another subject

```
# Correlation matrix code
correlation_matrix = data[['math_score', 'reading_score', 'writing_score']].corr()
print(f"\nCorrelation Matrix:\n{correlation_matrix}")

# Heatmap code
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix Heatmap')
plt.show()
```

Linear Regression

We used math scores to predict reading scores. The regression equation is:
Reading Score = 20.3523 + 0.7214 × Math Score
This equation tells us that math scores are a good way to predict reading scores. The higher the math score, the higher the reading score.
The results show that math scores can explain 66.7% of the variation in reading scores.

Linear regression code:

```
# Linear regression code
X = data['math_score']
y = data['reading_score']

# Add constant to the model
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
predictions = model.predict(X)

# Print out the regression summary
print(model.summary())

# Scatter plot with regression line
plt.figure(figsize=(10, 6))
sns.scatterplot(x='math_score', y='reading_score', data=data)
plt.plot(data['math_score'], predictions, color='red')
plt.title('Linear Regression: Math Score vs. Reading Score')
plt.show()
```

Shapiro-Wilk Test

The Shapiro-Wilk test checks if the math scores are normally distributed (a common pattern in many types of data). The p-value was 0.663, meaning the math scores are likely normally distributed.

Shapiro-Wilk test code:

```
# Shapiro-Wilk test code
stat, p = stats.shapiro(data['math_score'])
print(f'Shapiro-Wilk Test for Math Score: Statistics={stat}, p-value={p}')

# Check if data is normally distributed
if p > 0.05:
    print('Data is likely normally distributed.')
else:
    print('Data is likely not normally distributed.')
```

Discussion

The results highlight several key findings. First, there is a strong relationship between math and reading scores, as indicated by the high correlation coefficient and the linear regression model. This suggests that improving math skills could have a positive impact on reading performance. Furthermore, the Shapiro-Wilk test confirms that math scores are normally distributed, which is a useful property for many statistical analyses. The differences in score distributions between boys and girls suggest that gender may play a role in performance, which could be explored further in future research.

While the linear regression model provides a reliable way to predict reading scores from math scores, it explains only 66.7% of the variance. This leaves room for other factors that could influence reading performance, such as writing skills, socioeconomic status, or teaching methods. Future studies could incorporate additional variables to create a more comprehensive model.

Conclusion

This study helps us understand how student's scores in math, reading, and writing are connected. It shows that math scores can be used to predict reading scores. The results of the Shapiro-Wilk test suggest that the math scores follow a normal distribution. Understanding these relationships is useful for improving teaching methods and helping students perform better in all subjects.

References:

1. Shaiakhmetov, Dim, et al. "Criteria for Assessing the Quality of Educational Games and Their Impact on Student Learning Outcomes." Proceedings of the International Conference on Computer Systems and Technologies 2024. 2024.
2.
3. Shaiakhmetov, Dim, et al. "Morphological classification of galaxies using SpinalNet." 2021 16th International Conference on Electronics Computer and Computation (ICECCO). IEEE, 2021.
4.
5. Toktosunova, Akbiike, et al. "Developing an Artificial Intelligence Tool for Image Generation Using a Unique Dataset with Image-to-Image Functionality." Proceedings of the International Conference on Computer Systems and Technologies 2024. 2024.
6.
7. Sadriddin, Zuhra, Remudin Reshid Mekuria, and Mekia Shigute Gaso. "Machine Learning Models for Advanced Air Quality Prediction." Proceedings of the International Conference on Computer Systems and Technologies 2024. 2024.
8.
9. Gaso, Mekia Shigute, et al. "Utilizing Machine and Deep Learning Techniques for Predicting Re-admission Cases in Diabetes Patients." Proceedings of the International Conference on Computer Systems and Technologies 2024. 2024.
10.
11. Comparison and Selection of an Edge Detection Filters for ECG Applications
12. Dim Shaiakhmetov, Mohd. Tauheed Khan, Gulnaz Gimaletdinova
13. Alatoo Academic Studies 23 (1), 509 - 519