Research on the Influence of Students Gender on Students Examination Scores——Based on Logistic regression model

Sen Hu¹, Yue Hu²*
¹ TangShan No.1 High School, Tangshan City, Hebei Province, 063000, China
² School of Economics and Management, Beijing Forestry University, Beijing, 100089, China
*Corresponding author’s e-mail: 342833480@qq.com

Abstract. When students enter the middle school stage, students of different genders will show different advantages in different subjects. In order to explore the influence of students' gender characteristics on students' specialties, this paper uses logit regression to conduct an empirical analysis of students' performance. This article uses student achievement data, which uses different subject scores as explanatory variables and student gender as the explanatory variable. Finally, it is found that student gender factors will have a significant impact on student performance in different subjects. Based on the analysis of student performance data, this article believes that schools should formulate different education programs according to the gender factors of students, so as to teach students in accordance with their aptitude.

1. Introduction
After students enter the middle school stage, as their psychology and physiology continue to mature, students of different genders begin to show different talents and advantages in the learning process. Generally speaking, boys tend to be calm and rational, and are good at learning in abstract thinking; while girls tend to be more emotional and good at learning in image thinking. With the advancement of the education stage, the professionalism of various subjects continues to deepen, and the difficulty continues to increase. The gender differences of students may begin to have an increasingly obvious impact on the learning effects of different subjects.

According to experience, after students have completed the division of liberal arts and sciences in high school, the proportion of girls in liberal arts students is often larger, while the proportion of boys in science students the proportion is larger. From the perspective of the composition of the male to female ratio in Chinese universities, science and engineering colleges tend to have a higher proportion of men, while colleges such as finance and economics and normal colleges tend to have a higher proportion of women. It can be inferred from this that gender differences in students may have an impact on the learning effects of different types of subjects.

2. Research Objectives
In the middle school learning process, if the gender difference of students does have a significant impact on the learning effect of different subjects, then it is necessary for the school to take into account the differences between male and female students in learning, and formulate a more reasonable teaching and training program for this purpose, so as to "teach students in accordance with their aptitude."
This article will use Eviews software to analyze the results of various subjects of high school students of different genders to determine whether the gender differences of students will affect the learning effects of different subjects. First, perform a descriptive statistical analysis on the performance of students of different genders to simply judge this difference. Then, through the establishment of a Logistic regression model between the performance of each subject and the gender of the student, it will find subjects that are significantly affected by gender differences, and accurately measure the degree of influence. According to the characteristics of disciplines, each discipline is divided into disciplines with liberal arts characteristics: Chinese, English, politics, history, and geography. These disciplines tend to be memorized; other disciplines can be divided into disciplines with scientific characteristics: mathematics, physics, chemistry, biology. These disciplines are partial to computational reasoning. According to the positive and negative regression coefficients of the various variables of the regression equation, the relationship between academic performance and the gender of the student is determined, and finally whether boys are better at the science of computational reasoning, and whether girls are better at the liberal arts of recitation and memory.

3. Models And Theories

3.1. Logistic model
For a set of sample data \( X = (1, X_1, X_2, ..., X_p)^T \), The data in \( X \) is constant. For another set of sample data \( Y \), which contains data \( y \) with the same number of components as \( X \), and these data have only two values, 0 and 1. At this time, if you want to explore the correlation between \( X \) and \( Y \), you can find a continuous value of the latent variable \( Z \), so:
\[
Z = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = \beta^T X
\]
among them \( \beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T \), and the relationship between \( Z \) and \( Y \) is, \( Z > 0, Y = 1 \)
\( Z \leq 0, Y = 0 \)
For Logistic function:
\[
g(Z) = \frac{1}{1+e^{-Z}} = \frac{1}{1+e^{-\beta^T X}} = h_\beta(X).
\]
So as to establish the function of \( X \).
Because \( Z \in (-\infty, +\infty) \),
\[
g(Z) = h_\beta(\beta^T X) \in [0, 1]
\]
\( h_\beta(X) \) is called Logistic regression model, It converts the predicted value of the linear regression model through the nonlinear Logistic function into \([0, 1]\) probability value.
Establish a linear regression model between \( Z \) and \( X \): \( Z = \beta^T X + U \), so:
\[
P(Y = 1) = P(Z > 0) = P(\beta^T X + U > 0) = P(U > -\beta^T X) = P(U \leq \beta^T X) = h_\beta(\beta^T X)
\]
So, \( P(y = 0) = 1 - h_\beta(\beta^T X) \)
Set \( h_\beta(\beta^T X) = p \), among them \( p = P(Y = 1) \)
So, \( P(y = 1|X; \beta) = p, P(y = 0|X; \beta) = 1 - p \)
And, \( Z = \ln(p/(1-p)) \). In this way, the value of \( X \) is successfully linked to the probability of occurrence of the value of \( Y \). Through the regression of \( Z \) and \( X \), the probability of the value of \( Y \) is indirectly estimated, so that the correlation between \( X \) and \( Y \) can be found by establishing a Logistic model.

3.2. Solving the parameters of the Logistic model
The probability function of the value of \( y \) is \( P(y|X; \beta) = h_\beta(X)^y \times (1 - h_\beta(X))^{1-y} \)
In order to solve the position parameter \( \beta \) in the formula, it is necessary to construct an objective function, which is the likelihood function. The likelihood function can be expressed as:
\[
L(\beta) = P(y_1|X; \beta) \times P(y_2|X; \beta) \times \cdots \times P(y_n|X; \beta) = \prod_{i=1}^n h_\beta(X_i)^{y_i} \times (1 - h_\beta(X_i))^{1-y_i}
\]
Solve the unknown parameter \( \beta \) that maximizes the objective function, namely: \( \max_{\beta} L(\beta) \), To get the estimated value \( \hat{\beta} \).
4. Models And Theories

Next, the Logistic model is established to find the subjects whose learning effect is obviously influenced by the gender of the students. Denote the gender of the student as the dependent variable gen, which is a binary discrete variable. The male student is recorded as 1, and the female student is recorded as 0. The independent variable takes the grades of each subject and is recorded as Chinese, Math, English, Physics, Chemistry, Biology, Political, History, Geometry, respectively represent Chinese score, Math score, English score, Physics score, Chemistry score, Biology score, Political score, History score, Geography score, and record the latent variable as Z. Then can build a measurement model:

\[ Z = \ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 \text{Chinese} + \beta_2 \text{Math} + \beta_3 \text{English} + \beta_4 \text{Physics} + \beta_5 \text{Chemistry} + \beta_6 \text{Biology} + \beta_7 \text{Political} + \beta_8 \text{History} + \beta_9 \text{Geometry} + \varepsilon \]

And, \( P(y = 1|X; \beta) = g(Z) = h_p(\beta'X) = p, \ P(y = 0|X; \beta) = 1 - p \)

Furthermore, by setting different critical values p, a Logistic regression model of X and Y is established.

5. Select Sample

The November 2021 monthly test results of high school students in The First High School of Tangshan City of Hebei Province were selected as sample data, and a total of 976 valid samples were obtained. The gender ‘gen’ in the sample book is used as the dependent variable, and it is recorded as 1 for boys and 0 for girls. Subject scores in the sampled book are used as independent variables, and their values take the corresponding scores.

The test goal of this article is to explore subjects whose performance can be significantly affected by the gender of the students, and then determine whether the gender of the students can have an impact on the learning effects of different types of subjects.

6. Empirical Results

In the subject model in which test scores are affected by gender factors, the logistic regression is performed with gender as the dependent variable and academic performance of each subject as the independent variable. The specific regression results are shown in Table 1.

| Subject    | Coefficient | STD      | Z       | P       | Wald | Odd    |
|------------|-------------|----------|---------|---------|------|--------|
| C          | 1.362916    | 0.827481 | 1.647066| 0.0995  | 3.686477 | 3.919915 |
| English    | -0.04281    | 0.005927 | -7.22197| 0       | 63.4439 | 0.95651  |
| History    | 0.050711    | 0.010185 | 4.978984| 0       | 27.8925 | 1.051518 |
| Math       | 0.011542    | 0.004585 | 2.517501| 0.0118  | 8.657135 | 1.012825 |
| Physics    | 0.045711    | 0.008145 | 5.612327| 0       | 41.42952| 1.049535 |
| Politics   | -0.07163    | 0.010432 | -6.86696| 0       | 50.28005| 0.932027 |
| Biology    | -0.00382    | 0.007941 | -0.48062| 0.6308  | -     | -      |
| Chemistry  | 0.002601    | 0.007936 | 0.32774 | 0.7431  | -     | -      |
| Chinese    | -0.00912    | 0.008806 | -1.0354 | 0.3005  | -     | -      |
| Geography  | 0.014138    | 0.010018 | 1.411251| 0.1582  | -     | -      |

Note: Coefficient means coefficient, STD means standard error of coefficient, Z means Z statistic, P means P value.

According to the regression results, the regression coefficients of each variable were tested with Z test. Among them, the P value of Z test for Biology, Chemistry, Chinese, and Geography was too large, which obviously failed to pass the Z test, and the statistics were not significant. Biology and Geography are relatively mixed disciplines of liberal arts and sciences, so the statistics are not
significant, and the difference in performance is not affected by gender factors and is understandable. However, Chinese is a typical subject of liberal arts, and chemistry is a typical subject of science. However, the results show that its performance is not affected by the gender of students.

According to the Wald statistics, the Wald statistic of the explanatory variable English score English has the largest value, so this variable has the greatest impact on the probability that the sample is male. The coefficient of this variable is -0.04281, and the incidence rate is 0.95651, which means that every time the English score of English increases by one point, the incidence rate of boys in this sample will drop to 95.651% of the original.

The Wald statistic value of the explanatory variable Politics achievement Politics is the second, so this variable has the second influence on the probability that the sample is male. The coefficient of this variable is -0.07163, and the incidence rate is 0.932027. This shows that for every point increase in the Politics score Politics, the incidence rate of boys in this sample will drop to 93.2027%.

The explanatory variable Physics 's Wald statistic value is the third, so this variable has the third influence on the probability that the sample is male. The coefficient of this variable is 0.045711, and the incidence rate is 1.049535. This means that for every additional point in the Physics performance of Physics, the incidence rate of boys in this sample will rise to 104.9535%.

The Wald statistic value of the explanatory variable History is the fourth, so this variable has the fourth influence on the probability of the sample being male. The coefficient of this variable is 0.050711, and the incidence rate is 1.051518. This shows that for each additional point in the historical score History, the incidence rate of boys in this sample will rise to 105.1518% of the original.

The Wald statistic value of the explanatory variable math score Math is the smallest, so this variable has the smallest impact on the probability that the sample is male. The coefficient of this variable is 0.011542, and the incidence rate is 1.012825. This means that for every additional point in the math score Math, the incidence rate of boys in this sample will rise to 101.2825% of the original.

7. Conclusion
In summary, boys are better at logical reasoning subjects, and girls are better at language comprehension subjects. For subjects with strong scientific characteristics like Physics and Math, and subjects with strong liberal arts characteristics like Politics and English, there are obvious differences between male and female students in their learning effects. It is necessary for schools to take into account the impact of students’ gender factors on learning effects in these disciplines when formulating training and education programs, so as to "teach students in accordance with their aptitude."

References
[1] De La Maza Cristóbal,Davis Alex,Azevedo Inês.Welfare analysis of the ecological impacts of electricity production in Chile using the sparse multinomial logit model[J].Ecological Economics,2021,184(184).
[2] Sopitpongstorn Nithi,Silvapulle Param,Gao Jiti,Fenech Jean-Pierre.Local logit regression for loan recovery rate[J].Journal of Banking and Finance,2021,126(126).
[3] Rainer Winkelmann,Lin Xu.Testing the binomial fixed effects logit model, with an application to female labour supply[J].Empirical Economics,2021,(prepublish).
[4] Zhang Fan,Ji Yanjie,Lv Huitao,Ma Xinwei.Analysis of factors influencing delivery e-bikes’ red-light running behavior: A correlated mixed binary logit approach[J].Accident Analysis and Prevention,2021,152(152).
[5] Ghader Sepehr,Carrion Carlos,Tang Liang,Asadabadi Arash,Zhang Lei.A copula-based continuous cross-nested logit model for tour scheduling in activity-based travel demand models[J].Transportation Research Part B,2021,145(145).
[6] Hwang Sung-Ha,Rey-Bellet Luc.Positive feedback in coordination games: Stochastic evolutionary dynamics and the logit choice rule[J].Games and Economic Behavior,2021,126(126).