**Analysis of Factors Affecting Student Performance in Exams**

**Introduction** This report analyses a dataset containing various factors that influence student performance in exams. The dataset includes 6,607 observations and 20 variables, covering aspects such as study habits, attendance, parental involvement, access to resources, and extracurricular activities. The primary objective is to identify key factors that impact students' exam scores using descriptive statistics, visualization, and regression analysis.

The dataset was sourced from a publicly available educational performance study, ensuring reliability and comprehensiveness in assessing various academic success factors. The dataset satisfies the required condition of having at least 500 observations and more than five different variables, making it a suitable candidate for this analysis.

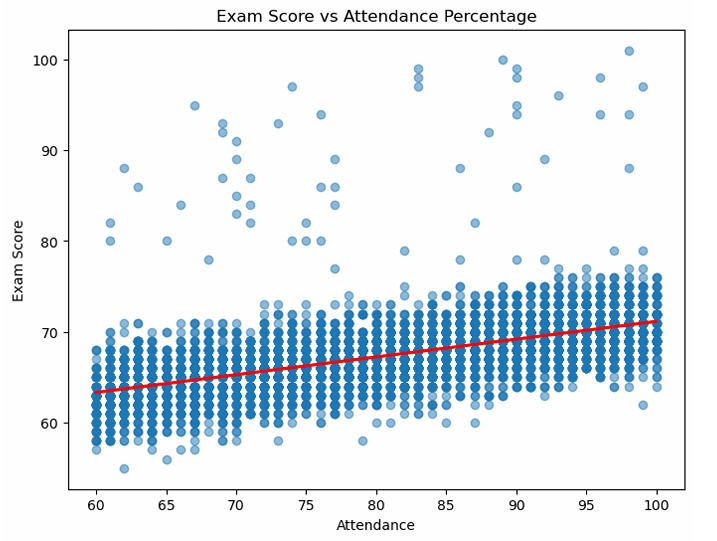
**Descriptive Statistics** The table below provides key summary statistics for some of the most relevant numerical variables in the dataset:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Count** | **Mean** | **Std Dev** | **Min** | **25%** | **50%** | **75%** | **Max** |
| Hours Studied | 6607 | 19.98 | 5.99 | 1 | 16 | 20 | 24 | 44 |
| Attendance (%) | 6607 | 79.98 | 11.55 | 60 | 70 | 80 | 90 | 100 |
| Sleep Hours | 6607 | 7.03 | 1.47 | 4 | 6 | 7 | 8 | 10 |
| Previous Scores | 6607 | 75.07 | 14.40 | 50 | 63 | 75 | 88 | 100 |
| Tutoring Sessions | 6607 | 1.49 | 1.23 | 0 | 1 | 1 | 2 | 8 |
| Physical Activity | 6607 | 2.97 | 1.03 | 0 | 2 | 3 | 4 | 6 |
| Exam Score | 6607 | 67.23 | 3.89 | 55 | 65 | 67 | 69 | 101 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |

From these statistics, we observe that the average number of study hours is around 20, and attendance averages nearly 80%. The average exam score is approximately 67, with a standard deviation of 3.89. Furthermore, tutoring sessions, physical activity, and sleep hours also show variation, potentially contributing to academic performance.

**Visualization** A scatter plot was generated to analyze the relationship between attendance and exam scores. The plot suggests a positive correlation, indicating that higher attendance is associated with better performance.

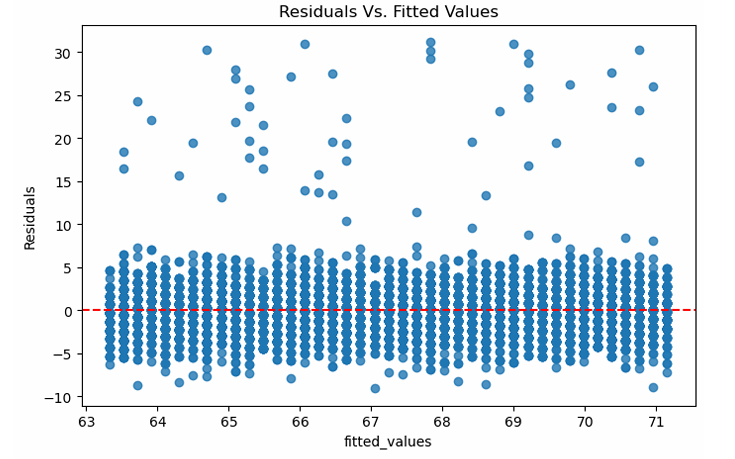


The correlation coefficient between attendance and exam scores is **0.581**, indicating a moderately strong positive relationship.

**Regression Analysis** A simple linear regression model was implemented using attendance as the independent variable and exam score as the dependent variable. The regression results are as follows:

* **Equation:** Exam Score = **51.58** + **0.1958 × Attendance**
* **R-squared:** 0.338 (34% of the variance in exam scores is explained by attendance)
* **P-value:** 0.000 (Highly significant relationship)
* **Breusch-Pagan Test for Heteroscedasticity:** p-value = 0.805 (No significant heteroscedasticity detected)

The residuals versus fitted values plot confirms that the model assumptions hold, as the residuals are randomly scattered.



To enhance the analysis, a multiple regression model incorporating variables such as hours studied, parental involvement, and access to resources could provide deeper insights. This would help in understanding how multiple factors collectively influence student performance.

**Conclusion:**

This analysis demonstrates that attendance is a significant predictor of student performance, explaining about 34% of the variance in exam scores. While attendance positively impacts exam scores, other factors such as study hours, tutoring, and motivation may also play an important role.

Future analyses could involve a multiple regression model incorporating these additional factors for a more comprehensive understanding of student performance determinants. Additionally, further exploration using classification models like logistic regression may help in categorizing students into performance brackets, enabling targeted interventions for academic improvement.