

Prediction of student attendance and its scheduling optimization in higher education based on multivariate regression analysis

Ying Jin¹, 

¹ Art Department, Fushun Vocational Technology College, Fushun, Liaoning, 113122, China

ABSTRACT

In order to improve the attendance rate of students and optimize the quality of teaching, this paper proposes a method of predicting the attendance rate of students in colleges and universities based on multivariate regression analysis. Firstly, we obtain the factors affecting students' attendance rate through sample survey and conduct correlation analysis, and then summarize and refine the three dimensions of students, teachers and schools. The above dimensions are used as independent variables to construct regression equations, and the regression equations are used to predict the attendance rate of students, so that teaching managers can optimize the management. The analysis found that the factors such as the college to which the truant students belonged, the reason for truancy, and the grade level showed diversity and complexity. Overall male students have more truancy rates than female students, and lecturers with higher titles have lower truancy rates. Regression modeling and prediction of truancy rate found that the prediction results are closer to the real results. Therefore, the method of this paper can be combined to optimize and adjust the attendance rate from the aspects of regulations, work allocation, teaching management and ideological education.

Keywords: multivariate regression analysis, class attendance prediction, class attendance scheduling optimization, correlation analysis

1. Introduction

The attendance rate of students is an important indicator of students' classroom learning. At present, the phenomenon of low student attendance is common in colleges and universities, which not only seriously affects the cultivation of students in schools, but also affects the normal order of education and teaching in schools [1]. The low attendance rate of students will seriously affect the quality of talent training. For students themselves, low attendance rate will affect students' knowledge acquisition and self-development [2-3]. As for teachers, low attendance of students also makes

 Corresponding author.

E-mail address: 15242746595@163.com (Y. Jin).

Received 20 February 2024; Revised 25 April 2024; Accepted 20 November 2024; Published Online 15 April 2025.

DOI: [10.61091/jcmcc127a-320](https://doi.org/10.61091/jcmcc127a-320)

© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

teachers less and less motivated to attend classes, which in turn makes the attendance of students even lower, resulting in a vicious circle, and finally leading to the waste of school resources [4-5].

There are various reasons for the low attendance rate of students. On the social side, in the form of employment, family education and other reasons, college students do not actively and consciously study, resulting in low class attendance; on the teachers' side, some teachers do not have a serious teaching attitude and poor teaching ability, which makes it difficult for students to understand and master the knowledge points, and then they do not want to continue to come to class [6]; on the curricula side, there are more theoretical courses than practical courses, which do not meet the real needs of students, and there are too many courses, and students lack time for independent learning, which also leads to the waste of school resources. Students lack the time for independent learning, and are also prone to the phenomenon of anorexia [7-8]; In terms of course assessment, the assessment method is single, mostly course paper assessment, such as paper checking and low quality requirements, resulting in students muddling through, in addition to the theoretical knowledge based on books for the preparation of question papers, students only need to recite the main knowledge points before the final exam in a surprise manner to pass the exam, the lack of course value embodied, resulting in students attaching little importance to the degree of [9-11]; in terms of the teaching mode, the students are not adapted to the new teaching mode, students do not adapt to the introduction of new teaching tools, online and offline teaching format changes, making the students learning effect decline, and then produce the idea that the classroom is no longer useful [12]; school, leadership figures, school philosophy, academic design, teaching facilities, etc. indirectly affect the student attendance rate [13-16]; Students' own reasons, such as physical state, part-time job, student work, not realizing the importance of the course, obsessed with online games, etc. [17-18]. In addition, there are certain patterns in the attendance rate, which is higher in colleges of science and engineering than in colleges of arts and sciences, the level of grade is inversely proportional to the attendance rate of students, and the attendance rate of elective classrooms is lower than that of specialized classrooms [19-20]. These reasons and patterns are important references for classroom teaching improvement. However, due to the many factors affecting students' attendance rate, how to improve it effectively is a debatable issue. It is necessary to analyze these factors and make predictions in order to make reference to the improvement of classroom teaching strategies for teachers, which can identify the problems of the school itself and target them for improvement so as to improve the effectiveness of classroom teaching.

This paper first investigated the reality of class attendance in college A. Statistical analysis was carried out with different colleges, grades, truancy orientation, reasons for truancy and influencing factors. Based on the analysis results, variables at the levels of students, teachers and schools were determined. The dimensions of students' gender, titles and nature of teachers, and school management were selected for correlation analysis, and then multiple linear regression equations were constructed between them and the class rate according to the factors of students, teachers, and schools, and the relevant parameters were derived from solving the model, so as to predict the attendance rate of students in colleges and universities. Several iterations of fitting were carried out while removing the outliers, and the predicted and real values were compared to reflect the predictive effect of the prediction model. Finally, the scheduling optimization strategy to improve the attendance rate is summarized.

2. Overview

In the study of student attendance, most scholars developed attendance management systems such as face recognition algorithms for intelligent identification to improve the efficiency and accuracy of attendance for student attendance in offline and distance education, e.g., literature [21] and [22]. Similarly, literature [23] designed an enterprise attendance tracking system using face recognition technology and added machine learning algorithms to predict employee attendance trends and optimize resource allocation. Although the study was not practically applied to classroom attendance,

its authors also mentioned in the article that the system offers potential for predicting attendance in educational institutions. There are fewer studies that focus on classroom attendance prediction alone. Literature [24] constructed a model for predicting student attendance by extracting features from a student database and using a decision tree algorithm, which provides an understanding of student attendance patterns. Literature [25] predicted the long and short-term attendance of students with autism by using long and short-term memory and multilayer perceptron algorithms through the students' previous attendance, and pointed out that the prediction algorithm is better than other machine learning algorithms. These algorithms are all hotspots in the current prediction direction, but the study is only the prediction of the results, and no explanation is given for the specific factors affecting the attendance rate, which is not conducive to the rectification of classroom teaching. Multivariate regression analysis can be used to predict the future value of the dependent variable, to determine which independent variables have the greatest impact on the dependent variable, and to help decision makers better understand the relationship between the variables [26]. This method can be a good solution to the above problems and assist teachers in making appropriate classroom decisions and plans.

3. Analysis of the current status of attendance at university A

University attendance rate is an important index to measure the learning culture and teaching status of the university, which can not only reflect the students' learning status, such as motivation and sense of responsibility for learning, but also reflect the classroom teaching situation, as well as the teaching management level of the university. In this paper, the attendance rate data of the spring semester of 2023 in college A are counted, and the current status of the attendance rate of this university is as follows.

3.1. Attendance rates vary by college

Based on the attendance of 72 classes in seven colleges of the university sampled during the spring semester of 2023, the attendance rates are shown in Table 1. It can be seen that the total average class attendance rate of the seven colleges is 78.7%, but there are disparities in the student attendance rates of different colleges, including two colleges with an average student attendance rate of more than 80%, namely, the College of Architecture and the College of Mechanical and Transportation Engineering. There are 5 colleges with attendance rates below 80%, namely, College of Economics and Trade, College of Journalism and Communication, College of Law, College of Electrical and Information Engineering, and College of Finance and Statistics. Among the seven colleges, the College of Architecture has the highest attendance rate of 82.39% and the College of Finance and Statistics has the lowest attendance rate of 74.76%. This indicates that there are differences in the attendance rates of students in different colleges.

Table 1. The attendance statistic of seven schools

School's name	Sampling number	Students that should have attended	Students that actually attend	Average attendance
School of economics and trade	7	338	266	78.7%
Architecture school	12	477	393	82.39%
Journalism and communication school	9	284	224	78.87%
Machinery and delivery engineering school	8	432	348	80.56%
Law school	13	822	653	79.44%
Electrical and information	17	1316	1019	77.43%

engineering school				
Financial and statistics school	6	206	154	74.76%
Overall	72	3875	3057	78.7%

3.2. Attendance rate decreases with grade levels

Table 2 shows the statistics of the attendance rate from the first year to the fourth year of university in this university. As can be seen from Table 2, there is a close relationship between the attendance rate and the grade level of college students, with an average attendance rate of 90.67% in the first year of college, 83.18% in the second year of college, 77.12% in the third year of college, and less than 60% in the fourth year of college, which is only 56.20%. The lower the grade of college students, the higher the attendance rate, while the higher the grade of college students, the lower the attendance rate. This suggests that college attendance rates plummet as the grade level of college students increases.

Table 2. The attendance statistic of different grades

Grade	Students that should have attended	Students that actually attend	Average attendance	Highest attendance	Lowest attendance
Freshman	1126	1021	90.67%	100%	83.42%
Sophomore	868	722	83.18%	100%	62.15%
Junior	1228	947	77.12%	97.92%	45.53%
Senior	653	367	56.20%	82.43%	18.44%
Total	3875	3057	78.89%		

3.3. Diversity of Truancy Destinations

The statistics of college students' skipping classes are shown in Table 3. about 44.38%% of college students skipped classes to go to the library to study, and another 13.20% of students skipped classes to participate in club activities. About 40% of the students skipped classes to do something not related to study, such as 14.30% of the students play games on the Internet, watch movies, chatting on the Internet, etc., 6.72% of the students go out to participate in friends' gatherings, 5.13% of the students sleep, 4.40% of the students go home in case of family emergencies that are too late to ask for leave, 3.42% of the students go shopping, and 8.44% of the students do other things that have nothing to do with study. The other 8.44% of students do other things not related to study. It can be seen that the low attendance rate of university is mainly caused by college students' skipping classes, but college students' skipping classes or low attendance rate is not mainly due to students' lack of love for learning.

Table 3. The statistics of where the students have gone after skipping

Ends	Number of students who skip the class	Proportion
Self-taught	363	44.38%
Surfing the internet	117	14.30%
Club activities	108	13.20%
Friends party	55	6.72%
Sleeping	42	5.13%
Emergency	36	4.40%
Shopping	28	3.42%
Other	69	8.44%
Sum up	818	100%

3.4. Primary and secondary reasons for truancy

According to the investigation of the reasons for truancy, the four main factors of students, teachers, schools and society were analyzed, and the results of the survey are shown in Table 4. The reasons for college students' truancy are categorized into four main factors: student factor, teacher factor, school factor and social factor. Among them: student factor is 35.82%, which is the main factor of students' truancy. Teacher factor is 35.21%, which also has a direct influence on students' truancy. The school factor is 20.54%, and the high or low rate of student truancy directly reflects the teaching quality and management level of a university. The social factor is 8.44%, which also has a certain influence on students' truancy.

Table 4. The survey of the reasons for skipping classes

Reasons	Count	Proportion
Student	293	35.82%
Teacher	288	35.21%
School	168	20.54%
Society	69	8.44%
Total	818	100%

3.5. Analysis of Factors Affecting Attendance Rates

3.5.1. Analysis of student factors. Many students have not set up lofty ideals, lack a sense of social responsibility, and have unclear learning objectives, unilaterally believing that learning is for the sake of getting good grades and finding a good job, and that learning is quite utilitarian, with selective attitudes and practices towards course learning.

Influenced by the employment situation, personal wishes and parents' expectations, many college students are blind in their choice of majors. After entering the university, only gradually understand the characteristics of the specialty studied and the focus of professional learning, but found that the gap between the specialty studied and their own expectations is too large, thus appearing to be uninterested in professional learning, learning enthusiasm has been seriously affected by the situation [27].

Many college students from childhood to accept the examination education, learning under the strict management of parents, teachers, emphasizing rote memorization, independent learning ability is not strong, into a more relaxed university learning environment, can not timely self-adjustment, change the learning concept and learning methods.

At present, many students in colleges and universities due to the face of all kinds of pressure, in order to cope with the enormous pressure, will weigh the pros and cons, take the rational way of skipping classes, or hidden skipping that is the use of class time to do their own opinion of the most valuable things, or explicit skipping that should be to the class really do not come to listen to the class, resulting in a low rate of attendance.

3.5.2. Analysis of teacher factors. Classroom teaching requires the participation and cooperation of both teachers and students, and the teaching level of teachers has a direct impact on student attendance. The low teaching level of teachers has a negative impact on the attendance rate of students.

Some teachers have been using the old version of the textbook, the teaching content is old, monotonous, irrational structure, seriously out of touch with the social practice, and seriously out of touch with the students' future work, the students think that there is no practical value in learning these outdated theories, which seriously affects the students' motivation to learn.

Many teachers talk to the end, students passively accept, lack of interaction between students and teachers, do not inspire students to think, guide students to think. Students are completely passive in

listening to the lessons, and feel that this teaching method makes them suffer, and thus they are often unwilling to come to the lessons and skip the lessons. Some teachers often do not prepare their lessons carefully, do not explain key points and difficulties, do not ask questions for students to think about, do not pay attention to classroom discipline, do not pay attention to classroom management, and encourage students' truanting behavior.

3.5.3. Analysis of school factors. Although many universities emphasize that the school is a talent training institution and that teaching is the central work of the school and the primary academic responsibility of the teachers, in practice, the level and quality of teaching are lightly regarded, and the central position of teaching is not well implemented.

Many universities in the curriculum of the theoretical courses are too much, experimental courses and internship courses are too little, professional courses and social practice is seriously out of touch with the problem. This not only aggravates the normal learning burden of students, but also seriously affects the quality of teaching and learning, resulting in students feeling too much pressure to learn the course and a decline in the attendance rate.

University course grades are mainly based on the results of the written test at the end of the course, and the written test focuses on the examination of memorized knowledge, so students can get satisfactory scores as long as they study suddenly for a few days, which makes most of the students improve their course grades by improving their written test scores through speculation, and they think that there is no need to attend the lectures, which aggravates the skipping behavior of the students.

4. Multiple linear regression analysis

Univariate linear regression limits the independent variable affecting a variable to a single variable, but this is not easy to do in most practical problems, and thus the application of correlation statistical analysis to solve practical problems usually requires the use of a mathematical model with a more comprehensive coverage, in which the effects of two or more explanatory variables are estimated separately within the model, which is the main mathematical idea of multivariate statistics.

4.1. Multiple linear regression correlation theory

Let y be a random variable that can be observed and recorded, and it is obviously influenced by $m-1$ a non-random factor $x_1, x_2, \dots, x_{m-1}$ and a random factor ε . There is a relationship between the non-random, random factor and y based on mathematical and statistical theory, if y and $x_1, x_2, \dots, x_{m-1}$ have the following linearly related relationship equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{m-1} x_{m-1} + \varepsilon \quad (1)$$

Where: $\beta_0, \beta_1, \beta_2, \dots, \beta_{m-1}$ is the coefficient of the independent variable. The ε mean is zero and the variance is $\sigma^2 > 0$. It is usually assumed $\varepsilon \sim N(0, \sigma^2)$. That the above model is a multiple regression model, where y is the dependent variable and $x_1, x_2, \dots, x_{m-1}$ is the independent variable.

We first need to make reasonable parameter estimates for the unknown coefficients $\beta_0, \beta_1, \beta_2, \dots, \beta_{m-1}$, so we need to make $n(n \geq p)$ observations of them first and record them, thus obtaining n sets of data:

$$(x_{i1}, x_{i2}, \dots, x_{im-1}; y_i) (i=1, 2, \dots, n) \quad (2)$$

They should satisfy $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{m-1} x_{m-1} + \varepsilon$, i.e., have:

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{m-1} x_{m-1} + \varepsilon_1 \\ y = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_{m-1} x_{2m-1} + \varepsilon_2 \\ \vdots \\ y = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_{m-1} x_{nm-1} + \varepsilon_n \end{cases} \quad (3)$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent of each other and obey a $N(0, \sigma^2)$ -distribution.

Let:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, X = \begin{bmatrix} 1, x_{1,1} \cdots x_{1,m-1} \\ 1, x_{2,1} \cdots x_{2,m-1} \\ \vdots \\ 1, x_{n,1} \cdots x_{n,m-1} \end{bmatrix}_{n \times m}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{m-1} \end{bmatrix}_{m \times 1}, \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \quad (4)$$

Then the system of multivariate equations can be abbreviated to the following form:

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I_n) \end{cases} \quad (5)$$

where Y is the observed value and X is the variable observation matrix, which are recorded from the previous data observations, while X is assumed to be the variable matrix with full rank of columns, i.e., $r(X) = m$. where β is the column vector of the unknown parameter to be estimated. ε is the column vector of random errors that cannot be effectively observed and recorded. The mathematical formula obtained above is the expression of the matrix of the multiple linear regression model and is denoted as $(Y, X\beta, \sigma^2 I_n)$.

The central problems to be solved for the above linear model $(Y, X\beta, \sigma^2 I_n)$ are [28]:

- 1) Effective estimation of parameters β and σ^2 to establish the linkage equation between y and $x_1, x_2, \dots, x_{m-1}$.
- 2) The assumptions of the proposed linear model and the assumptions of parameter β can be effectively tested.
- 3) Reasonable prediction of the dependent variable y is made, while the independent variables are effectively controlled and recorded.

4.2. Estimation of coefficients for multiple linear regression

As in the case of simple regression models, the common estimation of the linear coefficient $\beta_0, \beta_1, \beta_2, \dots, \beta_{m-1}$ in complex models is still the least squares method [29], denoted as:

$$Q(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2 \quad (6)$$

Find its minimum value at point $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$, i.e:

$$Q(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) = \min_{\beta_0, \beta_1, \dots, \beta_k} Q(\beta_0, \beta_1, \dots, \beta_k) \quad (7)$$

Then $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ is the least squares estimate of $\beta_0, \beta_1, \dots, \beta_k$.

The linear equation for $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ is given by the necessary condition for the differential method for the extremum:

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2 = 0 (j = 0, 1, \dots, k) \quad (8)$$

The solution of the Solve this equation and represent it as a matrix:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (9)$$

where: $\hat{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$ and $X^T X$ is required to be invertible.

In this way, the empirical regression equation is obtained as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad (10)$$

As with the simple regression model, Equation σ^2 can be estimated, noting:

$$S_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = Q(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) \quad (11)$$

Then the unbiased estimate of σ^2 is:

$$\hat{\sigma}^2 = \frac{S_E}{n-k-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1} \quad (12)$$

4.3. Significance test of multiple linear regression

As in the case of one-dimensionality, the significance of the empirical equation requires a significance test for the existence of a linear relationship between y and x_i . Similar to the one-dimensional case, the original hypotheses can be formulated $H_0: \beta_0 = 0, \beta_1 = 1, \dots, \beta_k = 0$. Common statistics used to measure the goodness of fit between the empirical regression equation and the observed quantities are the compound correlation coefficient R and the goodness-of-fit coefficient, where:

$$R = \sqrt{\frac{U}{U + Q_e}}, R^2 = \frac{U}{U + Q_e} \quad (13)$$

4.4. Prediction by multiple linear regression

In multiple linear regression analysis, when the variance of the regression equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$ is statistically significant for the test, the prediction of the value of the dependent variable y can be achieved relatively easily by the variance of the regression equation in a similar way as in the case of linear regression with the same element.

First assume that the prediction point is $x_0 = (x_{01}, x_{02}, \dots, x_{0k})^T$, then:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \dots + \hat{\beta}_k x_{0k} \quad (14)$$

It's right:

$$E(y_0) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \dots + \beta_k x_{0k} \quad (15)$$

The point estimate is also the right one:

$$y_0 = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \dots + \beta_k x_{0k} + \varepsilon_0 \left(\varepsilon_0 \sim N(0, \sigma^2) \right) \quad (16)$$

Of the point estimate. And the statistics can be proved:

$$t = \frac{y_0 - \hat{y}_0}{\hat{\sigma}^*} \sim t(n-p-1) \quad (17)$$

Among them:

$$\hat{\sigma}^{*2} = \frac{SSE}{n-p-1}, \Delta = \sqrt{1 + \frac{1}{n} + \sum_{i=1}^p \sum_{j=1}^p (x_{0i} - \bar{x}_i)(x_{0j} - \bar{x}_j) c_{ij}} \quad (18)$$

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki} (i=1, 2, \dots, p) \quad (19)$$

Thus, the marginal error of the point prediction is $\pm t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}\Delta$, i.e., the interval prediction at x_0 is:

$$\left(\hat{y}_0 - t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}^* \Delta + t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}^* \Delta \right) \quad (20)$$

To wit:

$$P \left\{ \hat{y}_0 - t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}^* \Delta + t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}^* \Delta \right\} \geq 1 - \alpha \quad (21)$$

When n is larger and $x_{0i} \approx \bar{x}_i (i=1, 2, \dots, p)$, $\Delta = 1$ can be taken to simplify the calculation.

5. Analysis of factors affecting school attendance and multiple regression prediction

5.1. Correlation analysis of factors affecting attendance at classes

Selected data on school administration, classroom teachers, and students were collected for correlation analysis in college B. Factors affecting class attendance were analyzed through the one-way analysis method, and the time span of the information included three semesters from 2022 to 2023.

5.1.1. Gender of students. Table 5 shows the correlation analysis between attendance rate and students' gender. From Table 5, it can be seen that the gender of the students has a greater effect on the attendance rate ($P < 0.001$), and the attendance rate of female students (64.29%) is significantly higher than that of male students (59.24%), which is a significant difference. This may be determined by the gender difference between men and women, where women are generally more patient and responsible, while male university students are generally more active.

Table 5. The correlation between the attendance and the gender of students

Gender	Sample size	Average attendance
Male	498	59.24%
Female	1036	64.29%
	F=126.921	P<0.001

5.1.2. Teachers in the classroom. Table 6 shows the correlation analysis between the attendance rate and the title of the instructor teaching the course. Teacher title had a significant effect on the attendance rate ($p < 0.001$), and the attendance rate of face-to-face courses for teachers with no title (59.43%) was significantly lower than that of teachers who had earned the title. A t-test analysis of the differences revealed that this difference existed between teachers with and without titles, while there was no significant difference between the means of the attendance rates of courses among teachers with different titles, i.e., the level of title did not have a significant effect on the attendance rate ($P > 0.05$).

Teachers with titles are mostly on-campus full-time teachers and off-campus college teachers with teaching experience and work experience as well as in-service personnel from enterprises, who are responsible and reputable in general. The non-titled personnel are mostly graduate students and new

teachers, young and inexperienced in teaching, not too familiar with the content of the textbook, coupled with the fact that graduate students mainly aim to make money, lack of professionalism and dedication, and their own studies are heavy and their energy is limited. It is inevitable that there are differences in the quality of teaching between the two.

Table 6. The correlation analysis between attendance and teacher's title

Title	Sample size	Average attendance
Senior	266	67.21%
Deputy senior	1935	64.33%
Intermediate	7498	64.47%
None	6474	59.43%
	F=63.125	P<0.001

The data in Table 7 shows the correlation analysis between the attendance rate and the nature of the instructor teaching the course, and there is a significant difference between the attendance rates of instructors of different natures ($p<0.001$). The mean attendance rate of face-to-face courses for full-time and permanent faculty was greater than 65%, which was significantly higher than that of part-time faculty (60.43%). This is mainly due to the fact that full-time and permanent faculty have a sense of job identity and career, and operate teaching as a lifelong business. Part-time teachers, on the other hand, mainly consider their own interests and are mostly short-term earners, and are obviously lacking in sense of enterprise and responsibility.

Table 7. The correlation between attendance and the teacher types

Types	Sample size	Average attendance
Full-time	5076	65.53%
Tenured	1261	65.12%
Part-time	9977	60.43%
	F=97.544	P<0.001

5.1.3. School management. Table 8 shows the correlation analysis between the attendance rate and semester order. It can be seen that semester sequence has a significant effect on attendance ($p<0.001$), with an upward trend in attendance in face-to-face tutoring courses from fall 2022 to fall 2023 semester, with an average increase of nearly twenty percentage points in attendance over the three semesters.

This is largely related to the emphasis and increased management of class attendance at University B in recent years. Since 2022, University B has implemented a random inspection system for class attendance rates, and since the fall semester of 2022, it has fully implemented the "daily reporting system for class attendance rates" and "weekly reporting system for disciplinary violations", and since the spring of 2023, it has been supplemented by a classroom teaching questionnaire survey, and the results of the inspection have been incorporated into the "Regulations on the Handling of Teaching Accidents", which are linked to the appointment and evaluation of teachers' professional titles, as well as the assessment management of tutors and colleges, and the teaching management system has been increasingly improved. In the fall semester of 2023, University B began to implement the unified management of public courses and the small class teaching system for language and art courses again, and the management of course teaching was refined. This shows that the emphasis on face-to-face tutorial teaching and the establishment and improvement of the system are the cornerstones of improving attendance and teaching quality.

Table 8. Correlation analysis of attendance and semester order

Semester	Sample size	Average attendance
2022 fall	4231	51.32%
2022 spring	6373	62.91%
2023fall	6048	69.72%
	F=873.339	P<0.001

5.2. Multiple regression-based prediction of class attendance rates

From this previous analysis, it is easy to find that the factors affecting student attendance can be mainly categorized into three levels: students, teachers and schools. Therefore, in this section, these three variables will be used to establish a multiple regression model in order to predict the student attendance rate, which is convenient for teaching and researchers to carry out scheduling optimization.

5.2.1. Research ideas. In the study of the prediction of class attendance rate, a multiple linear regression prediction model was established based on the data of University B, with the dependent variable of the model being the attendance rate (A) and the independent variable being the potential influencing factors at the three levels extracted above, namely, student quality index y_1 (weighted by learning interest, academic level, stress management, learning goals, etc.), teacher level y_2 (weighted by academic level, teaching innovation and teaching interaction), school management y_3 (considering the degree of teaching attention, curriculum setting, assessment methods, etc.). Grabbing the daily attendance rate of three semesters of University B from 2022 to 2023 as well as the related data of students, teachers and school administration, the Matlab software was used to calculate the

values of y_1 , y_2 and y_3 of the three independent variables, analyze the correlation between A and y_1 , y_2 and y_3 , and if the correlation is valid, the linear regression was used to derive the prediction model.

5.2.2. Analyze the correlation between attendance (A) and y_1 , y_2 , y_3 . Using Matlab software to draw a scatter plot of the attendance rate versus y_1 , y_2 and y_3 , as shown in Figure 1. It can be seen that the attendance rate is roughly in an upward trend with the increase in student quality, teacher level and school management performance, and the data points are roughly located on both sides of a straight line, which can be subjected to linear regression.

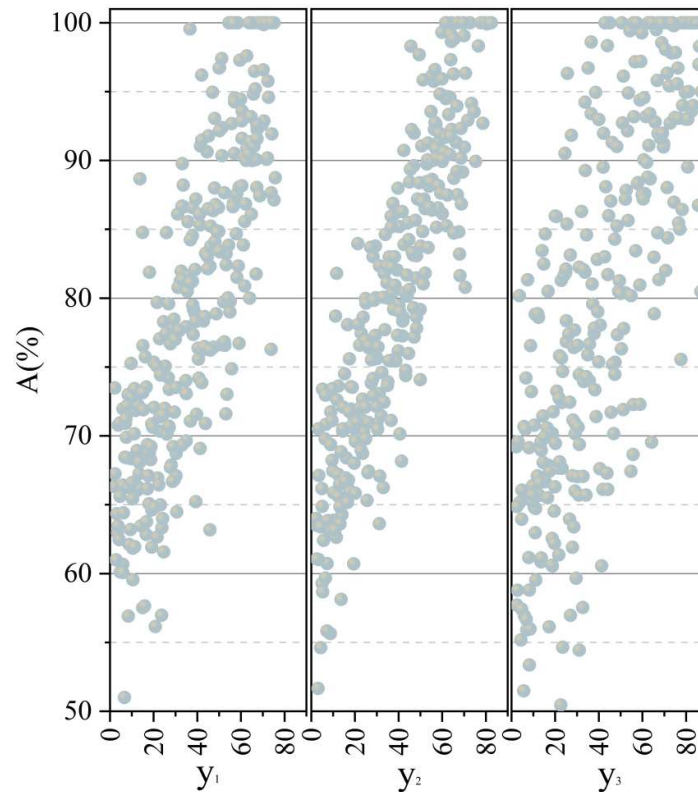


Fig. 1. The scatter point diagram of attendance

5.2.3. Modeling multiple linear regression equations. y_1 , y_2 , y_3 represent the quality of students, teacher level and school management level, respectively, A represents the value of the attendance rate, β_i ($i = 0, 1, 2, 3$) is the regression coefficient, which establishes the linear regression equation as:

$$A = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 \quad (22)$$

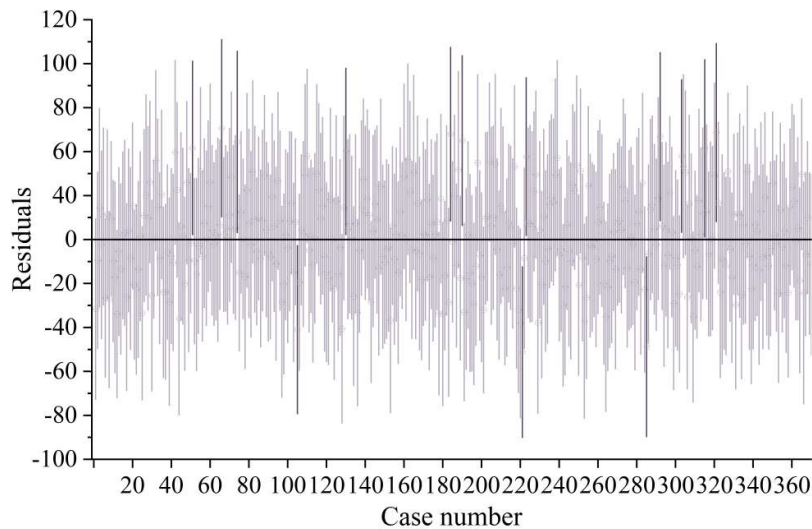
5.2.4. Model solving. The first linear regression was done using the regress function in Matlab to determine the coefficients of the linear regression equation. The specific data obtained (including regression coefficients, parameter confidence intervals, correlation coefficient R^2 , F values and their corresponding probabilities P) are shown in Table 9.

From the data in Table 9, the multiple linear regression equation was obtained as: $A = 12.5113 + 0.6239y_1 - 0.1513y_2 + 1.2721y_3$. In addition, Table R^2 shows the correlation coefficients, F is the value of the statistic, and P is the probability corresponding to the statistic. The goodness of fit of R^2 is 72.98%, and the closer R^2 is to 1 indicates a better fit, which means that the first fit is unsatisfactory.

Table 9. The first linear regression data

Regression coefficient	Regression coefficient estimation	The confidence interval of the regression coefficient
β_0	12.5113	[5.3731,19.6755]
β_1	0.6239	[0.4629,0.7746]
β_2	-0.1513	[-0.5911,0.2979]
β_3	1.2721	[0.4369,2.1035]
$R^2=0.7298, F=351.043, P=0.000, S^2=344.6321$		

The residual analysis is shown in Figure 2. In order to obtain the optimal multiple linear regression model, this paper eliminates the anomalies in Fig. 2, and then performs the 2nd fitting, and so on, a total of 12 times to eliminate the anomalies and then fit the operation, and the 13th time, we found that there are fewer anomalies, and the increase in the goodness-of-fit is not obvious, so we take the multiple linear regression equations obtained from the 12th time as the final solution. The regression equation obtained by removing outliers and then regressing is $A = 7.821 + 0.3451y_1 + 0.0506y_2 + 0.8641y_3$.

**Fig. 2.** Residual difference analysis

5.2.5. Model testing. In order to better verify the multiple linear regression model established above, the attendance rate data of a certain month is selected for prediction, and the prediction results are shown in Figure 3. It can be seen that the prediction accuracy of the model is good, the trend of the attendance rate matches the actual value, which also shows that the multiple linear regression model fits the data better, which is conducive to the subsequent optimization of the attendance rate.

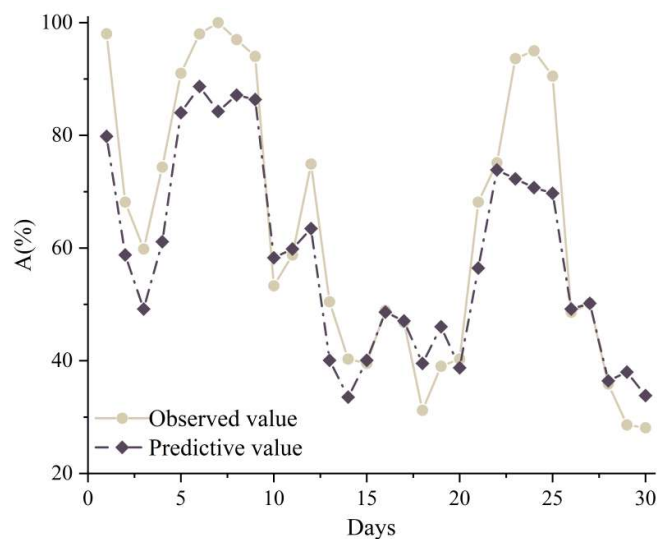


Fig. 3. The comparison between the estimated value and the observed value

5.3. Scheduling Optimization Strategies to Improve Attendance Rates

5.3.1. Implementing rules and regulations to improve enforcement. “No rule is not a circle”, the college should formulate rules and regulations specifically designed to improve the attendance of students, to strengthen the constraints on teachers and students, and once the implementation is determined, it must be fully implemented. The teacher’s constraints, can be linked to the evaluation of excellence, title evaluation. Constraints on students, must do the same, the scale can not be too wide, to cause students to pay attention to the students to experience the rigor of the system, do not allow students to have a loophole to drill. At the same time, in the implementation of regulations, there should be a specialized supervision department to improve the implementation of regulations.

5.3.2. Reasonable distribution of work and improvement of counselors' quality. At present, most of the student work management personnel in colleges and universities have a lot of affairs, high pressure, especially in the student management work in the first line of counselors, the work is complicated and miscellaneous, “the following a needle, above a thousand lines,” although the counselor to the management of the two ends: the Department of the Department and the Office of Student Affairs. But in practice, as long as any one of the matters related to students, are directly to the counselor task, which undoubtedly increased the counselor's work pressure, resulting in day-to-day handling of matters occupying most of the time of the counselor's work, so that the communication with the students to reduce. Counselors are the teachers who have the most contact with students, and the efficiency and quality of their work has a direct impact on students. Therefore, in order to improve the attendance rate and create an excellent academic style, colleges and universities must reasonably allocate the work, clarify the responsibilities, further improve the treatment and status of counselors, strengthen the training of counselors, and continuously improve the quality of counselors.

5.3.3. Strengthening teaching management and improving teaching quality. Attendance rate is directly related to the quality of classroom teaching, so teachers must, on the one hand, strictly implement the relevant teaching management rules and regulations and realize the attendance system. Students are disciplined through flexible roll-calls, such as before, during and after class, full or random roll-calls. Through the combination of the examination system and students' attendance, students with a high number of absences will be disqualified from the examination strictly according to the regulations. On the other hand, we should study and research hard, innovate teaching methods,

prepare lessons carefully, teach according to the students' abilities, and strive for high quality and efficiency in every lesson, so as to continuously improve the teaching efficiency and quality of teaching.

5.3.4. Emphasize ideological education and correct learning attitudes. Comprehensive understanding of the situation of students, accurately grasp the ideological dynamics of students, ideological education as an entry point, strengthen the concept of discipline, and strive to promote the transformation and enhancement of students' ideological awareness, correct students' learning attitudes, and really fundamentally solve the problem of low attendance rate. Ideological education must be carried out throughout the university career. The first year focuses on the cultivation of self-management ability, the ability to cultivate a positive mindset, interpersonal skills, and the ability to study and live independently, etc., and mainly carries out education on moral education, discipline, safety, sense of belonging, and education on learning and goals. The second year focuses on the cultivation of their professional learning ability, basic knowledge acquisition ability and practical ability, as well as the sense of innovation, etc., taking into account the characteristics of their majors. The third year focuses on internship and employment education to enhance students' internship and employment skills and establish a correct concept of internship and employment.

6. Conclusion

This study predicted the attendance rate of students in colleges and universities mainly through multivariate linear regression. The analysis of the attendance rate of college A revealed that the attendance rate of students varies from college to college and the attendance rate decreases gradually with the increase of grade. About 44.38% of the students skipped classes for the purpose of self-study and the reasons for skipping classes can be attributed to three main levels: students, teachers and schools. The attendance rate of female students is higher than that of male students, the attendance rate of students is higher in the classrooms of teachers with titles, and it is also found that school management helps to increase the attendance rate of students. The regression equation is derived through multiple regression analysis, and the prediction result of the attendance rate of this equation basically matches the real attendance rate. Through the above analysis, this paper proposes the following strategies to optimize the scheduling of class attendance rate: implement rules and regulations; reasonably allocate work; strengthen teaching management; and pay attention to ideological education.

References

- [1] Bergin, J., & Ferrara, L. (2019). How student attendance can improve institutional outcomes. *The Educause Review*.
- [2] Fadelelmoula, T. (2018). The impact of class attendance on student performance. *International Research Journal of Medicine and Medical Sciences*, 6(2), 47-49.
- [3] Rocque, M., Jennings, W. G., Piquero, A. R., Ozkan, T., & Farrington, D. P. (2017). The importance of school attendance: Findings from the Cambridge study in delinquent development on the life-course effects of truancy. *Crime & Delinquency*, 63(5), 592-612.
- [4] Cattán, S., Kamhöfer, D. A., Karlsson, M., & Nilsson, T. (2023). The long-term effects of student absence: Evidence from Sweden. *The Economic Journal*, 133(650), 888-903.
- [5] Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American educational research journal*, 55(6), 1163-1192.
- [6] Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, 56(2), 343-379.

- [7] Kwak, D. W., Sherwood, C., & Tang, K. K. (2019). Class attendance and learning outcome. *Empirical Economics*, 57, 177-203.
- [8] Moores, E., Birdi, G. K., & Higson, H. E. (2019). Determinants of university students' attendance. *Educational Research*, 61(4), 371-387.
- [9] Sloan, D., Manns, H., Mellor, A., & Jeffries, M. (2020). Factors influencing student non-attendance at formal teaching sessions. *Studies in Higher Education*, 45(11), 2203-2216.
- [10] Mokhtari, S., Nikzad, S., Mokhtari, S., Sabour, S., & Hosseini, S. (2021). Investigating the reasons for students' attendance in and absenteeism from lecture classes and educational planning to improve the situation. *Journal of education and Health promotion*, 10(1), 221.
- [11] James, T., & Seary, K. (2019). Why aren't they attending class like they are supposed to?: A review into students' perception of the value of class attendance. *Student Success*, 10(1), 115-129.
- [12] Haugom, E. (2022). The effect of changing from campus-based to digital teaching on student attendance: A case study of Norwegian business students. *Heliyon*, 8(11).
- [13] Bartanen, B. (2020). Principal Quality and Student Attendance. *Educational Researcher*, 49(2), 101-113.
- [14] Lavy, S., & Nixon, J. L. (2017). Applications, enrollment, attendance, and student performance in rebuilt school facilities: A case study. *International journal of construction education and research*, 13(2), 125-141.
- [15] Ashford-Hightower, K. A. (2018). An Analysis of the Relationship Between Teachers' Perceived Involvement in Secondary School Decision-Making and Concurrent Student Attendance Rate, Suspension Rate, Graduation Rate, and Dropout Rate. The University of Memphis.
- [16] Önder, E. (2017). Student absenteeism in secondary education, absenteeism-related school practices and recommended policies. *Egitim ve Bilim*, 42(190).
- [17] Miir, G., Rutakumwa, R., Nakiyingi-Miir, J., Nakuya, K., Musoke, S., Namakula, J., ... & Weiss, H. A. (2018). Menstrual health and school absenteeism among adolescent girls in Uganda (MENISCUS): a feasibility study. *BMC women's health*, 18, 1-13.
- [18] Durborow, A. (2017). Factors contributing to the problem of student absenteeism in a rural school. University of Arkansas.
- [19] Vlachopoulos, P., & Jan, S. (2020). Exploring modes of lecturing as a teaching method in higher education: Student attendance, preference and motivation. *Journal of University Teaching & Learning Practice*, 17(5), 15.
- [20] Alamoudi, W. A., Alhelo, A. F., Almazrooa, S. A., Felemban, O. M., Binmadi, N. O., Alhindi, N. A., ... & Mawardi, H. H. (2021). Why do students skip classroom lectures: A single dental school report. *BMC medical education*, 21(1), 388.
- [21] Madane, R., Pawar, S., Sangrampatle, S., Rode, A., & Pawar, Y. (2024). Attendance management system using face recognition. *Industrial Management Advances*, 2(1).
- [22] Ozdemir, D., & Ugur, M. E. (2021). Model proposal on the determination of student attendance in distance education with face recognition technology. *Turkish Online Journal of Distance Education*, 22(1), 19-32.
- [23] Acasamoso, D. M., & Sagum, R. A. (2023, December). SEAMS: A Smart and Efficient Attendance Management System with Attendance Prediction and Forecasting. In *Proceedings of the 2023 11th International Conference on Information Technology: IoT and Smart City* (pp. 148-154).
- [24] Muzaferija, I., Mašetić, Z., Jukić, S., & Kečo, D. (2020). Student attendance pattern detection and prediction. *Journal of Natural Sciences and Engineering*, 1.

- [25] Jarbou, M., Won, D., Gillis-Mattson, J., & Romanczyk, R. (2022). Deep learning-based school attendance prediction for autistic students. *Scientific Reports*, 12(1), 1431.
- [26] Stockemer, D., & Stockemer, D. (2019). Multivariate regression analysis. *Quantitative methods for the social sciences: a practical introduction with examples in SPSS and stata*, 163-174.
- [27] Mette Kaasgaard,Uffe Bodtger,Anders Løkke,Erik Jakobsen & Ole Hilberg. (2024). Attendance rate and perceived relevance related to type, content, and delivery of current rehabilitation programmes after surgical resection for non-small cell lung cancer. *Frontiers in Rehabilitation Sciences*1447767-1447767.
- [28] I N Aslam,N A Putri,A A Maulana,A Muslim & R Ayuni. (2024). Application of multiple linear regression (MLR) on copper ore separation based on grade and density. *IOP Conference Series: Earth and Environmental Science*(1),012027-012027.
- [29] Iliad Mehraein & Siavash Riahi. (2017). The QSPR models to predict the solubility of CO₂ in ionic liquids based on least-squares support vector machines and genetic algorithm-multi linear regression. *Journal of Molecular Liquids*521-530.