Study of the Association between Study Time and Final Grades

Yuyang Miao

Huan Yu Jiutian International Education Technology Co., Ltd, Beijing 100000, China

Abstract. Currently, many schools use test scores as a major criterion for measuring students' learning outcomes, and many students recognize increasing their personal performance as the main objective of education. Therefore, statistical learning method is utilized in this essay to examine the link between academic achievement and other variables. Using the student marks dataset, the link between students' grades and study time was investigated. To study if there is a direct correlation between study time and grades. With multiple regression, a proportionate relationship between study time and grades was determined. As predicted, study time had a substantial role in achieving great final marks. Additional investigation revealed that the impact of study time is greater than that of course load. Apparently, there is no evident association between the number of courses and the amount of time spent studying when viewed as independent variables for visual drawing. In conclusion, this paper suggest that if a student wants to increase his or her marks, increasing his or her study time is a beneficial option.

Keywords: Study Time; Final Grades; test scores.

1. Introduction

The final grades are affected by a variety of variables, such as the amount of stress brought on by family obligations, the number of classes taken, and the amount of time for studying. Allensworth and Luppescu's (2018) research was cited. William P. Young (1998) discovered via the use of the statistical tool known as Pearson's r Product Moment Correlation that there is a connection between grades and the amount of time spent studying. There is a strong correlation between the amount of time a student spends on studying for an examination and the quality of the examination results the student achieves. we obtain dataset from Kaggle. Then, we performed more research on Young's conclusions. Determining the degree to which final grades and the other two variables are correlated, then selecting the one with the highest connection as the focus of further study.

The results including how much time they spent on studying each day and their overall grade, were supplied via the official Kaggle website. With the use of this data, I investigated the multiple regression equation using a single variable. In the bottom part of the table, the data included a total of 100 sample numbers and 3 factors, however there were no group variables included. removing the data groups that were found to have missing values at the beginning, and then utilizing the Skimr skim function to build a skim description of the data. In the past, Skimr contained capabilities that combined the process of skimming with a secondary effect, such as the ability to rearrange the data, generate a list, or publish the results. These functions have since been removed. (Compact and Flexible Summaries of Data, not specified date) The data may be analyzed well by using the skmir skim function.

Table 1. The data after eliminating the missing value and a skim description of the data

Data summary	
Name	mark
Number of rows	100
Number of columns	3
Column type frequency:	
numeric	3
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
number_courses	0	1	5.29	1.80	3.00	4.00	5.00	7.00	8.00	
time_study	0	1	4.08	2.37	0.10	2.06	4.02	6.18	7.96	
Marks	0	1	24.42	14.33	5.61	12.63	20.06	36.68	55.30	

For the purpose of the visual sketching analysis, Mark served as the dependent variable, while the number of courses served as the independent variable. The x-values of the dots stand for the total number of classes taken, while the y-values indicate the grades earned in those classes. It is clear that there is only weak connection between the two variables based on the changing and nonlinear image of study time and course numbers.

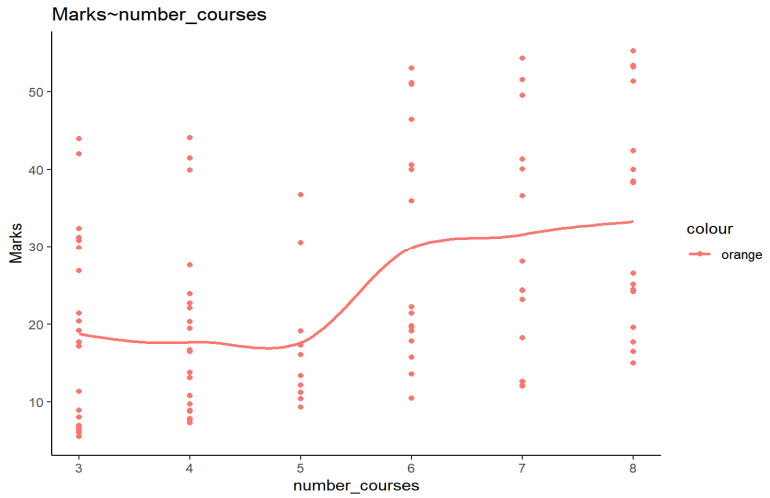

Fig. 1 The relationship between grades and the number of courses

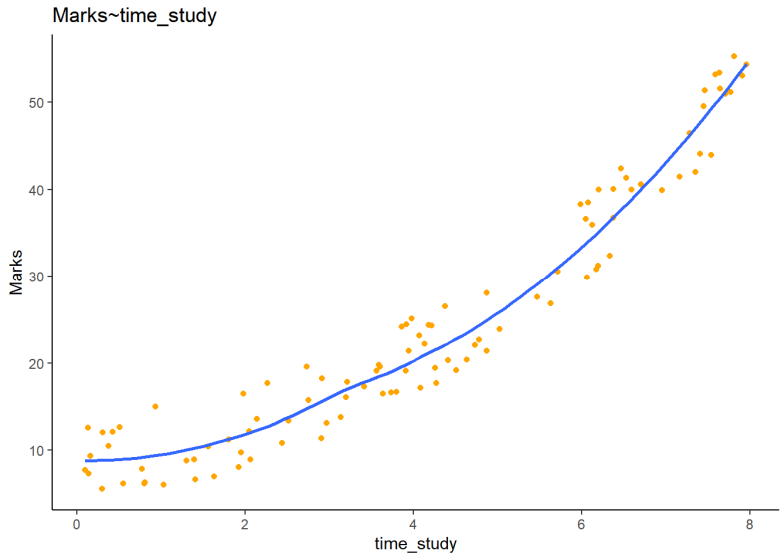

Fig. 2 The relationship between study time and marks

Then, using the grade as the dependent variable and the amount of studying time as the independent variable. By the visual drawing analysis, a graph demonstrates an increasing trend. It is evident that the point distribution on the picture of grades and time is more concentrated and demonstrates a direct proportional connection. The graph demonstrates a substantial relationship between study time and grades.

Based on the findings of the two scatter plots shown before, the association between study time and grades is more substantial. For the remainder of the analysis, I will focus mostly on the study time as the independent variable.

2. Modeling

First, the data is split into trainset and testset with :3 ratio. The Figure 2 shows that the distribution is almost exactly a straight line.

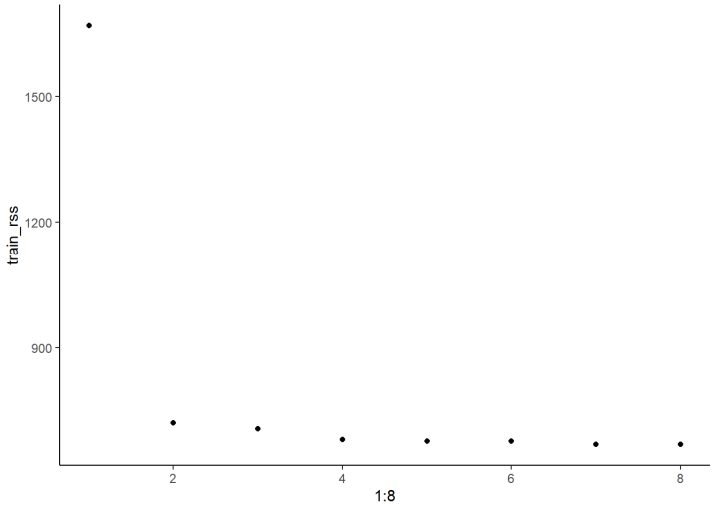

Fig. 3 The trainset and the testset with 7:3 ratio

The second step involves building a k = 1 to 8 training set model and sketching the RSS change curve. Iwhen K increases, RSS increases gradually at first and subsequently declines fast.

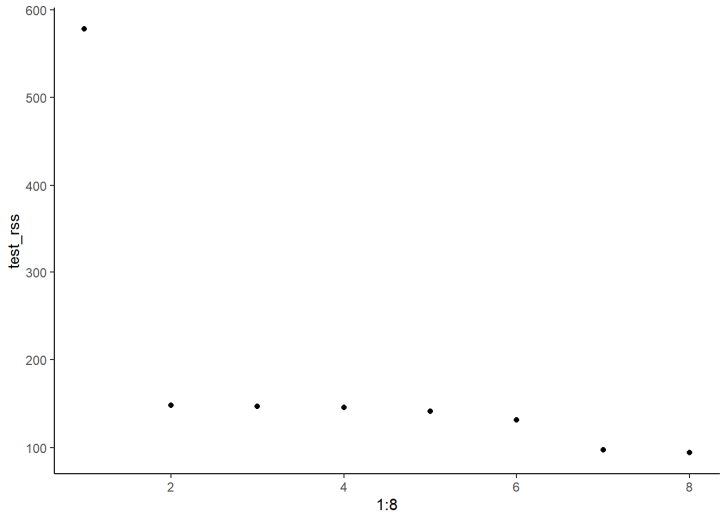

Fig. 4 The RSS change curve built by the training set model from 1 to 8

Comparing the aforementioned RSS values, the best model in the test set is LM8 TEST RSS, with the following residual standard deviation:

3. Plot Prediction

Forecasts for each model are shown below, with red dots indicating beginning values and black dots representing predictions.

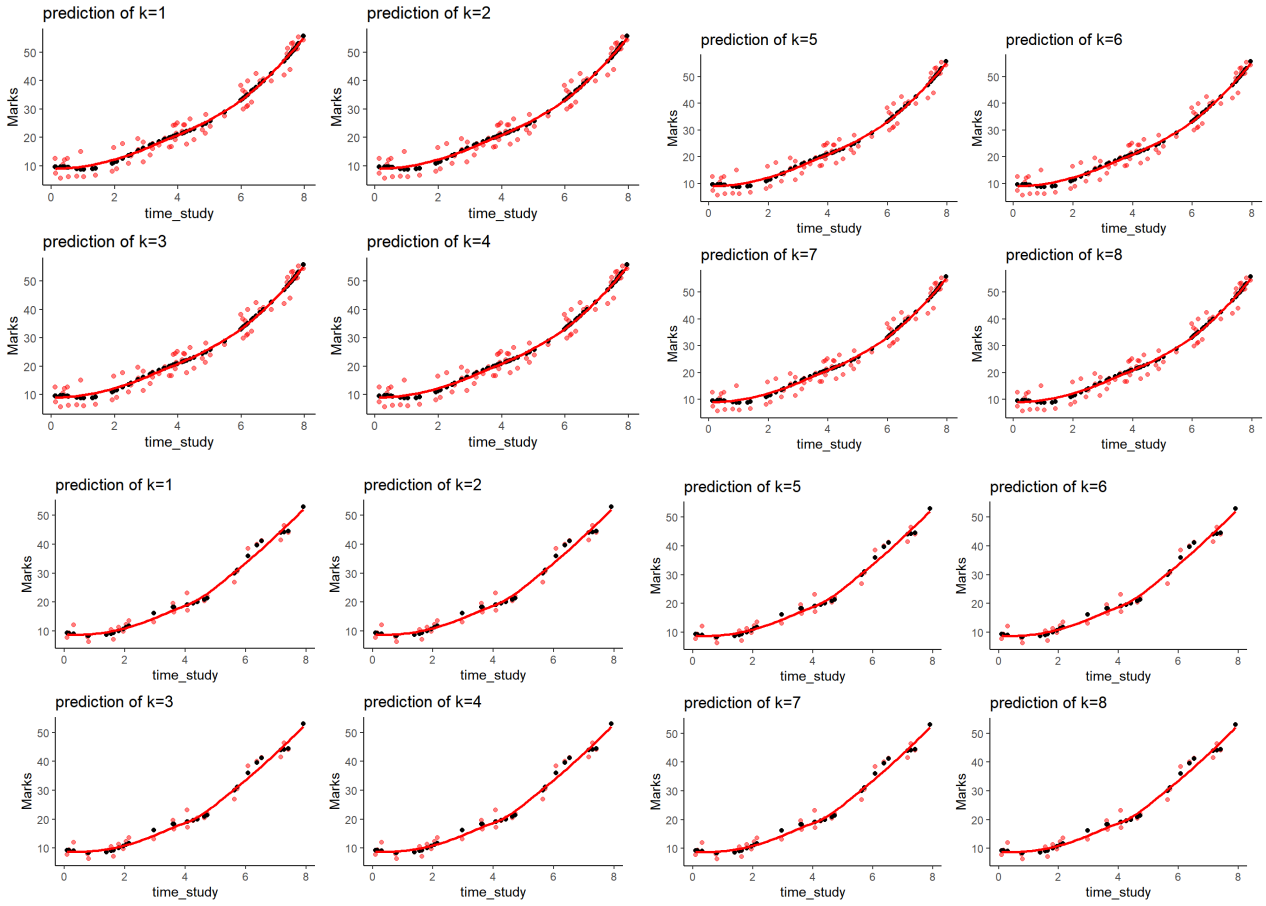

Fig. 5 The graph of the forecasts of each model

4. Conclusion

This article focuses mostly on analyzing the connection between the final grade, the total amount of studying time, and the total number of classes taken. In the first step of this process, the missing data points that were were filled in, then the multiple regression equation was used with only one variable. Creating a skim description of the data via the use of the skimr skim function. The second step is to make the cumulative grade point average the dependent variable, with the total number of classes serving as the independent variable. A graph was created using visual drawing analysis to demonstrate that there is only a tiny link between final grades and the number of courses taken by the student. Then we will assign the students' overall grade as the dependent variable and their amount of studying time as the independent variable. Repeating the process of using visual drawing analysis and discovering that the picture of the two variables is proportionate. As a matter of the fact, the connection between final grades and the amount of time spent studying is much greater than the connection between final grades and the total number of classes taken. As a result, the period of the research has been selected for more investigation. Third, separate the data into a trainset and a testset with a ratio of 7:3 between the two sets. The next step is to draft an RSS change curve for the trainset model, which ranges from k=1 to 8. The last step is to generate the predictions for each model. We found that the link between final grades and total number of courses had a poor relationship, however the correlation between final grades and total amount of time spent studying had a significant relationship. Therefore, increasing the amount of time spent studying is a viable strategy for enhancing academic achievement.

References

[1] Allensworth, E. M. A., & Luppescu, S. L. (2017). Why do students get good grades, or bad ones? The influence of the teacher, class, school, and student. University of Chicago Consortium on School Research. Retrieved September 21, 2022, from https://consortium.uchicago.edu/sites/default/files/2018-10/Why%20Do%20Students%20Get-Apr2018-Consortium.pdf.

[2] Compact and Flexible Summaries of Data. (n.d.). Retrieved September 21, 2022, from https://docs.ropensci.org/skimr/.

[3] Young, W. P. (1998). A Study to Determine the Correlation Between Extra Study Time After School to Grades Earned by Students. ODU Digital Commons. Retrieved September 21, 2022, from https://digitalcommons.odu.edu/ots_masters_projects/292/?utm_source=digitalcommons.odu.edu%2Fots_masters_projects%2F292&utm_medium=PDF&utm_campaign=PDFCoverPages.