Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/36/20240526

Study the influencing factors of passing or failing the exam

Yilin Tan

School of Statistics, Tianjin University of Finance and Economics, Tianjin, 300000, China

Allitan@stu.tjufe.edu.cn

Abstract. Previous studies have analyzed the factors affecting academic performance and concluded that the length of study is related to academic performance. However, there are other factors worth investigating. Data from 500 students were analyzed using the Binary Logistic Regression model. This paper studies the influence of Study Hours and Previous Exam Scores on pass or fail the exam. The results of the likelihood ratio test of the binary Logit regression model show the validity of the model and the significance of the construction of the model. In the Hosmer-Lemeshow fit test, the p-value is greater than 0.05, the model is considered to have good goodness of fit and passes the HL test. Through research, the author finally concludes that Study Hours and Previous exam Scores have a positive impact on passing the exam. Therefore, this paper argues that people can improve the likelihood of passing the exam by increasing their study time before the exam and getting better grades in the previous exam.

Keywords: Binary logistic regression model, study hours, previous exam scores, pass or fail the exam.

1. Introduction

Tests can assess students' academic performance, which in turn stimulates learning motivation, and ultimately urges students to progress. Ha and Lee concluded through research that exams had a positive impact on subsequent learning [1]. The ability to pass exams has always been an important part of a student's academic life. In many countries, the postgraduate examination is important in life, and the initial score of the postgraduate examination is the first threshold to determine whether a graduate can obtain a postgraduate degree [2]. The prediction of the variable score of the preliminary postgraduate examination can help candidates build up confidence, adjust strategies, and help decision-making, to achieve the purpose of improving the success rate of passing the postgraduate examination [2]. And many factors affect whether they can pass an exam, in this study, the study hour before the exam and the score of the last exam are selected as the variables that affect whether the exam can be passed. By studying the effect of these two variables on the ability to pass the test, people can better plan the study time before the exam, make more adequate preparation for the exam, and make more accurate test results.

Masui's research has shown that for most courses, study time can predict grades [3]. Studies have indicated a statistically significant correlation between academic achievement and online time limits. [4]. Most students in this survey believed that if they were not always able to access technology, they would devote more time to studying [4]. This showed from the side that academic success and learning time have a definite correlation.

DOI: 10.54254/2753-8818/36/20240526

There are numerous of literature on the study of influencing factors of students' scores. Sun et al. tested the results of the mediation model to show that allocation of study time played a significant dual mediating role between self-management and academic achievement [5]. The results showed that study time and active use of time have significant positive predictive effects on the academic achievement of college students [5].

Liao used multiple logistic regression method to analyze various factors affecting the academic performance of economic and management students in universities [6]. The results showed that factors such as the proportion of spare time have a significant impact on academic performance [6]. Based on binary logistic regression analysis of the data, Zhang analyzed the final path model fitting index and influence coefficient of the influence of learning motivation on learning time investment and grade ranking [7]. The results suggest that student effort can be measured by attendance and extracurricular learning engagement, both of which have important effects on student achievement [7]. From the survey results of data collected by Du et al., they can also see the fact that students with poor academic performance generally spend less time studying [8]. Eliasson et al. also compared the behavior patterns and academic performance of different students [9, 10].

To sum up, the research on the influencing factors of academic performance has attracted many scholars. However, In the above research on the influence factors of academic performance, the influence factor of the last exam score is missing. None of the above literature directly studies the relationship between study time, last test score, and whether the test can pass. This study mainly employs the binary logistic regression model in pursuit of analyzing and predicting the research results by studying the relationship between the two variables of the last exam score and study time and whether the exam can be passed and provides corresponding suggestions according to the predicted results.

2. Methods

A dataset consisting of data from 500 students was used. The variables are analyzed and selected. Finally, the logit binary regression model was chosen.

2.1. Data source

The dataset is designed to predict whether a student will pass or fail an exam. It is based on the number of study hours and their scores in the previous exam. The dataset includes data of 500 students. It ensures an extensive range of study patterns and the performances of previous exams. It contains three indicators, namely Study Hours, Previous Exam Score, and Pass/Fail. This dataset was last updated in 2024.

2.2. Variable selection

In this part, 100 observations were obtained from variables Study Hours and Previous Exam Scores of the original data set. The specific description of the 100 observations is shown in Table 1 and Table 2.

Table 1. Essential feature.

Name	Sample size	Minimum	Maximum	Mean	Standard deviation	Median
Study Hours	100	1.050	9.882	5.232	2.677	5.177
Previous Exam Score	100	40.873	99.983	71.518	17.827	71.692

Table 2. Analysis of variance.

	Pass/Fail (mean value standard deviation)		F	p
	0.0(n=59)	1.0(n=41)		
Study Hours	3.59±2.01	7.60±1.46	118.894	0.000
Previous Exam Score	66.14±19.45	79.25±11.57	14.930	0.000

DOI: 10.54254/2753-8818/36/20240526

2.3. Method introduction

In some cases, the dependent variable takes 0 or 1. Under the circumstances, it is easy to build a regression model using CDF. In this dataset, Y is a fixed type of data, the numbers are 0 and 1. So the logit binary regression model, which is applicable when the sample size is small, is chosen.

The formula of binary logistic regression is $P(Y=1)=\frac{1}{1+e^{-X\beta}}$. Where P(Y=1) represents the probability that the eigenvalue Y is 1, X represents the vector value of the input variable, and β represents the vector value of the regression coefficient.

3. Results and discussion

This part first draws the box plot of the data and does the related analysis. Then the data is preprocessed to exclude the case of missing values. Then the overall effectiveness of the model is analyzed, and the analysis result is obtained.

3.1. Descriptive statistical analysis

From Figure 1, the dispersion of the data is small, the data range is small, and there are no outliers. This shows that the data distribution is relatively concentrated, the range of change is small, and the representativeness of the average is large.

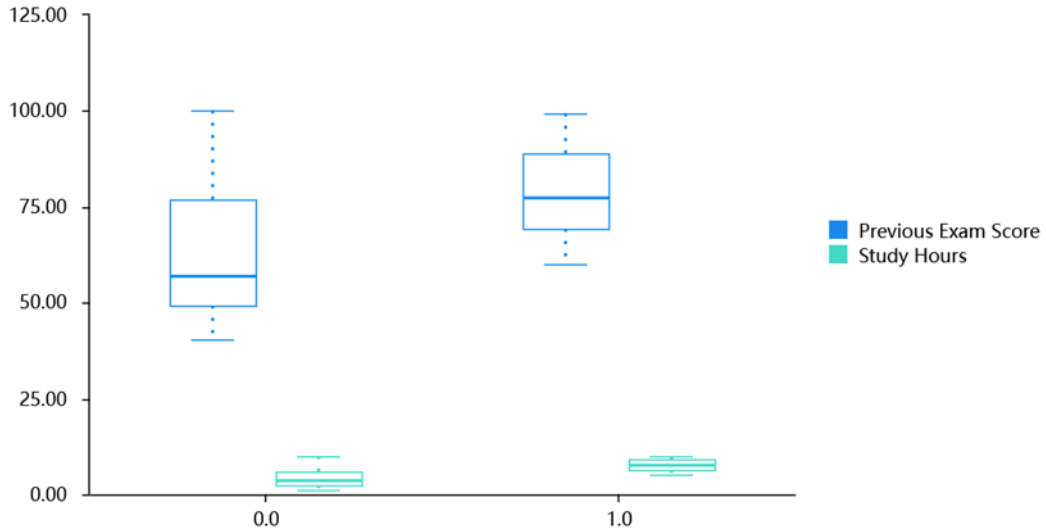

Figure 1. Boxplot of study hour and previous exam score.

3.2. Data preprocessing

The binary Logit regression analysis used study hours and previous exam scores as independent variables and pass/fail as dependent variables. A total of 500 samples participated, with no missing data. The summary is presented in Table 3.

Table 3. Binary logistic regression analysis basic summary

Name	Options	Frequency	Percentage
Pass/Fail	0	316	63.20%
	1	184	36.80%
	Total	500	100.0%
Summary	Valid	500	100.00%
	Missing	0	0.00%
	Total	500	100.0%

DOI: 10.54254/2753-8818/36/20240526

3.3. Model checking

The model's effectiveness is analyzed, rejecting the original hypothesis of equal model quality when Study Hours (Previous Exam Score) are involved. The P-value is less than 0.05. It indicates the validity of the independent variables (Table 4).

Table 4. Test results of the likelihood ratio of the binary logit model

Models	Logarithmic likelihood	Chi-square	do	p	AIC	BIC
Intercept only	657.883	-	-	-	-	-
Final model	269.134	388.749	2	0.000	275.134	287.778

The original assumption of the model test here is that the model fitting values are consistent with the observed values. As can be seen from the above table, the p-value here is greater than 0.05 (Chi=6.945, p=0.543>0.05), which indicates that the original hypothesis is accepted, that is, the model passes the HL test and the goodness of fit of the model is good. Table 5 shows this result.

Table 5. Hosmer-Lemeshow fit test

χ^2	do	p-value
6.945	8	0.543

3.4. Model results

We summarized the different indicators of the two variables in Table 6 and obtained the conclusion through analysis.

Table 6. Summary of results of binary logit regression

Item	RG	SE	z	Waldχ^2	p	OR	OR95% CI
Study Hours	1.146	0.114	10.061	101.231	0.000	3.145	2.516 ~ 3.931
Previous Exam Score	0.143	0.015	9.412	88.585	0.000	1.154	1.120 ~ 1.189
Intercept	-17.759	1.723	-10.305	106.191	0.000	0.000	0.000 ~ 0.000

Dependent variable: Pass/Fail

McFadden R square: 0.591

As can be seen from the above table, Study Hours and Previous Exam Score can explain 0.59 changes in Pass/Fail. The formula of the model is $\ln\left(\frac{p}{1}-p\right) = -17.759 + 1.146 * Study\ Hours + 0.143 * Previous\ Exam\ Score$. Where p represents the probability of Pass/Fail being 1 and 1-p represents the probability of Pass/Fail being 0.

The final concrete analysis shows that: The regression coefficient value of Study Hours was 1.146 and showed a significance level of 0.01 (The study found a significant positive impact of Study Hours on Pass/Fail, with an odds ratio of 3.145, indicating a significant positive effect.) in Pass/Fail is 3.145 times when the Study Hours are increased by one unit. The regression coefficient indicates a significant positive impact of the Previous Exam Score on Pass/Fail, with an odds ratio of 1.154, indicating a significant increase in Pass/Fail. The model results are shown as follows (Figure 2):

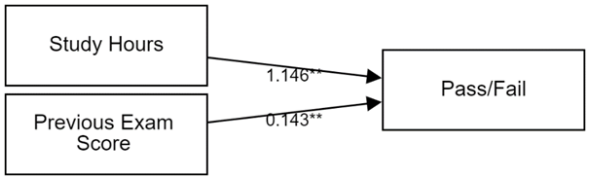

Figure 2. Model results

DOI: 10.54254/2753-8818/36/20240526

Summary analysis shows that both Study Hours and Previous Exam Scores have a significant positive impact on Pass/Fail.

4. Conclusion

This Study selected a dataset containing three variables: Study Hours, Previous Exam Score, and passing or not of 500 students. The influence of Study Hours and Previous Exam Scores on passing the exam was analyzed. The method (binary Logistic regression analysis) is accurate, valid, and comprehensive. Because the number of independent variables and the type of dependent variables (qualitative data) of the data meet the conditions of using binary logistic regression, the goodness of fit of the model is good. In the analysis stage, we used the Binary logistic regression model. We used the model to find out the possible relationship between Study Hours, Previous Exam Scores, and passing the exam. The formula for binary logistic regression is $P(Y=1)=\frac{1}{1+e^{-X\beta}}$. Among them, P means the probability. The probability that the eigenvalue Y is equal to 1 is called $P(Y=1)$, X represents the vector value of the input variable, and β represents the vector value of the regression coefficient.

Through analysis, Study Hours and Previous Exam Scores have a positive impact on whether the exam can be passed. Admittedly, due to the limited amount of data and the fact that the sample did not cover all ages and ethnicities, there can be errors. However, there is still a lot of value and merit in this research. With this research, people who aspire to pass exams can have a reference point in terms of both Study Hours and Previous Exam scores. The research could also help people make more accurate judgments about whether they will pass exams. Furthermore, people can also obtain their satisfaction by changing the above two variables.

References

[1] Ha H, Lee H S 2019 Effect of Interim Testing on Learners' Metacognitive Judgments, Study Time, and Learning Performance in Concept and Category Learning. The Korean Journal of Educational Psychology, 125-152.

[2] Wang H H, Zhang Y, Jiang Y Q 2024 Prediction Analysis of college students' Postgraduate Entrance Examination Results based on Data Mining. Wuyi College Journal, 93-97.

[3] Chris M, Jan B, Sarah D, Anne G, Geert M 2014 Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education. Studies in Higher Education, 621-643.

[4] Adoración D L, Ana B M, Javier M 2021 Adolescents' Perceptions of Their Problematic Use of ICT: Relationship with Study Time and Academic Performance. International Journal of Environmental Research and Public Health, 6673-6673.

[5] Sun L Z, Zhang H, Liu Q L 2022 College students after school time distribution and the impact on the academic achievement of empirical studies. Journal of Chongqing liberal arts college (social science edition), 127-140.

[6] Liao Y J 2020 Analysis on the influencing factors of college students' Academic Performance. Journal of Shanxi University of Energy, 33(5), 45-47.

[7] Zhang X 2018 Study on the influence of learning motivation and Effort on Students' Academic Achievement. Educational Theory and Practice, 38(15), 31-33.

[8] Du X Y, Xu Y, Li X N, et al. 2017 Late Guangzhou college students learning situation and influencing factors of. Journal of School Health in China, 38(4), 592-594.

[9] Achamma G 2012 Influence of Behavioral Patterns on Academic Performance- Comparison between Students of the Morning session and Afternoon session at the College of Health Sciences (CHS), Kingdom of Bahrain. Revista Internacional de Sociología de la Educación, 54-70.

[10] Eliasson A H, Lettieri C J 2017 Differences in sleep habits, study time, and academic performance between US-born and foreign-born college students. Sleep & Breathing, 529-533.