

Study on the influencing factors of the student-teacher ratio based on linear regression-take China as example

Ziyan Peng

School of the Gifted Young, University of Science and Technology of China, Hefei, 238000, China

pengziyan@mail.ustc.edu.cn

Abstract. In recent years, the student-teacher ratio (STR) has received widespread attention in the field of education. By now, the STR has become an important indicator of education. This article first theoretically analyzes the relationship between the STR and the proportion of the school-age population in the middle school stage to the total population of the country, teacher gender ratio, the proportion of education expenditure in GDP, and GDP per capita, predicting that there is a positive correlation between STR and the proportion of school-age population in the total population, and there is a negative correlation between STR and the rate of women teacher, the proportion of education funds in GDP, and GDP per capita. Then this article establishes the regression equation of STR and the above variables through the data of secondary schools in China in the past 30 years. And it finds that there is a positive correlation between STR and the proportion of the school-age population in the middle school stage to the total population of the country, and there is a negative correlation between STR and the rate of women teachers, and GDP per capita, and there is no linear relation between STR and the proportion of education funds in GDP. Also, this article finds that proportion of the proportion of the school-age population in the middle school stage to the total population of the country, rate of women teachers and GDP per capita have collinearity relationship.

Keywords: Student-teacher ratio, influencing factors, linear regression.

1. Introduction

In the field of education, the concept of student-teacher ratio (STR) is often mentioned. Chen's research showed that there is an obvious linear relationship between the number of teachers and the number of students in higher education institutions in seven countries: China, France, Germany, India, Japan, the United Kingdom, and the United States, and this slope is called the STR [1]. Further, STR can be defined as the ratio of the number of students to teachers at a specific time and a certain education level.

The STR is an important indicator of education. Chen emphasized the importance of student teacher ratio by discussing the research and development status of student teacher ratio in domestic and foreign universities [2]. A lower STR means a teacher has fewer students to teach and the teacher can spend more time with the students. Based on this idea, Hu and Zhao established a mathematical model for teachers to answer questions after class and discussed that the STR is closely related to the teaching effect [3]. Guo pointed out that the number of teachers in China's higher education is insufficient, resulting in teachers' heavy workload, affecting the quality of teaching, and restricting the development

of China's higher education [4]. Shao's research showed that reducing the STR in universities can significantly increase the number of papers published by students and enhance their research abilities [5]. Nowadays, the STR has become an important indicator to measure education. Many institutions, including the national statistical offices of various countries and the United Nations Educational Scientific and Cultural Organization (UNESCO), have included the STR as one of the educational statistical indicators. The STR is also an important indicator when ranking institutions of higher learning. Fu's research showed that it is a high-probability event for American universities with low STR to enter the ranks of world-class universities [6]. The U.S. News and World Report uses average class size (total weight of 8%) and STR (total weight of 1%) in its national rankings of U.S. colleges and universities. In international rankings, QS Ranking lists the weight of STR as 20% of the total score. Meanwhile, STR is also used to estimate the demand for the number of teachers [7].

Due to the importance of STR, it is also important to study its influencing factors. Chen discussed that the STR is greatly affected by school level and subject type. There is no correlation with school size and little impact on the passage of time. Hu analyzed the STR and teaching form, teacher ability, student ability, and educational facilities. However, the above research is only based on theory, without data support, and it is difficult to quantify. Lyu's research showed that economic growth has led to a decline in the STR in China's primary schools [8]. Zhang's research showed that there is an urban-rural gap in the STR [9]. However, the above studies only explored correlations based on statistical data. Liu's research established a regression analysis of the STR, education funding, and per capita GDP [10]. Among them, education funding is calculated based on numerical values rather than the proportion of funding in GDP, making this regression equation unable to be applied elsewhere. Regarding the STR, few papers have studied its relationship with the proportion of the school-age population in the middle school stage to the total population of the country, as well as the male and female teacher ratio.

This article will theoretically analyze the relationship between the student-teacher ratio, population structure (The proportion of the school-age population in the middle school stage to the total population of the country), teacher-gender ratio, the proportion of education expenditure in GDP, and GDP per capita and then establish the regression equation of STR and population structure, the gender ratio of teachers, the proportion of education expenditure in GDP, and GDP per capita, through the data of secondary schools in China in the past 30 years.

2. Methodology

2.1. Data source

The data comes from the UNESCO Institute of Statistics (UIS). This includes the STR, and other information of the secondary school of some countries in the past 30 years. The data of the UIS Institute of Statistics comes from official statistical agencies of various countries reporting relevant data to UIS, so it's officially endorsed. However, UIS didn't verify the authenticity of the data, so this dataset may conflict with other datasets. For missing data, this article uses the linear interpolation method to fill in missing values.

2.2. Variable selection

This article uses these data below (Table 1). This article analyzes the following 5 indicators: Y, X_1, X_2, X_3, X_4 . The regression equation is $Y = \sum a_i \cdot X_i + b$. The reasons for picking these indicators are as following:

First, the proportion of the proportion of the school-age population in the middle school stage to the total population of the country. If a country has a relatively high proportion of the school-age population, the corresponding proportion of the labor force will be lower, which will result in the country not having a sufficient educational workforce to teach a large school-age population, and the SFR will be higher at this time.

Second, the Gender ratio of teachers. Perhaps based on people's biases, it is believed that female teachers can better handle issues in student life and take care of students because people believe that

women do better than men in taking care of children. This may lead to a difference in the number of students that each female teacher can take care of while teaching compared to male teachers, resulting in a different student-teacher ratio. Also, there is wage discrimination in the field of teaching, where the average salary of female teachers is lower than that of male teachers. This means that hiring female teachers with the same funding will result in more teachers, which may increase the total number of teachers and lower the student-teacher ratio in schools.

Third, The proportion of education funds in GDP. If the country's expenditure on education is high, it indicates that the country places a higher emphasis on education and has more education funds to hire more teachers, resulting in a decrease in the SFR. Last, GDP Per Capita. Similar to the proportion of education funds in GDP, higher GDP per capita will hire more teachers, which leads to a lower SFR.

Table 1. Variable introduction.

Notation	Data name and definition	Range
Y	$\frac{B}{C}$ (STR)	14.245-30.975
X_1	$\frac{A}{D}$ (the proportion of the school-age population in the middle school stage to the total population of the country)	11.274-20.702
X_2	Percentage of female teachers in secondary education	33.182-58.185
X_3	Government spending on secondary education as a percentage of GDP	0.522-0.717
X_4	GDP per capita, PPP (current international \$)	1660-21476

2.3. Method introduction

First, this paper uses linear interpolation to fill in missing data. Second, descriptive analysis of Y and each X_i . Third, for multiple linear regression, a very important point is to determine whether there is a collinearity relationship between the dependent variables, so it is necessary to make collinearity judgments on each independent variable. Last, this paper establishes a regression model between the remaining variables and Y,

$$Y = \sum a_i \cdot X_i + b \quad (1)$$

3. Result and Discussion

3.1. Description analyzes

This paper draw scatter plots for Y and X_1 as above, the linear fitting formula for scatter data is $Y = -6.651 + 1.911 \cdot X_1$, and its R-squared value (R^2) is 0.903. So, there is a significant positive correlation between Y and X_1 . Which means that STR will increase when proportion of the proportion of the school-age population in the middle school stage to the total population of the country increases (Figure 1).

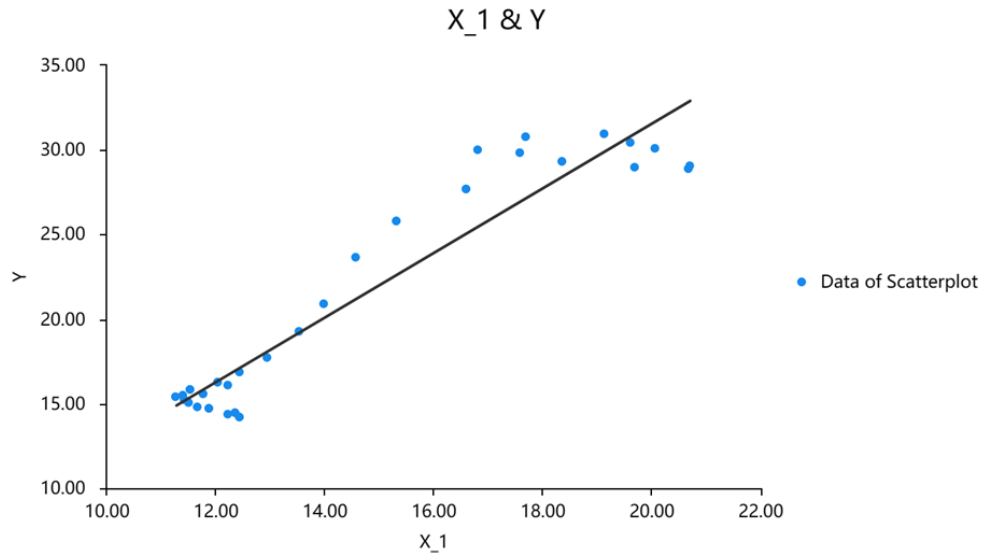


Figure 1. The scatter plot of Y and X_1 (STR and the proportion of the school-age population in the middle school stage to the total population of China)

This paper then draw scatter plots for Y and X_2 as above. The linear fitting formula for scatter data is $Y = 65.215 - 62.731 \cdot X_3$, and its' R^2 is 0.136. Thus, there is a significant negative correlation between Y and X_2 . Which means that STR will decrease when percentage of female teachers in secondary education increases (Figure 2).

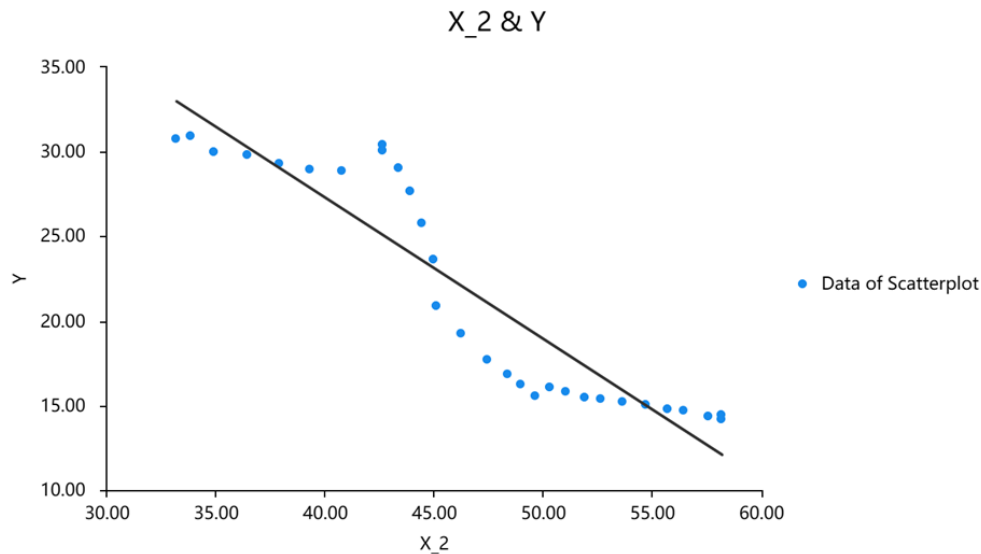


Figure 2. The scatter plot of Y and X_2 (STR and percentage of female teachers in secondary education).

This paper then draw scatter plots for Y and X_3 as above. The linear fitting formula for scatter data is $Y = 65.215 - 62.731 \cdot X_3$, and its' R^2 is 0.136. So, there is no significant correlation between Y and X_3 . Which means that government spending on secondary education as a percentage of GDP has almost no impact on STR (Figure 3).

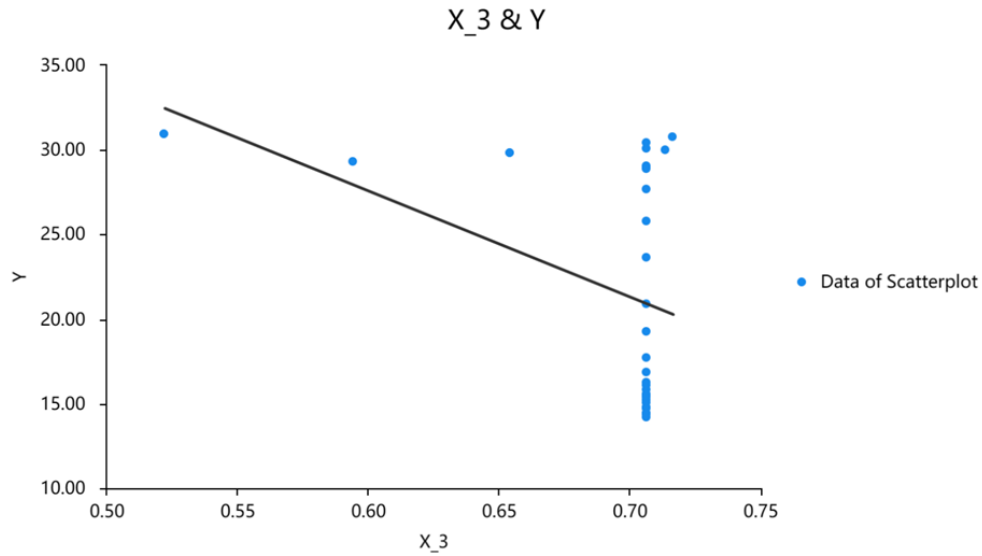


Figure 3. The scatter plot of Y and X_3 (STR and government spending on secondary education as a percentage of GDP).

This paper draw scatter plots for Y and X_4 as above. The linear fitting formula for scatter data is $Y = 30.488 - 0.001 \cdot X_4$, and its R^2 is 0.828. So is a significant negative correlation between Y and X_4 . Which means that STR will decrease when GDP per capita increases (Figure 4).

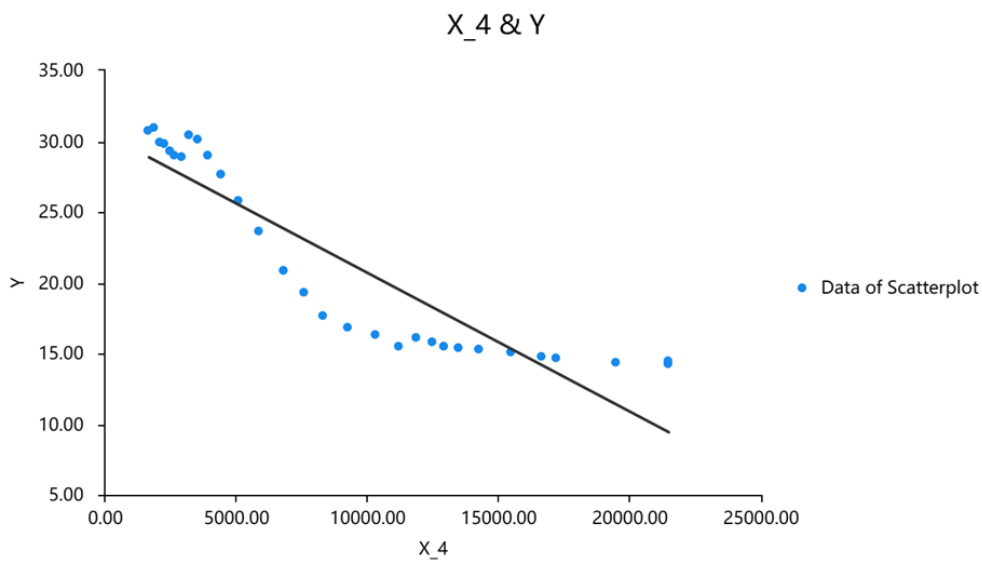


Figure 4. The scatter plot of Y and X_4 (GDP per capita)

From the result above, it can be found that there is a significant positive correlation between Y and X_1 , a significant negative correlation between Y and X_2 , X_4 , and no relationship between Y and X_3 .

3.2. Correlation analysis

This paper uses Pearson correlation coefficient and VIF value to test the collinearity of variables (Table 2).

Table 2. Pearson Correlation Coefficient

Variables	average value	standard deviation	X_1	X_2	X_3	X_4
X_1	14.783	3.360	1			
X_2	46.809	7.436	0.809**	1		
X_3	0.695	0.040	-0.337	0.434*	1	
X_4	9068.900	6270.344	0.816**	0.958**	0.319	1

For X_1 , the absolute correlation coefficient between it and X_2 , X_4 appears to be more than 0.8. For X_3 , the absolute correlation coefficients between it and the other one are all less than 0.8.

In summary, there is a collinearity issue with the two independent variables X_2 and X_4 . Therefore, in subsequent linear regression, these two variables should be removed from the model (Table 3).

Table 3. Collinearity Diagnosis

Variables	VIF	Tolerance
X_1	3.097	0.323
X_2	15.945	0.063
X_3	1.444	0.693
X_4	15.195	0.066

Using VIF value more than 10 as the criterion for collinearity, it can be seen from the above table that the corresponding VIF values of X_2 and X_4 is greater than 10. Therefore, in subsequent regression analysis, these two items should be removed from the model. Both methods show that there are a total of 2 items (X_2 , X_4) with collinearity issues, so it should remove these 2 items from the model during analysis such as regression.

3.3. Model results

Using X_1 and X_3 as independent variables and Y as the dependent variable for linear regression analysis, the model formula can be obtained as follows: $Y = 0.296 + 1.874 \cdot X_1 - 9.210 \cdot X_3$, and the model's R^2 is 0.906. At the same time, the model passed the F-test ($F=130.028$, $p=0.000 < 0.05$), indicating that at least one of X_1 and X_3 will have an impact on Y . The regression coefficient value of X_1 is 1.874 ($t=14.868$, $p=0.000 < 0.01$), indicating that X_1 will have a significant positive impact on Y . The regression coefficient value of X_3 is -9.210 ($t=-0.863$, $p=0.396 > 0.05$), indicating that X_3 does not have an impact on Y .

This means that as the proportion of school-age population in the total population increases, STR will increase linearly. Considering the collinearity of independent variables, STR also varies linearly with the percentage of female teachers in secondary education and GDP per capital. However, the government spending on secondary education as a percentage of GDP has little impact on STR (Figure 5).

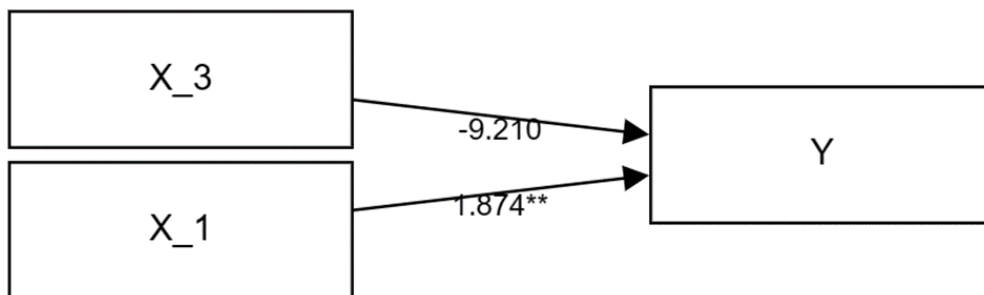


Figure 5. Model Results.

3.4. Discussion

This article first theoretically analyzes the correlation between STR and several variables, and predicts that there is a positive correlation between STR and the proportion of school-age population in the total population, and there is a negative correlation between STR and the rate of women teachers, the proportion of education funds in GDP, and GDP per capita.

This article examines the STR of various countries and finds their influencing factors. Through regression analysis, this article finds that there is a positive correlation between STR and the proportion of the proportion of the school-age population in the middle school stage to the total population of the country, that is STR will increase when the proportion of the proportion of the school-age population in the middle school stage to the total population of the country increases; there is a negative correlation between STR and the rate of women teacher, that is STR will decrease when the rate of women teacher get increased; STR doesn't have linear relationships with the proportion of education funds in GDP; there is a negative correlation between STR and GDP per capita, that is STR will decrease when the GDP per capita get increased.

However, there are shortcomings in this article. First, the dataset from the UIS has many missing values, so this article has to use the linear interpolation method to fill in missing values. If there is a dataset with fewer missing values, the regression formulas will be more accurate. Second, this article just uses the method of linear regressions, due to the limitations of the author's knowledge reserve and technical ability, the depth of research is insufficient, and the diversity of usage methods is lacking. There may be subjectivity and incompleteness in the interpretation of certain results. Third, there are many theoretically unquantifiable influencing factors on the STR at all levels, such as environmental factors, teacher utilization rate, and so on. Fourth, this article theoretically analyzes the negative correlation between STR and expenditure on secondary education, but data analysis shows that there isn't, it needs to find the reasons.

Future studies should analyze the STR with a more comprehensive dataset, and try to use more advanced methods to delve deeper into research and find more influencing factors on the STR. Future research can also start from variables with collinearity relationships and investigate the reasons for the existence of collinearity.

4. Conclusion

STR has become a very important indicator in the field of education in recent years, so it's important to find the influencing factors of STR. Few papers have studied its relationship with the proportion of the proportion of the school-age population in the middle school stage to the total population of the country, as well as the male and female teacher ratio. So, this article theoretically analyzes the relationship between STR, population structure, and teacher-gender ratio, and then establishes the univariate regression equation of the above variables (and added GDP per capita) through the data of secondary schools in China in the past 30 years. Then, this article performed collinearity analysis on each variable and established a multiple regression model for STR. The model shows that there is a positive correlation between STR and the proportion of the proportion of the school-age population in the middle school stage to the total population of the country, and there is a negative correlation between STR and the rate of women teachers, and GDP per capita, also, these three variables have collinearity relationship.

Future studies should analyze the STR with a more comprehensive dataset, and try to use more advanced methods to delve deeper into research and find more influencing factors on the STR. At the same time, the author's understanding of the reasons for the impact of the proportion of the proportion of the school-age population in the middle school stage to the total population of the country and teacher gender ratio on STR is very superficial, and further research should be conducted on how the teacher gender ratio affects STR.

References

- [1] Chen M 1991 University educational efficiency and teacher-student ratio. Higher Education Research, 64-67+78.

- [2] Chen Z and Hu B C 2013 Student teacher ratio: an important indicator of talent cultivation quality. *University Education Science*, 118-124.
- [3] Hu Y N and Zhao J H 2006 Research on the Teacher-Student Ratio Issue. Ministry of Education's National Teaching Steering Committee for Educational Technology in Colleges and Universities. *Innovation, Development and Service of Educational Technology - Proceedings of the Fifth International Forum on Educational Technology*. Central China Normal University Press, 4.
- [4] Guo D 2014 Analysis of the number of teachers in higher education in my country from the perspective of teacher-student ratio. *Henan Social Sciences*, 22(8), 97-101.
- [5] Shao Y J, et al. 2020 Empirical study on the impact of "student teacher ratio" on students' research abilities in domestic universities. *Science and Technology and Innovation*, 24, 90-92.
- [6] Fu W L and Jia J P 2019 Characteristics of teacher-teacher ratio among world-class college students in the United States. *Comparative Education Research*, 41(1), 24-31.
- [7] Zhou Q, Chen D and Qiu D F 2023 Prediction of the Demand for Teacher Quantity in Compulsory Education Stage: An International Comparison Based on Student Teacher Ratio and Class Size. *Education Research*, 44(7), 134-149.
- [8] Lv X 2018 Looking at the development characteristics of basic education in my country from the perspective of primary school enrollment rate and teacher-student ratio. *Journal of Science and Education*, 16, 132-133+138.
- [9] Zhang H M 2013 Research on the teacher-student ratio and teacher allocation in rural schools - an empirical investigation based on Hai'an County, Jiangsu Province. *Journal of Jiangsu Institute of Education*, 29(6), 47-50.
- [10] Liu L Y 2020 Analysis and prediction of influencing factors of teacher-student ratio at all levels. Chongqing University.