ISSN 2959-6157

Research on the Factors Influencing High School Achievement

Qiya Wang[1, *]

[1]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, 999077, China
*Corresponding author: s1139476@s.eduhk.hk

Abstract:

There have been many studies on the factors that affect students' academic performance, but scholars have reached different conclusions, and many studies and analyses are still needed. The purpose of this essay is to identify the variables influencing high school students' academic achievement. Using the method of multiple linear regression, 12 variables related to academic performance were selected, and a significant analysis of 2,392 high school students was carried out in 2024. In order to check the effectiveness of the operation, the residual test was conducted in this study. The results indicate that high school students' academic achievement is positively correlated with the weekly study time, whether to participate in tutoring, whether to participate in extracurricular activities, whether to participate in sports, and the degree of parental support, and negatively correlated with the number of absenteeism. The students' age, gender, ethnicity, parents' education level and whether they participate in volunteer activities did not pass the significance test.

Keywords: Academic performance; high school students; multiple linear regression.

1. Introduction

Nowadays, education has become a major force to promote the progress of human civilization. Education plays an important role not only for individuals, but also for the development of society and a country. Education instills in the individual the acquisition of basic skills and knowledge and promotes the individual's material and spiritual pursuits [1]. For society and country, education provides human resources guarantee for national construction and promotes the overall growth of society [2]. How well students master knowledge is usually judged by their academic performance. Academic achievement affects students' future development to some extent. There are numerous factors that influence the students' academic achievement. In order to assist relevant instructors in raising students' academic accomplishment, the purpose of this study is to investigate the elements that influence high school students' academic achievement.

On the one hand, a variety of internal characteristics, such as gender and involvement, influence students' academic achievement. The effect of gender on the academic achievement of secondary school pupils is a complicated topic. It is generally believed that male and female students perform differently in different subjects. Girls excel in the humanities, whereas boys excel in science [3]. However, a survey reveals that male and female students' performance in the humanities and STEM fields-science, technology, engineering, and mathematics—does not significantly differ from one another [4]. In many developed countries, far fewer men than women receive higher education. The reason for the difference in college enrollment rates between men and women may be the result of boys' poor reading ability and negative social attitudes towards girls' education [5]. In addition, there is a close relationship between high school students' academic performance and their investment in learning. An analysis of students' behavioral, emotional, and cognitive engagement areas showed that there was a positive correlation with academic performance in almost all students [6]. Students tend to do better academically the more time they devote to studying each week. Even if students spend a lot of time studying each week, if the learning method is incorrect or inefficient, it may lead to poor learning results.

On the other hand, a variety of outside variables have an impact on how well high students achieve academically. Children will perform better academically when their parents have greater education. Parental education often affects occupation and income, which provides children with cognitive stimulation inside and outside the family environment, indirectly supporting children's academic success [7]. More educated parents mean they are more knowledgeable and can provide more guidance and help to their children. Children's academic achievement is positively impacted by parental education [8]. Parents with higher education usually pay more attention to their

children's education, and they provide better learning resources. Nevertheless, it is not that poorly educated parents are unable to provide support to their children, and their children are unable to achieve well. Furthermore, the study demonstrates a strong positive correlation between student test results and private tutoring [9]. Students may not be able to fully master all the knowledge points in school, and tutoring can help students check for and fill in the gaps in a targeted manner. Conversely, a study found only weak evidence that private tutoring is effective [10]. Students' performance and their involvement in extracurricular interest classes are somewhat correlated. Students that participate in extracurricular activities develop their character and do better academically [11]. Participating in extracurricular interest classes can help students broaden their horizons, cultivate students' comprehensive abilities, and help improve their academic performance. Despite that, excessive participation in extracurricular interest classes may lead to students' academic burden being overloaded and have a negative impact on academic performance [12].

In summary, this paper will use a multiple linear regression model to analyze the internal and external factors that affect the academic achievement of high school pupils, compare the degree of influence of each factor on academic performance, and thus provide references and suggestions for relevant educators to improve high school students' academic achievement.

2. Methods

2.1 Data Source

The data used in this article is from the Kaggle website, which was released in 2024. This dataset presents elements that impact students' academic success and includes detailed information on high school students between the ages of 15 and 18.

2.2 Variable Selection

The data counts a total of 2,392 high school students, and the data contains 12 variables (Age, Gender, Ethnicity, ParentalEducation, StudyTimeWeekly, Absences, Tutoring, ParentalSupport, Extracurricular, Sports, Music and Volunteering). The specific description of this dataset is shown in Table 1:

Table 1. List of variables

Variable	Logogram	Meaning
Age	x1	The students' ages range from 15 to 18 years old
Gender	x2	Male (0) and female (1)
Ethnicity	x3	Caucasian (0), African American (1), Asian(2) and other (3)
ParentalEducation	x4	None (0), high school (1), some college (2), bachelor's (3) and higher (4)
StudyTimeWeekly	x5	Hours spent studying per week, ranging from 0 to 20
Absences	x6	Total number of absences from 0 to 30 in a school year
Tutoring	x7	No (0) and yes (1)
ParentalSupport	x8	None (0), low (1), moderate (2), high (3) and very high (4)
Extracurricular	x9	No (0) and yes (1)
Sports	x10	No (0) and yes (1)
Music	x11	No (0) and yes (1)
Volunteering	x12	No (0) and yes (1)
GPA	y	Grade Point Average (GPA) on a 2.0-4.0 scale

2.3 Method Introduction

Multiple linear regression can quantitatively explain how several independent variables relate to dependent variables through regression functions. This regression function is called a linear regression model. The regression equation estimated using sample data is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad (1)$$

In the above formula: β_0 is the regression constant, β_1, β_2, …, β_k is the regression coefficient, and ϵ is the random error term. This paper will use SPSS to analyze the impact of x on y, that is, the impact of the 12 factors on the academic performance of high school students. After taking GPA (y) as the dependent variable, and the 12 factors (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11} and x_{12}) as the

independent variables, the formula can be established as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{12} x_{12} + \epsilon \quad (2)$$

3. Results and Discussion

3.1 Correlation Analysis

This article's investigation demonstrates the variety of factors influencing high school pupils' academic achievement. As shown in Table 2:

Table 2. Relevance analysis between dependent and independent variables

	y
x1	0.000
x2	-0.013
x3	0.028
x4	-0.036
x5	0.179**
x6	-0.919**
x7	0.145**
x8	0.191**
x9	0.094**
x10	0.058**
x11	0.073**
x12	0.003
* p<0.05 ** p<0.01	

The Pearson correlation coefficient between these 12 parameters and high school pupils' academic achievement is displayed in the table above. The correlation coefficient values between y and x_1, x_2, x_3, x_4 and x_{12} are all close to 0, and the p value is greater than 0.05, which shows that these five variables have no bearing on how well high school pupils succeed academically. Research data shows that there is no significant correlation between high school students' academic performance and their age, gender, ethnicity, parents' education level, and whether they participate in volunteer activities. This finding suggests that students' academic performance is not directly affected by these factors, which may be different from previous perceptions of factors affecting academic performance. Students' academic success is not solely determined by their background information such as age, gender, ethnicity, or parents' education level, but may be more dependent on students' personality and learning style.

The correlation coefficient values between y and x_5, x_7, x_8, x_9, x_{10} and x_{11} are all greater than 0, and show significance at the 0.01 level, which shows that high school students' academic achievement is positively correlated with their weekly study time, whether to participate tutoring, degree of parental support, whether to participate in extracurricular activities and sports. This means that students' active participation in these aspects may have a positive impact on their academic performance. Devoting more study time per week generally promotes the accumulation of knowledge and depth of understanding, thereby improving academic performance. Participating in tutoring and extracurricular activities can broaden students' knowledge and provide them with more support in their studies. In addition, participating in sports helps to maintain physical health and energy, and has a positive impact on learning status and performance. Finally, parental support is a crucial factor in students' learning process, which can motivate students to better engage in learning and achieve better results.

There is a negative link between high school students' academic performance and absences, as indicated by the correlation coefficient value of -0.919 between y and x_6, which is significant at the 0.01 level. This means that students' academic performance tends to decline as their absences increase. Increased absences may cause students to miss important learning content and classroom interactions, which affects their mastery and understanding of knowledge. In addition, absences also affect students' communication and interaction with teachers and class-

mates, reducing the effectiveness and enthusiasm of learning. Therefore, reducing absences is crucial to improving students' academic performance.

3.2 Model Results

According to the results in Table 3, the relevant multiple linear regression equation can be obtained:

$$y = 2.597 + (-0.006)x_1 + 0.015x_2 + \cdots + (-0.009)x_{12} \quad (3)$$

Table 3. Regression coefficient table

	B	S.E.	Beta	T	significance	VIF
Constant	2.597	0.061	-	42.407	0.000**	-
x1	-0.006	0.004	-0.007	-1.542	0.123	1.008
x2	0.015	0.008	0.008	1.854	0.064	1.004
x3	0.003	0.004	0.003	0.669	0.503	1.005
x4	0.002	0.004	0.002	0.432	0.666	1.006
x5	0.029	0.001	0.179	40.571	0.000**	1.003
x6	-0.100	0.000	-0.923	-209.572	0.000**	1.005
x7	0.250	0.009	0.125	28.473	0.000**	1.005
x8	0.151	0.004	0.186	42.188	0.000**	1.005
x9	0.191	0.008	0.101	23.042	0.000**	1.002
x10	0.194	0.009	0.098	22.138	0.000**	1.005
x11	0.143	0.010	0.062	14.121	0.000**	1.004
x12 -0.009		0.011	-0.003	-0.771	0.441	1.004
R^2	0.954					
Adj R^2	0.954					
F	F (12,2379) =4117.844, p=0.000					
D-W	1.984					

With an R-squared of 0.954, the model can account for 95.4% of the variation in y that can be explained by the independent variables. Upon doing the F test on the model, it is discovered that the model passes (F=4117.844, p=0.000<0.05), suggesting that y would be impacted by at least one of the 12 components. The model's multicollinearity test revealed that all of the VIF values are less than 5, indicating the absence of a collinearity issue. Furthermore, the D-W value is in proximity to 2, indicating the absence of both autocorrelation in the model and correlation among the sample data. The model has a good fit.

The results of the study remind that education should be fair, and students should not be judged unfairly because of their personal backgrounds. In addition, academic achievement can be positively impacted by parental support and students' active participation in learning. Educators and parents should work together to provide students with a better learning environment and support system to promote their all-round development and learning achievement. At the same time, schools need to take corresponding measures, such as providing make-up opportunities for classroom content or online learning resources, to help students who are absent due to special reasons to resume their learning progress as soon as possible, so as to attract students' attention to attendance and learning outcomes (Figure 1).

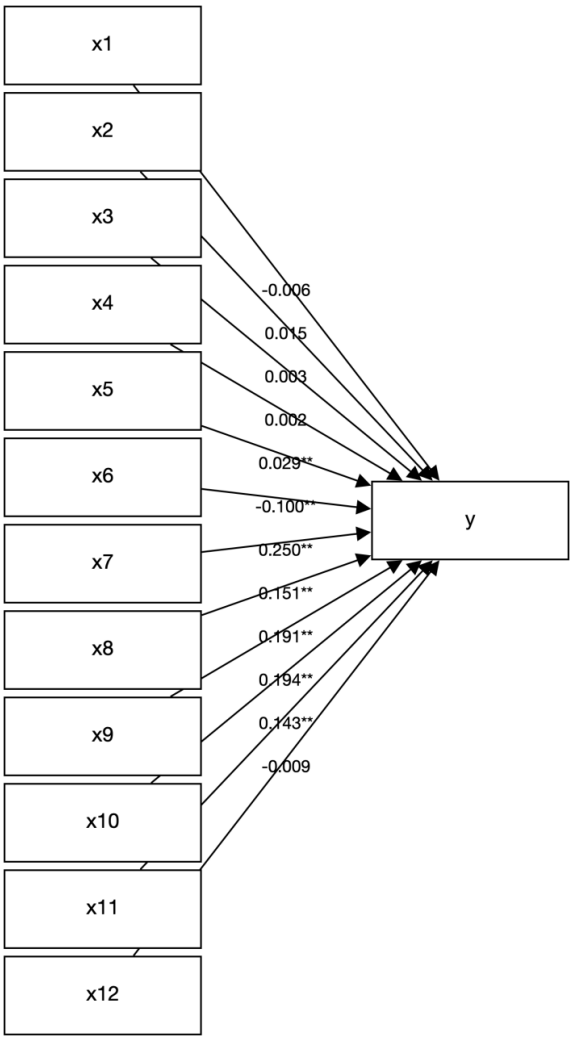

Fig. 1 Model results

3.3 Residual Analysis

The figure 2 below is a normality test of the residuals. The residual histogram can be used to determine whether the residuals conform to normality.

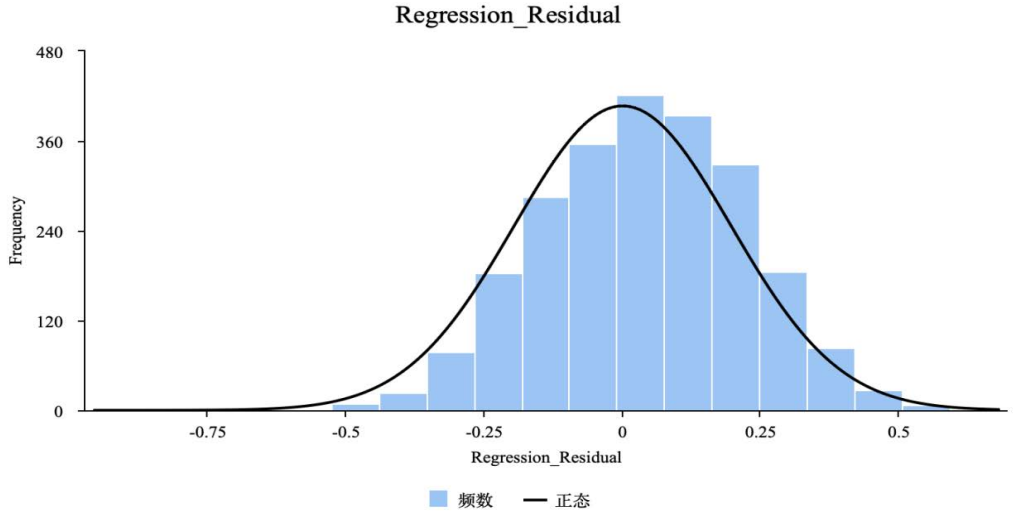

Fig. 2 Residual normality testing

The expected value was used as the x-axis and the residual value as the y-axis in the residual equivariance test. The figure below is a scatter plot of the linear regression predicted value and the regression residual value. All the points in Figure 3 are evenly distributed on both sides of the straight line $y = 0$, which shows that the residuals of the linear regression model established this time meet the homogeneity of variance.

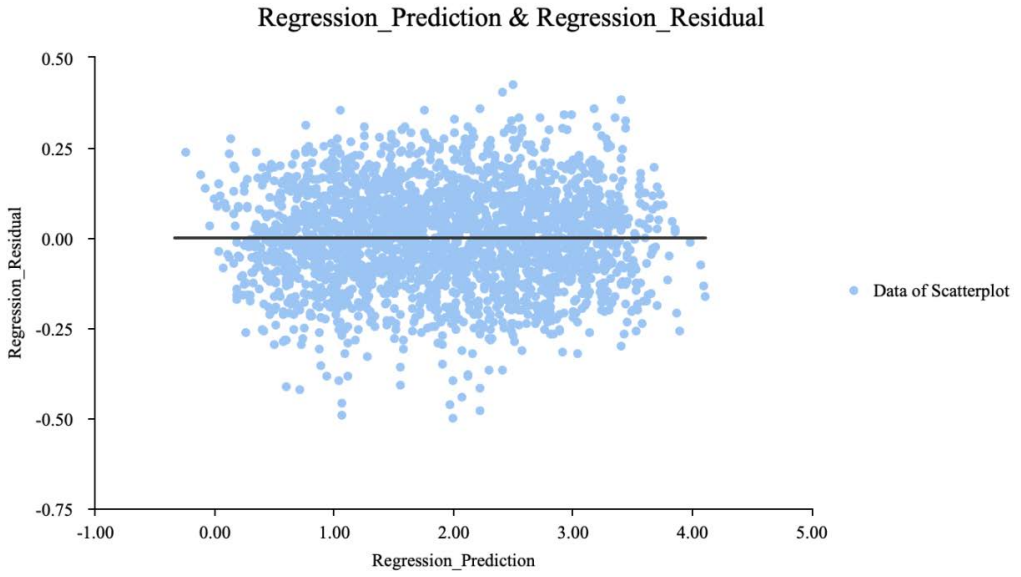

Fig. 3 Residual equal variances testing

The QQ diagram in Figure 4 is a straight line, indicating that the residual can be considered as white noise.

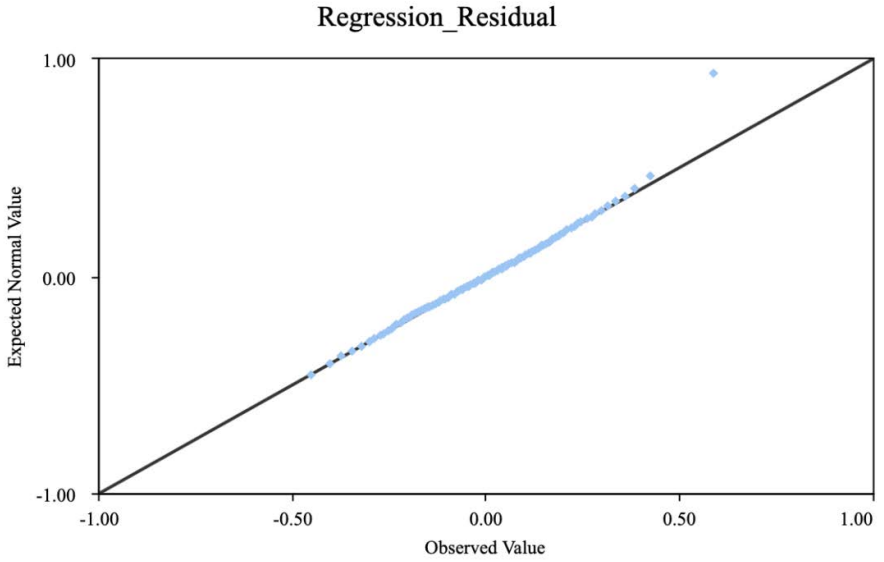

Fig. 4 Normalized Q-Q plots of regression standardized residuals

4. Conclusion

The study selected data from 2024, including 2,392 samples and 12 variables. The conclusion is that the academic performance of high school students may not be related to the students' age, gender, race, parents' education level, and whether they participate in volunteer activities. It may be positively correlated with weekly study time, whether they participate in tutoring, whether they participate in extracurricular activities, whether they participate in physical exercise, and the degree of parental support, and may be negatively correlated with the number of absences. The

multivariate linear regression analysis method used in this study is accurate, effective, and comprehensive, because it performed multivariate analysis and then obtained the Pearson correlation coefficient for each variable.

Although the findings of this study may contradict preconceived notions about what influences the academic achievement of high school students, the results also provide more inspiration for educators to cogitate. The data shows that there is no obvious correlation between high school students' academic performance and their background, which means that educators need to ensure that every student has equal learning opportunities. In addition to the students' own efforts, educators and parents should work together to create a better support system for students to promote their academic achievement. However, due to the limited amount of data, the sample did not cover all high school student groups, so the accuracy of the results may be affected. To address this issue, more research into new data sources and the use of control variable techniques are required in order to ascertain any potential causal relationships between various factors and the academic achievement of high school pupils. Such a research method will help to have a more thorough knowledge of the influencing factors behind academic performance, thereby providing more effective guidance for education policies and practices.

References

[1] Ben-Eliyahu A. Sustainable learning in education. Sustainability, 2021, 13(8): 4250.

[2] Chankseliani M, Qoraboyev I, Gimranova D. Higher education contributing to local, national, and global development: new empirical and conceptual insights. Higher Education, 2021, 81(1): 109-127.

[3] Trusz S. Why do females choose to study humanities or social sciences, while males prefer technology or science? Some intrapersonal and interpersonal predictors. Social Psychology of Education, 2020, 23(3): 615-639.

[4] Chang D F, Chang Tzeng H C. Patterns of gender parity in the humanities and STEM programs: The trajectory under the expanded higher education system. Studies in Higher Education, 2020, 45(6): 1108-1120.

[5] Stoet G, Geary D C. Gender differences in the pathways to higher education. Proceedings of the National Academy of Sciences, 2020, 117(25): 14073-14076.

[6] Lei H, Cui Y, Zhou W. Relationships between student engagement and academic achievement: A meta-analysis. Social Behavior and Personality: an international journal, 2018, 46(3): 517-528.

[7] Davis-Kean P E, Tighe L A, Waters N E. The role of parent educational attainment in parenting and children's development. Current Directions in Psychological Science, 2021, 30(2): 186-192.

[8] Idris M, Hussain S, Ahmad N. Relationship between parents' education and their children's academic achievement. Journal of Arts & Social Sciences, 2020, 7(2): 82-92.

[9] Zheng X, Wang C, Shen Z, Fang X. Associations of private tutoring with Chinese students' academic achievement, emotional well-being, and parent-child relationship. Children and Youth Services Review, 2020, 112: 104934.

[10] Ömeroğulları M, Guill K, Köller O. Effectiveness of private tutoring during secondary schooling in Germany: Do the duration of private tutoring and tutor qualification affect school achievement? Learning and Instruction, 2020, 66: 101306.

[11] Yousry N. Policy Evaluation of the role of Extracurricular activities on students' Character building and Academic Excellence: A Case Study of Cairo's Schools. Journal of Arts & Social Sciences, 2022.

[12] Carmona J V. Teacher Perception about Excessive Amounts of Extracurricular Activities that Interfere with Regular Classes. Revista Educación, 2020, 44(1): 533-548.