

Analyzing the Relationship Between Study Time and Academic Performance

Rajesh Upadhayaya, Manjil Pradhan, Thomas Hynes and Jonathan LLoyd

May 2025

1 Introduction

In this study we investigate the relationship between study time and their academic performance. The performance of student can be influenced by factors such as study hours, family support, teacher factor and many more. We perform exploratory data analysis, model creation, hypothesis formulation, and testing.

2 Dataset

We selected student performance dataset from the University of California Irvine Machine Learning Repository. This dataset is about the student achievement in secondary education of two Portuguese schools. The dataset contains 395 students and includes data for numerous factors including, but not limited to: sex, age, familial size, study time, and academic grades. The extensive data collected for each student enables numerous research studies that investigate how various life style attributes may affect academic performance.

3 Research Question

Does weekly study time affect the student academic performance ?

4 Exploratory Data Analysis

4.1 Visualization

Each student in the dataset has three grades: G1,G2, and G3. We computed the average of these three grades for each student, and visualized the data as a function of weekly schedule time. The results are shown in Figure 1. The figure demonstrates that there is an increase in average grades with increasing study time, but we don't know if this result is statistically significant. We also visualized a scatter plot of the data, and plotted a linear regression, see Figure 2.

The visualization showed a positive correlation between study time and average grades, and motivates us to explore this research question further. In Figure 3, we plotted a heat map of the correlation of various student variables with average grades. The plot shows that study time has one of the highest positive correlation coefficients with average grades for our dataset, which provides further motive to explore this research question.

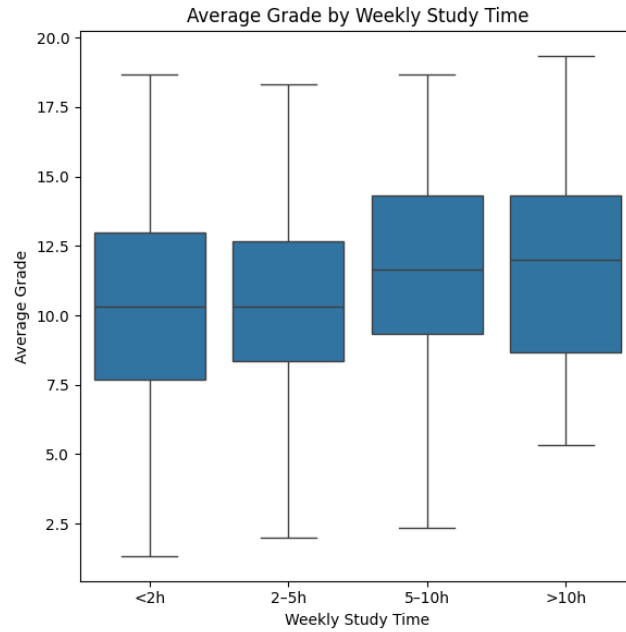


Figure 1: Box plot for average grade of student.

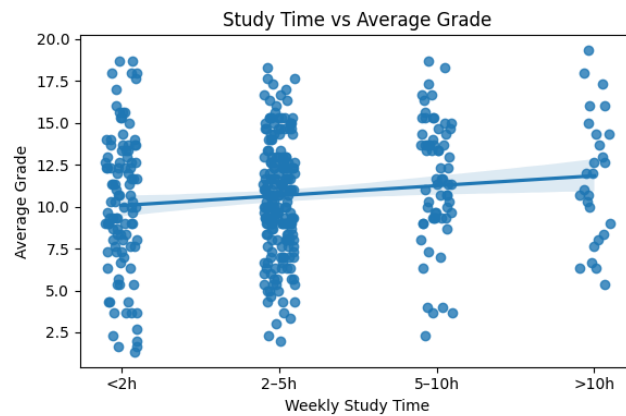


Figure 2: Scatter plot for average grade of student.

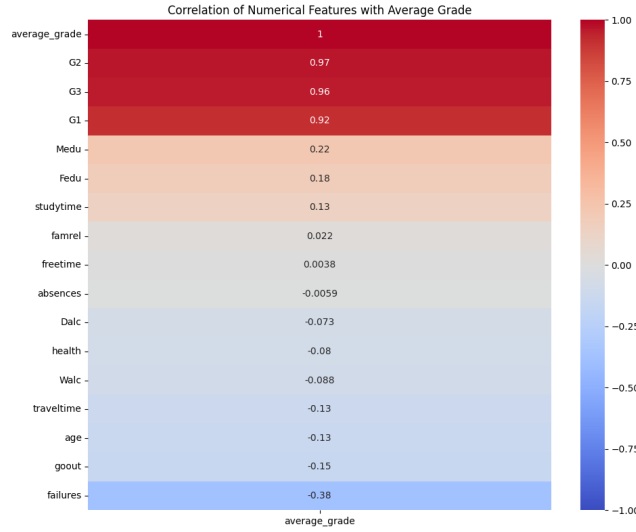


Figure 3: Correlation of Numerical Features with Average Grade

5 Hypotheses and Methods

Different method are applied to answer our research question:

- **Correlation:** Pearson and Spearman tests are conducted to examine the correlation between study time and average grade. The Pearson correlation calculates the strength of a direct linear correlation, while the Spearman correlation calculates the strength of a monotonic relationship between two variables. Both produce a correlation coefficient between -1 and 1, where -1 represents a perfect negative correlation, 1 represents a perfect positive correlation, and 0 indicates no correlation. Thus, our null hypothesis for each test is as follows:

$$H_0: \text{Correlation Coefficient} = 0$$

$$H_1: \text{Correlation Coefficient} \neq 0$$

- **Linear Regression:** This test captures the relationship between study time and average grades.

$$H_0: \text{Coefficient} = 0$$

$$H_1: \text{Coefficient} \neq 0$$

- **ANOVA:** Analysis of Variance test is used to quantify if the difference between average grades for the four categories of study time is statistically significant.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

- **Logistic Regression:** A logistic regression model was developed to predict if a student will pass or fail depending on their study time. The model considers a student to be passing if their average grade is greater than 10.

6 Results

The Pearson correlation coefficient between study time and average grades is 0.135, with a P-value of 0.0074. The Spearman coefficient is 0.125, with a P-value of 0.0132. The linear regression model shows a positive coefficient of 0.593 with a 95% confidence interval of 0.160 to 1.026, and a computed P-value of 0.007. The output from the linear regression is shown in Figure 4. The ANOVA test yielded a F-statistic of 3.001, and a P-value of 0.0305. The output from the logistic regression model is shown in Figure 5.

OLS Regression Results						
Dep. Variable:	average_grade	R-squared:	0.018			
Model:	OLS	Adj. R-squared:	0.016			
Method:	Least Squares	F-statistic:	7.248			
Date:	Sun, 11 May 2025	Prob (F-statistic):	0.00740			
Time:	11:56:52	Log-Likelihood:	-1072.8			
No. Observations:	395	AIC:	2150.			
Df Residuals:	393	BIC:	2158.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	9.4728	0.485	19.545	0.000	8.520	10.426
studytime	0.5927	0.220	2.692	0.007	0.160	1.026
Omnibus:	3.629	Durbin-Watson:	2.054			
Prob(Omnibus):	0.163	Jarque-Bera (JB):	3.037			
Skew:	-0.115	Prob(JB):	0.219			
Kurtosis:	2.638	Cond. No.	6.83			

Figure 4: Linear regression model

	precision	recall	f1-score	support
False	0.00	0.00	0.00	30
True	0.62	1.00	0.77	49
accuracy			0.62	79
macro avg	0.31	0.50	0.38	79
weighted avg	0.38	0.62	0.47	79

Figure 5: Logistic regression model

7 Discussion

The results of the Pearson and Spearman coefficient tests indicate that there is a statistically significant positive correlation between study time and average student grades, giving us confidence to reject the null hypothesis for both tests. Similarly, the results of the linear regression demonstrate that there is a statistically significant positive correlation between study time and average grades. However, the linear regression also yielded a very small R-squared value, indicating that the effect of study time on average grades may be quite small. The results of the ANOVA test indicate that the difference between average grades of the different study time groups is statistically significant. Applying the logistic regression model trained on the data of average grades vs. study time produced a model that has an overall accuracy of 62%. The model accurately predicts when a student will "pass" in every case, but did not correctly predict a single student "fail".

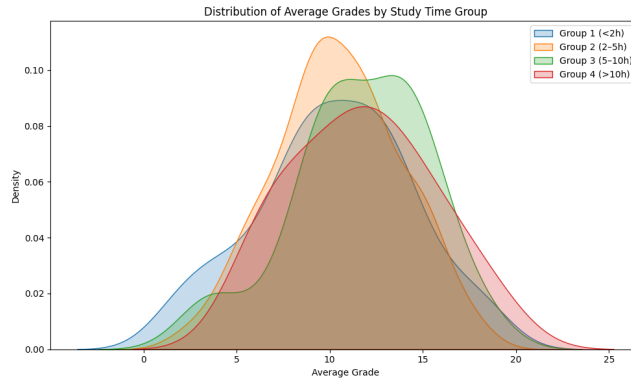


Figure 6: Distribution of average Grades

8 Conclusion

In this study, we conclude weekly study hours have a positive impact on their grades. Study time is associated with higher average grades. From our model,

we can infer that study time along with other factor plays a significant influence in student grades.