

# *Research and Analysis Based on Cloud Class Data*

Hengrui Pan, Zixia You, Xinhong Liu\*, Piao Yang, Yiming Liu, Sirui Ding

*Beijing Institute of Petrochemical Technology, Beijing, 102617, China*

*\*Corresponding author: liuxinhong@bipt.edu.cn*

**Keywords:** Cloud class, correlation analysis, regression analysis

**Abstract:** This article applies the theoretical methods of probability and statistics, using probability and statistics cloud class courses as carriers, to study and analyze a large amount of data generated in cloud class courses, establish a mathematical model of related variables, and predict the final grades of students in cloud class courses. By presenting a learning situation data model, students can timely see their own problems and recognize the importance of process learning. At the same time, it provides a basis for teachers to conduct teaching behavior analysis and provide personalized and precise learning guidance for students.

## 1. Introduction

With the continuous development of modern information technology and the continuous updating of educational and teaching concepts, teaching models are also quietly changing. Cloud class courses, as a new type of teaching model, are gradually receiving people's attention and use. The cloud class teaching platform based on the network environment has broken the traditional mode of offline classroom teaching and replaced it with a modern "student-centered" teaching mode using mobile intelligent devices. This mode can fully demonstrate a new teaching platform of "interactive" and "participatory" teaching methods. At present, "Cloud Class" has become an intelligent mobile teaching platform for all theoretical and practical courses in our school. "Cloud Class" not only achieves real-time interaction, resource push, and homework tasks between teachers and students, but also records students' learning situations through "experience value" of cloud class, which helps teachers track and analyze students' learning trajectory, enabling teachers to grasp students' learning effects in real time, to provide a basis for conducting teaching behavior analysis and providing personalized and precise learning guidance for students.

## 2. The embodiment of cloud class in blended teaching

Blended teaching is a commonly used teaching mode in curriculum teaching in various higher education institutions in recent years. It is an "online" and "offline" teaching method that combines the advantages of online teaching and traditional teaching. Through the organic combination of online and offline teaching organizational forms, the optimization of teaching objectives is achieved, while guiding learners' learning from shallow to deep to deep learning. Cloud class is a bridge for blended teaching, which fully leverages the advantages of blended teaching through the implementation of three stages: pre class, in class, and post class.

## 2.1 Before class

Pre class preparation is a necessary task for teachers before teaching. Firstly, various teaching resources, such as syllabus and teaching calendar, are uploaded in the resource area to help students understand the main content and teaching progress of the course; For the activity area, teachers can also use the "brainstorming" and "homework" functions to analyze certain theoretical knowledge and typical cases, present them in the form of problems and tasks, and add learning guidance.

## 2.2 In class

During the teaching process, teachers can release classroom quizzes in the activity area based on the specific situation of the chapters, or use "brainstorming" to require students to take photos and upload typical cases. The teacher can directly display the students' problem-solving situation through the large screen. For protruding problems, students can communicate and discuss with each other, and the teacher can answer questions and explain relevant knowledge points. Additionally, targeted communication and interaction can be achieved through the use of "light live streaming". Students can also be tested for their mastery of the course content by randomly selecting individuals, answering questions, or raising their hands. Classroom interaction has greatly improved the effectiveness of students' listening in class and promoted their enthusiasm for learning.

## 2.3 After class

After class, teachers use the "homework activity" function of cloud class to publish assignments and set a deadline. Teachers can comment on the assignments submitted by students in the comment area. In addition, a "light live broadcast" is set up in cloud class classes for interactive Q&A, where students can ask their own questions at any time during the light live broadcast. Students can answer each other's questions, and the teacher will give appropriate experience rewards to students who ask and answer questions. For questions that students cannot answer, the teacher not only provides the answer process during the light live broadcast, but also explains them in the next class. This interactive Q&A method fully mobilizes the feelings of mutual assistance, unity, friendship, and common progress among classmates.

Students obtain corresponding experience values representing their own achievements by participating in various activities before, during, and after class. Teachers can preliminarily determine the learning status of students based on the experience values, and provide precise assistance to individual students. At the same time, appropriate adjustments can be made for the next step of teaching.

## 3. Correlation analysis of data variables<sup>[1-3]</sup>

At the end of the semester, the cloud class will generate complete data information. Through analysis and research of past data, the total experience value mainly consists of four parts, namely check-in, testing, homework, and daily interaction (including resource learning, classroom interaction, homework, brainstorming, etc. likes). This article collects and organizes data from a probability statistics cloud class in 2019, extracts experience values for testing, homework, and daily interaction, and performs corresponding score transformations to obtain daily scores corresponding to each experience value, namely  $X_1$  for daily interaction,  $X_2$  for homework, and  $X_3$  for test. Due to the fact that almost all students who sign in for class have full marks, they are ignored in variable correlation analysis and model fitting. In order to study the influence of each usual score

$X_k (k=1,2,3)$  on the student's Final examination score  $Y$ , we use the statistical software SPSS to conduct correlation analysis on variables  $X_1, X_2, X_3$  and  $Y$ , obtain the correlation coefficient between the Final examination score  $Y$  and each usual score  $X_k (k=1,2,3)$ , and conduct hypothesis testing on the correlation coefficient, as shown in Table 1 below.

Table 1: Correlation analysis between Final examination scores and other usual scores

	$X_1$	$X_2$	$X_3$
$Y$	0.700**	0.693**	0.329
	0.000	0.000	0.057

Table 1 shows that there is a positive correlation between the Final examination scores and the usual scores, which is statistically significant. The correlation values are 0.700, 0.693 and 0.392, respectively. It can be seen that the usual scores of each item have a significant impact on the Final examination scores, and the correlation between the usual interaction scores and the homework scores and the Final examination scores is relatively high. Therefore, it is necessary to remind students to actively participate in daily interactions and complete homework diligently[5].

#### 4. Establishment of regression forecasting model

In order to better study the relationship between final exam scores and daily interaction scores, homework scores and test scores, and to provide better suggestions and help for teachers and students, this paper conducts regression analysis on the final exam scores and daily scores of students in this cloud class. Multiple regression analysis is an effective mathematical method to deal with the interdependence of multiple variables. By using this method, a regression prediction model is established with the help of statistical software SPSS, and the rationality of the model and the significance of the regression coefficient are tested.

##### 4.1 Multiple linear regression model

In practical problems, random variables are often related to multiple ordinary variables  $x_1, x_2, \dots, x_n$ , and the multiple linear regression model is

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + \varepsilon, \varepsilon \sim N(0,1)_{[4]}$$

The maximum likelihood estimation and least square method are used to obtain the estimated values of the coefficients of the regression model. Get regression equation

$$\hat{y} = \hat{b} + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3 + \dots + \hat{b}_n x_n$$

##### 4.2 Regression prediction model between final exam scores and ordinary scores

The data of a probability and statistics cloud class in 2019 is still adopted, and the normal interaction score  $x_1$ , homework score  $x_2$  and test score  $x_3$  are all set to follow a normal distribution and are independent of each other. Statistical software SPSS was used to conduct regression analysis on the final grades  $y$  of variables and the results  $(x_1, x_2, x_3)$  of ANOVA and parameter estimation were obtained as shown in Table 2 and Table 3.

Table 2: Analysis of variance

Source	Sum of squares	Df	Mean square	F	p value
Regression	1933.388	3	644.463	48.627	0.000
Residual	397.583	30	13.253		
Aggregate	2330.971	33			

Table 3: Parameter estimation and significance test

	Nonnormalized coefficient		Standardization coefficient	t	p
	B	Standard error	Beta		
$x_1$	4.248	2.033	0.395	2.090	0.035*
$x_2$	3.638	1.892	0.363	1.923	0.044
$x_3$	2.198	3.089	0.092	0.711	0.082
Constant	-9.723	28.951	-	-0.336	0.739

It can be seen from the analysis of variance in Table 2,  $F = 48.627$ ,  $p < 0.05$ , indicating that the model construction has statistical significance, that is, it indicates that at least one item in daily interaction, homework and test will have an impact on the overall evaluation. The coefficient of

determination of the regression equation is  $R^2 = \frac{1933.388}{2330.971} = 0.829$ , indicating that the variable  $x_1$ ,  $x_2$ ,  $x_3$  has better fitting effect and significant regression effect. Table 3 shows the estimation of regression coefficient by regression and the significance hypothesis test. From the P value, it can be seen that the interaction scores  $x_1$  and homework scores  $x_2$  have a significant impact on the final exam scores, while the test scores  $x_3$  have no significant impact on the final exam scores. The constant term is significantly zero, but the overall regression effect is good.

This equation can be used as :

$$y = -9.723 + 4.248x_1 + 3.638x_2 + 2.198x_3 \quad [6-10]$$

This equation can be used as a reference for students to predict the final scores of Probability theory and mathematical statistics courses.

### 4.3 Model back generation results

The scores of the first 5 people were extracted from the sample data, and the above regression prediction model was used to calculate their actual final exam scores, predicted scores and the prediction interval of 0.95 confidence level. The results are shown in the Table 4.

Table 4: Final exam score and prediction

Final grade	Predicted score	Predicted value interval
94	91	(90.7, 96.6)
60	64	(58.4, 66.7)
91	91	(89.6, 95.8)
96	97	(95.0, 97.5)
77	74	(73.3, 79.0)

Through the actual calculation and comparison of some data, all the predicted values are within the prediction interval, and the error between the predicted values and the actual values is not large,

which indicates that the regression prediction model established has certain reference value.

## 5. Conclusions

(1) Through the statistics and analysis of the big data learning situation generated by a cloud class in one semester, it is not difficult to see that students' daily interaction and homework completion have a great impact on the final exam score, and both teachers and students should pay attention to it;

(2) For normal tests, through correlation analysis and regression analysis, this index has no significant impact on the final exam score. The analysis shows that the test questions are released after class, and students can finish them at any time before the next class, and the number of answers is more than once, most of the questions are single choice, and the questions are simple. There was little difference in students' scores on the test as a result of multiple rounds of questions. In addition, there is a great difference between the question type of the usual test and the final exam, which has no significant impact on the final exam score. The results of this analysis remind teachers that it is best to consider the way of testing and the type of testing questions when testing, so that the test can give full play to its best results<sup>[6-10]</sup>;

(3) The establishment of the model needs to be improved and perfected. This model only considers daily interaction, homework and testing, especially the daily interaction is not decomposed in detail.

## Acknowledgements

The authors gratefully acknowledge the financial support from URT project - Data Analysis and Grade Prediction Based on Cloud Class Courses of Beijing Institute of Petrochemical Technology in 2023 (2023J00097).

## References

- [1] Sheng Ju, Xie Shiqian, et al. *Probability theory and mathematical statistics* [M]. Beijing, higher education press, 2009
- [2] Wang Jiale, Wang Shizhi, Huang Qian. *Prediction of coal mine water inflow based on correlation analysis* [J]. *Shaanxi Coal*, 2021, 40 (02): 99-102
- [3] Chen Wei, Pan Liying, Lin Chaoran. *Research on the Competitiveness of Knowledge Intensive Manufacturing Industry Based on Canonical Correlation Analysis* [J]. *Learning and Exploration*, 2021 (03): 113-119
- [4] Zhang Houcan, Xu Jianping. *Modern Psychology and Educational Statistics. 3rd Edition* [M]. Beijing Normal University Press, 2009
- [5] Li Qiuyao. *Application of Multiple Linear Regression Model in River Water Quality Prediction* [J]. *Information Systems Engineering*, 2023 (07): 79-82
- [6] Yuan Qi. *Empirical Study on the Factors Influencing China's Foreign Exchange Reserves Based on Multiple Linear Regression Models* [J]. *China Business Review*, 2023 (13): 8-11
- [7] Ying Xiyuan, Sa Binhan. *Prediction of Strawberry Volume and Quality Based on Linear Regression Models* [J]. *Advanced Mathematics Research*, 2023, 26 (03): 86-90
- [8] Miao Hui, Luo Lulu. *Prediction of Logistics Demand in Guizhou Province Based on Multiple Linear Regression Model* [J]. *Logistics Technology*, 2023, 46 (08): 75-78
- [9] Yang Xinchun, You Wei, Wan Xiangyu, et al. *Reconstruction of Land Water Storage Changes in Nine Major Watersheds in China Using Multiple Linear Regression Models* [J]. *Geodesy and Geodynamics*, 2023, 43 (02): 116-120
- [10] Wan Jianguo. *Application of Multiple Linear Regression Model in Owner Satisfaction Evaluation: A Case Study of Residential Owners in Nanning City* [J]. *Real Estate World*, 2023 (03): 103-105