Adult and Higher Education (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/aduhe.2024.060515
ISSN 2523-5826 Vol. 6 Num. 5

A Study of Influential Factors of College Students Studying Abroad Based on Multivariate Statistical Methods

Keyu Chen

Eurasia International School of Henan University, Kaifeng, 475004, China

Keywords: Factors influencing admission to study abroad; multivariate statistical analysis; random forests

Abstract: This paper presents a comprehensive analysis of the factors that influence college students' decisions to study abroad. The goal is to provide a scientific basis for developing relevant study abroad policies, improving study abroad services, and enhancing the international student experience. The analysis focuses on the admission success rate of 400 international students studying abroad and considers factors such as GRE scores, TOEFL scores, statement of purpose letters, strength of recommendation letters, undergraduate GPAs, university ratings, and research experience. After conducting exploratory statistical analyses of the variables, the model was corrected using the robust standard error method to account for the heteroskedasticity present in a direct multiple linear regression. The results derived from the model were further supplemented with the random forest algorithm. This research demonstrates that the key to improving the acceptance rate of studying abroad lies in excellent undergraduate academic performance. Applicants should actively participate in research, aim to improve their language test scores (such as TOEFL and IELTS), and obtain strong letters of recommendation and other relevant background information. These initiatives aim to improve the competitiveness of applicants and increase the success rate of study abroad applications. They also aim to help higher education institutions understand the motivation and background of students who choose to study abroad, as well as improve the quality of teaching and service systems. These initiatives play an important role in optimizing the domestic education system and cultivating and importing high-quality talent.

1. Introduction

In the context of global internationalization, universities are developing exchange programs abroad. These programs aim to diversify exchanges at different levels and in different disciplines, including joint research, exchange of teachers and students, and practical competitions. The number of students participating in exchange programs abroad is increasing every year. The destinations for international students are diverse and vary for different reasons, such as international relations and family economic conditions. The UK, US, Australia, and Canada are the main destination countries for international students, but there is also an increasing number of niche international student destination

countries in Europe and Southeast Asia. Figure 1 shows the increasing number of students studying in China in recent years, based on data from the National Bureau of Statistics (NBS). The total number of students studying abroad was 540,000 in 2016 and has reached 700,000 in 2019. The research presented in this paper is significant as it can assist students in understanding the factors that increase their chances of admission. By doing so, they can plan their studies and development direction in a targeted manner, enhance their individual competitiveness, and lay the foundation for their future professional and personal development. The study of admission factors for studying abroad can also help higher education institutions better understand the motivations and backgrounds of students who choose to study abroad. This understanding can lead to improvements in the quality of teaching, service systems, and the academic and life experiences of international students. Ultimately, this can enhance the attractiveness and international competitiveness of the schools. It is important for optimizing the domestic education system and cultivating and attracting high-quality talents.

Currently, scholars are increasingly interested in students' motivations and goals for studying abroad. These motivations include personal growth, career development, and cultural experience. Understanding these factors is crucial in the decision-making process of studying abroad. Studies[1] consistently report a growing number of university students pursuing education abroad. These studies also indicate a correlation between students' socioeconomic backgrounds and their likelihood to study abroad. Specifically, individuals from more affluent families tend to prefer higher-tier institutions overseas. Academic performance is another important factor, as students with excellent scholastic achievements are more likely to pursue education abroad. Researchers[2] have explored the push-pull theory to better understand the factors affecting Chinese students' decision-making when it comes to studying abroad. This theory highlights the complex interaction between domestic educational policies, economic development levels, incentives for returnees, and the external socio-political landscape. Variables such as foreign social dynamics, legal frameworks, safety indices, educational quality, and admission criteria play a crucial role in shaping students' choices. Moreover, academic excellence and proficiency in standardized assessments (e.g., TOEFL, IELTS, GRE) are key factors in the admission process for international education. It is worth noting that[3] postgraduate students with outstanding academic records and fluency in foreign languages are more inclined to pursue education abroad.

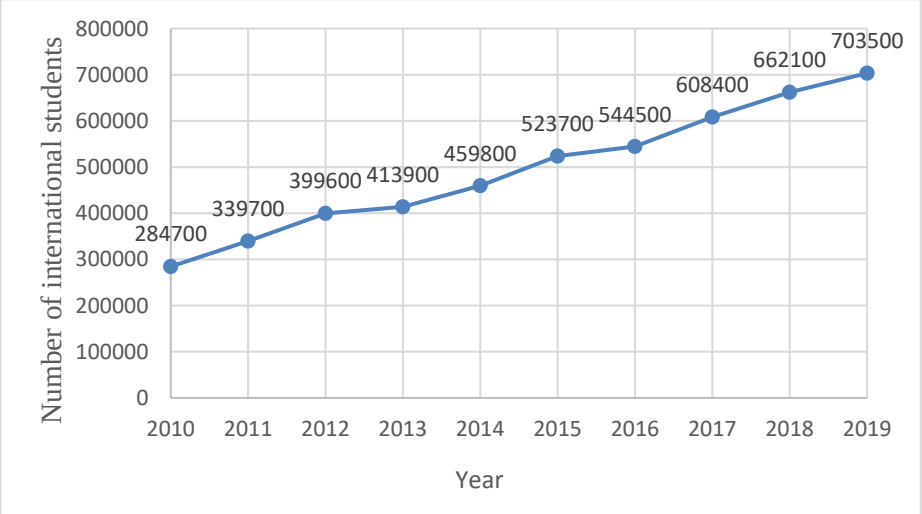

Figure 1: Number of studying abroad students from 2010 to 2019

However, it is worth mentioning that the aforementioned studies solely focus on factors influencing students' willingness to study abroad. It is essential to consider that many college students may lack a clear understanding of how to improve their chances of being accepted into a study abroad

program. Furthermore, some study abroad organizations may excessively highlight the benefits of studying abroad, which can lead to confusion and a lack of clarity during the application preparation stage.

In summary, this paper selected 400 international students as research subjects. Correlation analysis and VIF tests were conducted, revealing a strong correlation between admission success rate and GRE score, TOEFL score, statement of purpose letter, strength of recommendation letter, undergraduate GPA, university rating, and research experience. The VIF value was less than 10. We established a robust standard error for the multiple regression model and found that undergraduate GPA has the most significant impact on admission success rate. Additionally, the strength of recommendation letters and language scores also significantly impact admission success rate. Finally, this paper employs the Random Forest algorithm to provide secondary evidence and reinforce the reliability of the conclusions.

2. Variable selection and descriptive analysis

2.1 Variable section

This paper utilizes comparable variables to those utilized in Yu-Ho Chang's study [4]. GRE and TOEFL scores are significant indicators of an applicant's academic and linguistic proficiency. High scores suggest that applicants possess strong academic and language skills, making them more likely to excel in an academic environment and, consequently, more appealing. The ranking of an applicant's university reflects the quality and reputation of the education they received. Applicants who have graduated from reputable universities are often perceived as having higher academic standards and potential, and are therefore more likely to be admitted. The Statement of Purpose (SOP) demonstrates the applicant's academic interests, goals, and motivations. A clear and well-defined SOP that meets the requirements of the program to which it applies demonstrates the applicant's professionalism and academic enthusiasm, thereby increasing their chances of admission. Letters of recommendation provide an objective evaluation of the applicant's abilities and qualities, endorsing their academic ability and character. The credibility of the letters of recommendation is crucial for the applicant's admission. Additionally, the applicant's undergraduate GPA serves as an important indicator of their academic performance at the undergraduate level. A higher GPA suggests that the applicant has a strong academic record and learning ability, making them more likely to be admitted. Research experience demonstrates an applicant's academic ability, research potential, and ability to think independently. Applicants with research experience are often considered more suitable for research-based work or further study and are therefore preferred.

Table 1: Description of variables

Variable name	Description of variables
GRE Score(GS)	GRE score (ranging from 0 to 340))
TOEFL Score(TS)	TOEFL score (ranging from 0 to 120)
University Rating(UR)	university rating (on a 5-point scale)
SOP	statement of purpose (up to 5 points)
LOR	strength of recommendations (up to 5 points)
CGPA	undergraduate GPA (ranging from 0 to 10)
Research	research experience (coded as 1 for yes and 0 for no)
Chance of Admit(COA)	the probability of admission

The paper utilized 400 specific data sets which are described below. These data sets consist of a

dependent variable known as "Chance of Admit," which ranges from 0 to 1. Additionally, there are seven independent variables that were used to evaluate applicants for this program. These variables include the applicant's GRE score (ranging from 0 to 340), TOEFL score (ranging from 0 to 120), university rating (on a 5-point scale), statement of purpose or personal statement (up to 5 points), strength of recommendations (up to 5 points), undergraduate GPA (ranging from 0 to 10), and research experience (coded as 1 for yes and 0 for no). The purpose of this research paper was to examine the relationship between the probability of admission (Chance of Admit) and the aforementioned seven independent variables. Table 1 provides a comprehensive list of the variables utilized in this study.

2.2 Descriptive statistical analysis

The data were processed using R to generate the median, mean, first and third quartiles, and maximum and minimum values for each data. In order to have a clearer understanding of the data and to determine the model developed, the minimum, mean, median, maximum, and first and third quartiles were obtained for each of the influencing factors as well as the probability of being admitted.

Table 2: Descriptive analysis

Stats	**GRE Score**	**TOEFL Score**	**UR**	**SOP**
Min	290.00	92.00	1.00	1.00
1st Qu	308.00	103.00	2.00	2.50
Median	317.00	107.00	3.00	3.50
Mean	316.80	107.40	3.09	3.40
3rd Qu	325.00	112.00	4.00	4.00
Max	340.00	120.00	5.00	5.00
Stats	**LOR**	**CGPA**	**Research**	**Chance of Admit**
Min	1.00	6.80	0.00	0.30
1st Qu	3.00	8.17	0.00	0.64
Median	3.50	8.61	1.00	0.73
Mean	3.45	8.60	0.55	0.72
3rd Qu	4.00	9.06	1.00	0.83

Standardized tests, such as GRE and TOEFL, are used to assess a student's readiness for study abroad and English proficiency, respectively. Higher scores on these tests generally result in higher chances of admission. The mean GRE and TOEFL scores are 317 and 107, respectively. It is important to maintain objectivity and avoid subjective evaluations. Meanwhile, college ratings indicate the prestige or ranking of the university from which an applicant received a prior degree. Higher college ratings may positively impact admission chances. As shown in Table 2, the mean and third quartile of the University Rating (UR) are 3.09 and 4.00, respectively. This suggests that students tend to choose universities with more balanced ratings for their studies. While SOP and LOR are crucial components of graduate school applications, it is important to note that higher scores in these areas do not guarantee admission. CGPA, on the other hand, reflects an applicant's academic performance during their undergraduate studies and is often a significant factor in the admissions process. Therefore, a higher CGPA can increase an applicant's chances of admission. The Research variable is a dichotomous variable indicating whether the surveyed undergraduate student had survey

experience (1 for yes, 0 for no). Admissions committees view this positively and provide students with a probability of admission. The Chance of Admit, the dependent variable, increases as each score indicator increases.

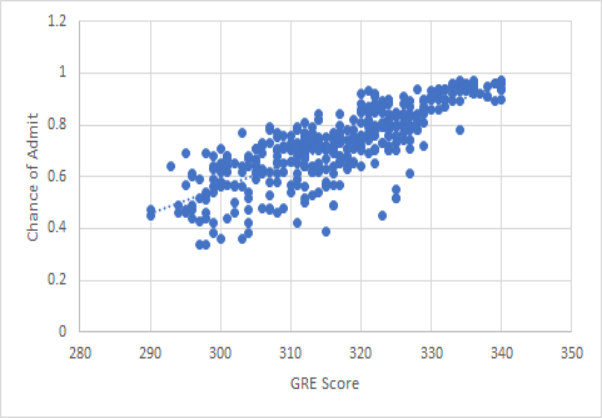

Figure 2: GRE vs ChanceofAdmit

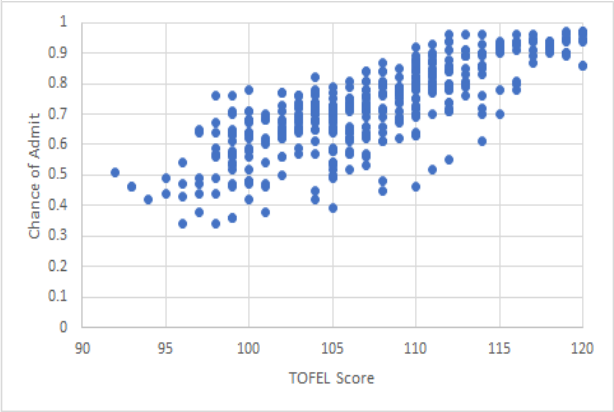

Figure 3: TOFEL vs Chance of Admit

This paper explores the relationship between independent and dependent variables, as well as the distribution of variable values. The language used is clear, objective, and unbiased. It maintains a formal tone and employs precise terminology. Figure 2 presents a scatter plot that demonstrates a positive correlation between GRE scores and admission rates. Within the score range of [330, 340], the admission rate is higher and remains more stable compared to the range of [300, 310]. It is evident that as the GRE score increases, the expected admission rate also increases. Technical term abbreviations are defined upon first usage. The text is grammatically correct and free of errors. No changes have been made to the content. The scatter plot in Figure 3 indicates a positive correlation between TOEFL scores and acceptance rates. While it is possible to achieve high acceptance rates with lower scores, the overall pattern aligns with that depicted in Figure 3.

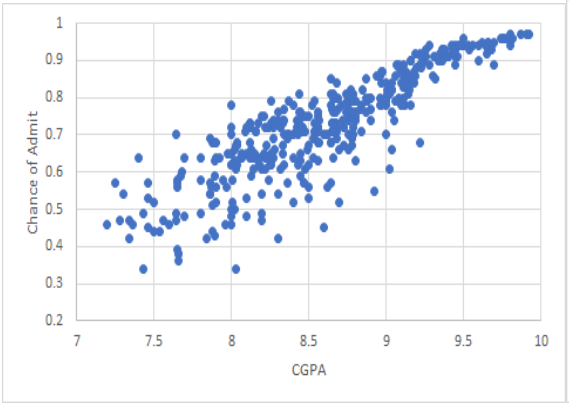

Figure 4: CGPA vs Chance of Admit

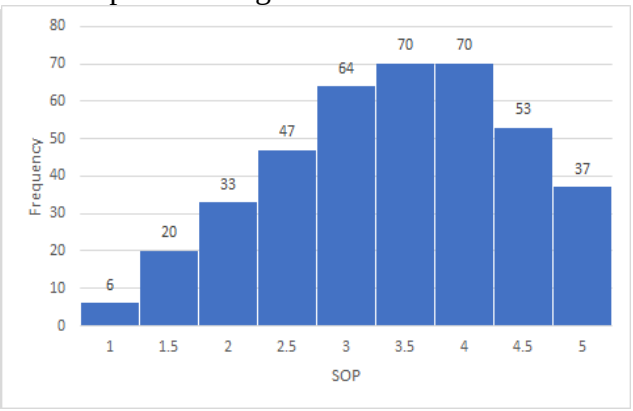

Figure 5: SOP vs Chance of Admit

From Figure 4, it is evident that there is a positive correlation between GPA and acceptance rate. Furthermore, upon observing the trend depicted in the graph, it becomes clear that as GPA increases, the expectation of a higher acceptance rate becomes more pronounced. The histogram in Figure 5 illustrates the overall distribution of recommendation letter strengths, with scores primarily concentrated in the range of 3 to 4. This indicates that the majority of individuals in the sample data received similar scores.

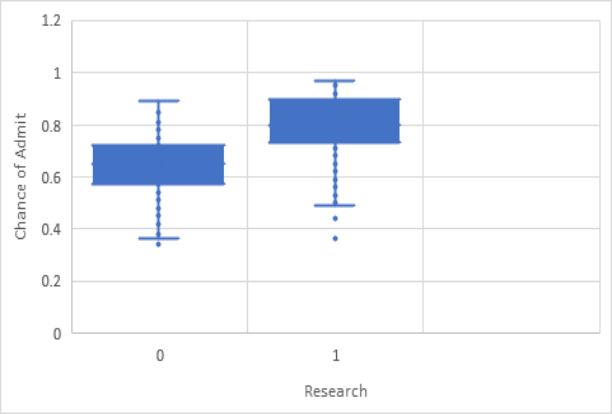

Figure 6: Research vs Chance of Admit

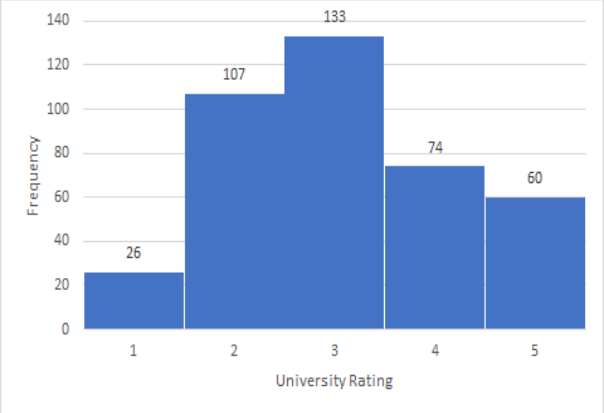

Figure 7: University Rating vs Chance of Admit

As can be seen from Figure 6, the mean acceptance rate of students with research experience is significantly higher than that of students without research experience. This indicates that research experience is an important factor in determining the acceptance rate of college students studying abroad. Figure 7 also displays a frequency histogram of university ratings, showing that medium-ranked universities have the highest number of ratings in the interval [2,3], followed by lower-ranked universities, and top-tier universities have the fewest ratings. This aligns with the popular perception that there are fewer top universities and suggests that international students tend to prefer highly-ranked institutions when applying for admission, which could potentially influence the acceptance rate. It is worth noting that the number of top-tier universities is the lowest, which further supports the common notion that there are fewer top universities. This factor can also have an impact on the choice of studying abroad, as international students tend to favor the best institutions when applying for exams, thereby affecting the acceptance rate.

In practice, the probability of acceptance is influenced by a combination of factors, which are listed as independent variables. If there is a high correlation between these variables, the model may suffer from multicollinearity. Therefore, when establishing the probability of admission and multiple regression model, it is important not to include all the factors simultaneously for analysis. Otherwise, it may lead to confusion in interpreting the regression analysis results, and even result in incorrect overall analysis and affect the interpretation of the parameter signs. In such cases, it is necessary to calculate the correlation between the independent variables. The results of correlation coefficients calculated for the above seven independent variables are shown in Figure 8:

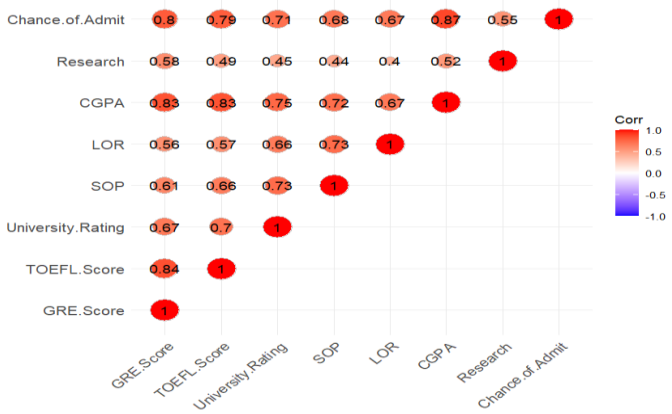

Figure 8: Heat map of variable correlation

The correlation coefficient between certain independent variables exceeds 0.8, indicating a high correlation. If these independent variables are simultaneously included in the analysis model, it may lead to the issue of multicollinearity. As a result, a VIF test was conducted on seven independent variables, yielding the following results:

Table 3: Value of vif

Variable	VIF
CGPA	5.21
GRE Score	4.62
TOEFL Score	4.29
SOP	3.08
University Rating	2.92
LOR	2.43
Research	1.54

The VIF values of the seven independent variables in Table 3 are all less than 10. Therefore, it is reasonable to assume that there is no significant multicollinearity in the model. As a result, the interpretation of the regression coefficients can generally be considered reliable. Taking this into account, the paper establishes a multiple regression model.

3. A study of the mechanisms influencing the probability of admission

Regression analysis is a statistical method used to examine the correlation between variables. In this study, a multiple linear regression model is established based on data collected from 400 international students [5]

$$y = \beta_0 + \beta_1 GS + \beta_2 TS + \beta_3 UR + \beta_4 SOP + \beta_5 LOR + \beta_6 CGPA + \beta_7 Research + \varepsilon \quad (1)$$

Where$\beta_0$is the regression coefficient, $\beta_1 \ldots \beta_7$are regression constants, GS, TS, UR, SOP, LOR, CGPA, and Research are the seven independent variables, while y is the predicted value, referred to as the dependent variable, and ε is the random error term. In the case of multiple linear regression, the general situation is solved using the least squares method to obtain the best result. This is achieved by minimizing the sum of squares of errors to reduce the difference between real data and predicted data. To test whetherβ_1 is significantly non-zero, the t-test is used with the following formula:

$$t = \frac{\widehat{\beta_1} - \beta_i}{SE(\widehat{\beta_i})} \quad (2)$$

Where $\widehat{\beta_1}$ is the sample estimate of the regression coefficientβ_1, and $\beta_i$is the hypothesis being tested. $SE(\widehat{\beta_1})$is the standard error of the regression coefficientβ_1. If the absolute value of the t-statistic is greater than a certain critical value (corresponding to the chosen level of significance, usually 0.05), the original hypothesis can be rejected. This means that the effect of the independent variable on the dependent variable is considered significant. In this paper, regression analysis was performed using Stata, and the model was adjusted using the robust standard error method. The following results were obtained:

After applying the robustness correction for heteroskedasticity to the model, it is evident from Table 4 that the goodness of fit, R2, is 0.803. This value indicates that the model effectively captures the observed data, specifically explaining 80.3 percent of the variance in the dependent variable (study abroad acceptance rate).

Table 4: Regression results

Linear regression	
Number of obs	400
F(7, 392)	319.64
Prob > F	0
R-squared	0.8035
Root MSE	0.06378

Table 5: Regression results

Chance of Admit	**Coefficient**	**std. err.**	**t**	**P>t**
GRE Score	0.001737	0.000637	2.73	0.007
TOEFL Score	0.002920	0.000957	3.05	0.002
University Rating	0.005717	0.004716	1.21	0.226
SOP	-0.003305	0.006371	-0.52	0.604
LOR	0.022353	0.005122	4.36	<0.01
CGPA	0.118940	0.012683	9.38	<0.01
Research	0.024525	0.009441	2.6	0.01
_cons	-1.259432	0.143870	-8.75	0

From Table 5, it is evident that the p-values of the independent variables GRE Score, TOEFL Score, CGPA, and Research are all <0.05. This indicates that they have a significant impact on the university admission rate and are statistically significant at the 5% level. For example, when examining the correlation between GRE Score and the Chance of Admission, it is important to pay close attention to its p-value, which should be less than 0.05. The p-value of GRE Score is 0.0007, significantly lower than 0.05. Therefore, we can conclude that it rejects the original hypothesis and accepts the alternative hypothesis at the 5% level of significance. Similarly, the p-value of TOEFL Score is 0.0002<0.05, which is also highly significant. However, when evaluating the p-values of University Rating and SOP, they are 0.226 and 0.604 respectively. These p-values do not reject the original hypothesis at the 5% significance level. Therefore, we can assume that the coefficients of University Rating and SOP are not significant for the acceptance rate of studying abroad.

To further validate the regression results, this paper employs the random forest algorithm as secondary evidence. Random Forest[6] is a comprehensive learning method that uses multiple decision trees to perform classification and regression tasks. In Random Forest, the concept of Increase Node Purity (INP) represents the IncNodePurity. Node Purity refers to the proportion of samples of the same category in a node. Higher node purity indicates that the samples in the node are more likely to belong to the same category, while lower node purity suggests a mixture of categories. IncNodePurity measures the contribution of each feature to the node's purity during node splitting. Larger values indicate higher importance of the feature in the node splitting process. Therefore, in random forests, IncNodePurity helps understand the contribution of each feature to the model and the importance of features in the decision-making process. Based on this, this paper implements Random Forest using the R language and presents the results in Table 6 below:

Table 6 shows that GRE Score, TOEFL Score, and CGPA have a greater impact on the model, which is consistent with the results obtained in multiple regression in this paper. Therefore, the three independent variables, GRE Score, TOEFL Score, and CGPA, have the greatest influence on the dependent variable.

Table 6: Random Forest Results

importance(rf_model)	IncNodePurity
GRE Score	1.1837059
TOEFL Score	0.9680919
University Rating	0.4624199
SOP	0.404232
LOR	0.3349353
CGPA	1.7016878
Research	0.1918977

4. Conclusions

The study of market data has found that GPA has a significant impact on the acceptance rate of studying abroad. Therefore, for students who wish to pursue further studies abroad, it is crucial to maintain a high GPA in order to significantly improve their chances of acceptance. Additionally, applicants should actively engage in research, strive to improve their scores on language tests such as TOEFL and IELTS, and obtain strong letters of recommendation. These initiatives will enhance the competitiveness of the applicants and increase the success rate of their study abroad applications.

In terms of future research direction, there are a few considerations. Firstly, while the linear probability model regression used in this paper allows for easy utilization, it is important to note that the impact of an independent variable on the acceptance rate may not necessarily be linear. There may be a certain point at which the marginal benefit diminishes. For instance, when the grade point average is close to 10, further increases in GPA may have little to no significant impact on the probability of acceptance. Secondly, this paper provides an analysis of the data on the study abroad market, which can offer recommendations to educational institutions and students who are interested in studying abroad. For example, in order to increase the acceptance rate of study abroad applications, it is crucial to prioritize academic performance during undergraduate studies. Based on this foundation, students should engage in professional research, improve their language test scores (such as TOEFL), and obtain strong letters of recommendation as well as other aspects of soft power background. As for study abroad educational institutions, they should provide accurate and reliable information about studying abroad, avoiding exaggerated advertising. Additionally, personalized study abroad planning and consulting services should be offered to students, and service quality should be continuously improved to meet the students' needs.

References

[1] Chen Yanglin, Yue Changjun. Trends and influencing factors of Chinese university students studying abroad: 2005-2017[J]. Education Academic Monthly, 2020, (05): 40-45+111.

[2] Feng Jiawen. "Analysis of the Factors of Domestic College Students Choosing to Study Abroad under the Push-Pull Theory [J]. Employment and Security, 2021, (09): 154-155.

[3] Fu Yun, He Yuxing, Liu Lijun. Study on the Willingness of Graduate Students of Local Agricultural Universities to Study Abroad and Its Influencing Factors--Taking Fujian Agriculture and Forestry University as an Example [J]. Journal of Fujian College of Education, 2023, 24 (01): 43-47.

[4] Zhang Yuho. An Analytical Study of Factors Influencing College Students' Admission to Study Abroad [J]. Knowledge Base, 2023, 39 (15): 179-182.

[5] Yang XL, Wang MX, Li XJ, et al. Machine learning-multiple linear regression technique to predict water ecological benchmarks for copper [J/OL]. China Environmental Science, 1-11, 2024.

[6] Han Shuaibing, Su Cheng, Kou Xiaokang, et al. Risk assessment of heavy rainfall flooding on mountain railway lines based on random forest model - A case study of Yuan ping section of Shuo Huang Railway[J/OL]. Geography and Geographic Information Science, 1-7, 2024.