Research on Correlation Analysis of Student Achievement and Family Social Status under "Double Reduction" Policy Background

Yanchen Zhou

{1546663261@qq.com}

The University of Queensland ,Australia Queensland Brisbane 4072

Abstract. Currently, teaching design has shown significantly impacts on the continuous development of student achievements and abilities. Since the implementation of "double reduction" policy, the teaching model has changed and become an important factor for students future achievements. Additionally, family social status can guarantee the stuent to obtain tremendous education resources and affects students education. In this paper, we based on the "double reduction" policy and utilize the mathematical correlation analysis model to obtain the relationships among student achievements and their family social status. Furthermore, the analysis results can assist the schools to provide different teaching methods and guarantee the fairness of education. Additionally, we compare our proposed model with existing correlation analysis methods to evaluate the accuracy of analysis results and estimate the performance of our model. From our extensive experiment results and analysis, we can observe that our model can achieve the acceptable correlation prediction for students achievement and their family social status with reasonable computation costs.

Keywords: Correlation analysis, "Double reduction" policy, Students achievement, Education.

1 Introduction

Initially, the "double reduction" policy is a novel policy, which is proposed by the Chinese government to reduce the academic burden of primary and secondary school students and decrease the excessive profitability of off-campus training institutions. The primary contents of the policy include reducing the academic burden, regulating off-campus training and strict supervision [1]. As for academic burden, primary and secondary schools are not allowed to set up and open remedial schools and training courses inside and outside the school that are linked to enrollment and further education. Schools must not organize or covertly organize primary and secondary school students to participate in all kinds of competitions and examinations. At the same time, the curriculum is further optimized, and the structure of class hours is adjusted to ensure that students have enough rest and leisure time.

Additionally, the off-campus training institutions, the policy clearly prohibits them from conducting initial public offerings (IPOs) and listing financing, not allowing financing through the issuance of securities, and not financing projects related to primary and secondary school subject education and training business through the issuance of securities or other means [2]. In

addition, off-campus training institutions must not recruit preschool children, must not teach primary and secondary school subject content in advance, and must not organize subject training for primary and secondary school students during winter and summer vacations or statutory holidays.

Furthermore, the supervision requires government departments to increase supervision of off-campus training institutions to ensure their legal and compliant operations, and strictly investigate and deal with violations of laws and regulations [3]. The introduction of the double reduction policy aims to protect the physical and mental health of students, so that they have more time to participate in sports, art and social practice activities and develop in an all-round way [4]. At the same time, it is also hoped that by reducing excessive tuition and training, education will return to its essence and pay more attention to cultivating students' comprehensive quality.

Correlation analysis is a pivotal statistical technique, which can offer insights into the linear relationship between two quantitative variables. Correlation analysis is instrumental for researchers, analysts, and businesses in making predictions, pinpointing potential causes, and revealing concealed patterns in data. Central to this is the correlation coefficient, a value between -1 and 1 that signifies the relationship's intensity and direction [5]. When the coefficient is close to 1 or -1, it indicates a strong positive or negative relationship, respectively, while a value near 0 suggests a weak or nonexistent relationship.

Indeed, correlations can be divided into positive, negative, and zero, denoting the nature of the association between variables. However, one must tread cautiously, acknowledging that correlation doesn't equate to causation. There is always the possibility of external influences or mere coincidences leading to observed correlations. For calculating these correlations, the most utilized method is Pearson's correlation coefficient for linear associations [6]. In cases of non-linear relationships, alternatives like Spearman's rank correlation or Kendall's tau are preferred.

Correlation analysis is instrumental in informing instructional strategies, developing educational policies, and identifying areas that may benefit from targeted interventions [7]. However, it is crucial to note that correlation does not imply causation; therefore, while it reveals patterns and associations, it does not establish a cause-and-effect relationship between the variables studied [8]. Nonetheless, the insights derived from correlation analysis in education lay the groundwork for further research and experimentation.

In this paper, we adopt above mentioned analysis methods to systematically obtain the relationships between students achievement and family social status and investigate the complex interactions between these parameters. The reminder of this paper is arranged as an introduction about related correlation analysis methods and primary symbols that are used in the proposed model shows in Section 2. Indeed, we show our model framework and detail procedures in Section 3. Moreover, the analysis results and experimental comparisons are demonstrated in Section 4. Finally, Section 5 concludes our proposed model and provide several possible improvements mechanisms for future research.

2 Backgrounds

In this section, we initially introduce several existing correlation analysis methods and explain its principles. Subsequently, the primary parameters of our proposed model in provided.

2.1 Related Works

Traditional methods including Pearson correlation coefficient and Spearman rank correlation laid the groundwork for understanding linear and monotonic relationships between different variables [9]. Its applications span diverse fields especifically for finance, where it aids in portfolio diversification, as for the biology and medicine area, where analysis can assist decipher relationships between biological markers.

The realm of time series analysis also leverages correlation to understand autocorrelation in data. With the continuous investigation on big data and establishment of databases, modern techniques have emerged to analyze vast datasets, though it is imperative to recognize the inherent limitations of correlation, notably that it does not equate to causation [10].

2.2 Primary Parameters Description

Following Table 1 describes the primary used parameters and corresponding explanations in the proposed model.

Table 1. Primary parameters descrption.

Parameters symbol	Utilization description
P	Input data of students achievement
S	Social status parameters
C	Pearson correlation analysis
R	Regression prediction results

3 Model Framework

After kowing the basic backgrounds, we subsequently demonstrate our proposed model through explaining the detail execution procedures and general model framework.

3.1 Model Structure

Initially, we introduce the model framework in following Figure 1 demonstration with specific components. According to our model framework, we can significantly observe that we primarily concentrate the relationship between student achievement and their family social status. However, there exists confounding factors to influence the correlation analysis, which is mainly consisted by the shcool, teachers and student aspects. Therefore, these variables are also quantified by the proposed model, which can enhance the analysis accuracy.

3.2 Model Procedures

Above all, we introduce the model execution sequences and deatil utilizations of these modules in following items.
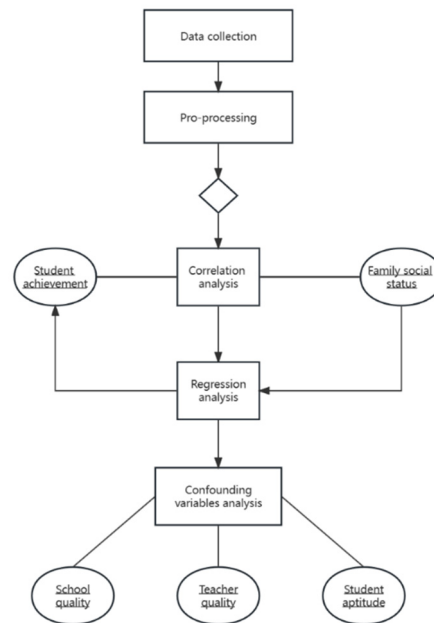

Fig. 1. Architecture of proposed analysis model.

- Objective Definition includes the primary Objective is aimed to determine the correlation between student achievement and family social status under the "Double Reduction" policy. Secondary Objective will identify any other factors that might influence student achievement in the analysis purpose.

- Data Collection includes student achievement data will collect data on student grades, test scores, and other academic achievements. Family social status data will gather the data on family income, parental education levels, occupation, and other relevant indicators of social status. Policy data will extract the time-period of the "Double Reduction" policy and how it has been implemented in the region of study.

- Data Cleaning will remove any missing or inconsistent data. Normalize scores if they come from different grading systems. Categorize family social status into groups for easier analysis if necessary.

- Exploratory Data Analysis (EDA): Use descriptive statistics to understand the distribution of student achievements and family social statuses. Visualize the data using histograms, and box plots to identify any patterns or outliers.

- Correlation Analysis will use Pearson correlation coefficient to measure the linear relationship between student achievement and family social status. A positive coefficient

indicates that as one variable increases, the other also does. A negative coefficient indicates the opposite.

- Regression Analysis will predict student achievement based on family social status through using regression analysis. This method will allow model to quantify the relationship and understand how much variation in student achievement can be explained by family social status.

- Control for Confounding Variables, It is possible that other factors including school quality, teacher quality, or individual student aptitude, could influence student achievement. Use multivariate regression to control for these variables and isolate the effect of family social status.

3.3 Methodology Principles

As for correlation analysis, we adopt Pearson correlation coefficient, which is a statistic that measures the strength and direction of the linear relationship between two quantitative variables. Its calculation formula is shown in following Equation 1 where the x and y represent the student achievement (PISA scores in our analysis) and their family social status.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \tag{1}$$

4 Experimental Results and Analysis

4.1 Experimental Setups

Above all, wo adopt the dataset of PISA (Programme for International Student Assessment), which is an international student assessment program led by the Organisation for Economic Co-operation and Development (OECD). The program takes place every three years, assesses 15-year-olds' knowledge and skills in three areas: reading, math and science. The average PISA score is around 500 points, and the standard deviation is around 100 points. The score of family social is divided into 3 classes including low(<20000 dollars/year), middle(20000 dollars/year ~ 40000 dollars/year) and high (>40000 dollars/year). We utilize the data from 2000 to 2020 of China mainland.

4.2 Correlation Analysis

Initially, we demonstrate the general description about the used data in the PISA dataset in following Table 2.

Table 2. Dataset desription.

Analysis Participants	Statistics Ratio
Students	Male(48.2%), Female(51.8%).
Ages	13(15.8%), 14(27.7%), 15(28.9%), 16(27.6%)
Student Achievement	Low(35.2%), Middle(46.5%), High(18.3%)
Family Income	Low(29.6%), Middle(57.1%), High(13.3%)
Parents Education	Primary School(10.4%), High School(37.6%), Colleges(48.0%)

Subsequently, we calculate the correlation analysis by using the PISA data from 2000 to 2020 in the china four regions and following Figure 2 shows the detail correlation results.
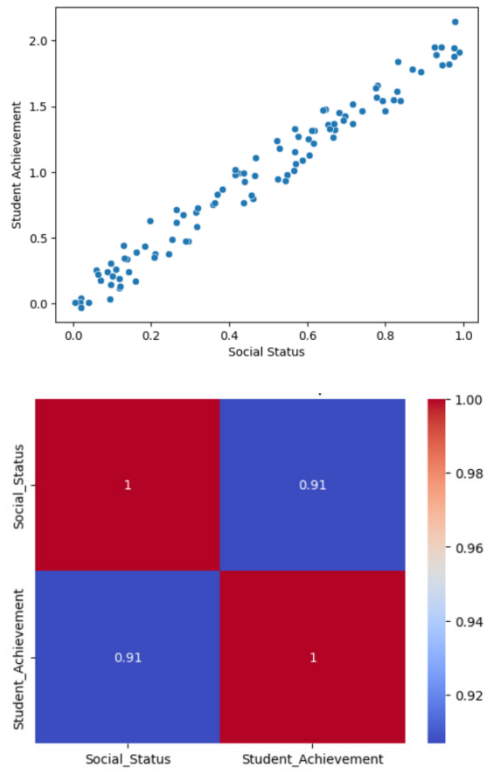

Fig. 2. Correlation Analysis Results.

Additionally, we estimate the regression of the analysis results and following Figure 3 shows the results of non-linear regression. From our regression prediction within high social status families, we can observe the higher status can lead higher achievement of students.
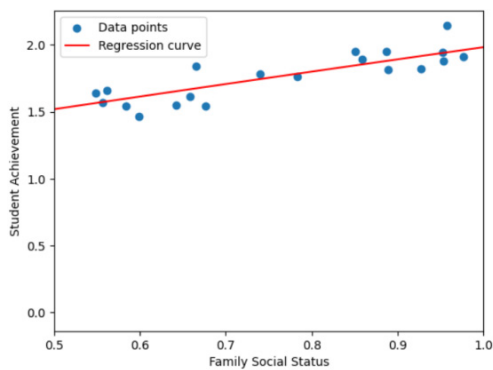

Fig. 3. Regression Analysis Results of High Social Status.

From our simulation analysis, we can significantly observe the student achievement has a strong positive relationship with their family status by removing the confounding variables. According our analysis, we can provide students extra assistance for low family social statues students to guarantee each student can get fairness education environment. Additionally, the detail analysis values can assist the school to measure the student achievement and deploy suitable education costs for each student.

5 Conclusion

In conclusion, our analysis indicates that under the "Double Reduction" policy, family social status remains a significant determinant of student achievement. This indicates that in addition to reducing homework and private tutoring, policymakers need to address other underlying factors contributing to educational inequality to ensure equal opportunities for all students. As for future inestigation, the utilization of learning methods can extract the essential information from complex data and improve the correlation analysis results of proposed model.

References

[1] Guo, Yihan. "The Current Impact of the Double Reduction Policy." 2021 International Conference on Education, Language and Art (ICELA 2021). Atlantis Press, (2022).

[2] Ibbotson, Roger G., and Jay R. Ritter. "Initial public offerings." Handbooks in operations research and management science 9: 993-1016, (1995).

[3] Li, Jia, et al. "The impact of the double reduction policy." 2021 International Conference on Education, Language and Art (ICELA 2021). Atlantis Press, (2022).

[4] Chen, Gaoyu, et al. "Parents' educational anxiety under the "double reduction" policy based on the family and students' personal factors." Psychology research and behavior management: 2067-2082, (2022).

[5] Gogtay, Nithya J., and Urmila M. Thatte. "Principles of correlation analysis." Journal of the Association of Physicians of India 65.3: 78-81, (2017).

[6] Hardoon, David R., Sandor Szedmak, and John Shawe-Taylor. "Canonical correlation analysis: An overview with application to learning methods." Neural computation 16.12: 2639-2664, (2004).

[7] Karim, Ibnu Nur. "CORRELATION ANALYSIS: Are Religious Scavengers More Motivated?." AKADEMIK: Jurnal Mahasiswa Humanis 1.2: 59-67, (2021).

[8] Ariastuti, Melani Dwi, and Achmad Yudi Wahyudin. "Exploring academic performance and learning style of undergraduate students in English Education program." Journal of English Language Teaching and Learning 3.1: 67-73, (2022).

[9] Lindley, D. V. "Regression and correlation analysis." Time series and statistics. London: Palgrave Macmillan UK. 237-243, (1990).

[10] Qu, Guangzhi, Salim Hariri, and Mazin Yousif. "A new dependency and correlation analysis for features." IEEE Transactions on Knowledge and Data Engineering 17.9: 1199-1207, (2005).