ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 205
The 2nd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2018)

Application of Multivariate Regression Analysis in Teaching Management*

Xiuli Li
School of Mathematics and Physics
Qingdao University of Science and Technology
Qingdao, China 266042

Pengcheng Zhao
School of Mathematics and Physics
Qingdao University of Science and Technology
Qingdao, China 266042

Yuxue Yang
School of Mathematics and Physics
Qingdao University of Science and Technology
Qingdao, China 266042

***Abstract*—This thesis made multivariate statistical analysis of academic scores of freshmen, sophomores and juniors and constructed a multivariate regression model, with the purpose of revealing the main factors influencing teaching quality management of colleges and universities as well as putting forward solutions to major influencing factors.**

Keywords—multivariate regression analysis; regression equation; significance testing

I. INTRODUCTION

In 1980s, western developed countries like Britain and the United States started to make theoretical discussions of the application of quality management practice in the education field, and now they have made pretty deep research on this. Since 1980s, quality management system standard has been widely popular with foreign scholars and colleges and universities that started to introduce the quality management system standard into teaching. In practical application, multivariate statistical analysis is an effective method that uses the mathematical statistical method to study theories and solutions to multivariate problems and also is the popularization of univariate statistics. In many education problems, a education phenomenon involves not one but more variables which have certain relationships among them, so it requires to handle the observed data of multiple variables. If using the univariate statistical method, you have to analyze multiple aspects respectively, that is, one aspect at a time, but this analysis method ignores the correlation between any two aspects and will lose a lot of information, so the analysis result obtained cannot reflect the real situation objectively and fully. If univariate statistical analysis is a course that studies the statistical regularity of a random variable, then multivariate statistical analysis will be a statistics course studying the interdependent relationship and internal statistical regularity of multiple random variables. With the popularization of computers, various statistical software has been launched, and multivariate statistical analysis method has been widely used in every field of educational management. All kinds of statistical software packages such as SAS and SPSS make it easy for practitioners to solve practical problems by using multivariate statistical analysis method.

As it is known, the management result of teaching quality is mainly reflected in students' academic scores, so based on current education management situations, the author established a regression equation of students' academic scores when giving the course "Multivariate Statistical Analysis" to study the correlation between academic sores of different courses, make a reasonable analysis of main factors influencing teaching quality so as to act appropriately in teaching management as well as applying multivariate statistical analysis in high-efficiency learning management. The essence of multivariate regression analysis model is to establish the relationship of a random variable y and p non-random variables $x_i\left(i=1,2,\mathrm{L}\ ,p\right)$. It can be expressed as:

$$y=f\left(x_1,x_2,\mathrm{L}\ ,x_p\right)+\varepsilon,$$

where $y=f\left(x_1,x_2,\mathrm{L}\ ,x_p\right)$ is a regression function which describes the leading functions of the factors to the observed quantity y; ε is chance fluctuation or model residual error, leading to inaccurate prediction of the value of y. After solving the linear regression equation, it has to make significance statistical hypothesis testing to verify if there is a correlation between y and $x_i\left(i=1,2,\mathrm{L}\ ,p\right)$. Considering that there are two major reasons that cause the difference of y: one is that an error is inevitable in every observation; the other is that in every observation, there might be other factors, except x_i, that can influence the value of y, so it has to distinguish the two aspects of factors properly to make objective judgment.

*Fund project: High-level Teaching Staff of Qingdao University of Science and Technology ---- Teaching team of basic Mathematics course in Science and Engineering.

The thesis made multivariate regression analysis of academic scores of 116 students in total including 62 Mathematics and Applied Mathematics majors and 54 Information and Computer Science majors of Class of 2011 of a college. There are two kinds of analysis of students' academic scores: one is the regression analysis of students' scores of a subject in freshman year and in later periods; the other is the regression analysis of overall scores of similar courses and those in later periods. In this way, we can understand students' basic learning situation, and then take proper measures in teaching management to arouse the initiative and enthusiasm of learning so as to improve learning quality. In case analysis, use the SPSS software to conduct statistical computing.

II. Regression Analysis of Students' Scores of A Subject in the Freshman Year and in Later Periods

Take four types of courses for Mathematics and Applied Mathematics and Information and Computer Science majors as the study objects, as shown in the "Table I".

TABLE I. Part of the Class Schedule of Class of 2011 of a College

Course	First Semester	Second Semester	Third Semester
1 (English)	College English B1	College English B2	College English B3
2 (Computer language)	C Language	C++ Programming	Java Language
3 (Analysis)	Mathematical Analysis 1	Mathematical Analysis 2	Mathematical Analysis 3
4 (Algebra)	Advanced Algebra 1	Advanced Algebra 2	Abstract Algebra

First, take the Mathematical Analysis scores of students of this major, use SPSS software to make regression and standardized residual normal p-p plot, as shown in "Fig. 1".

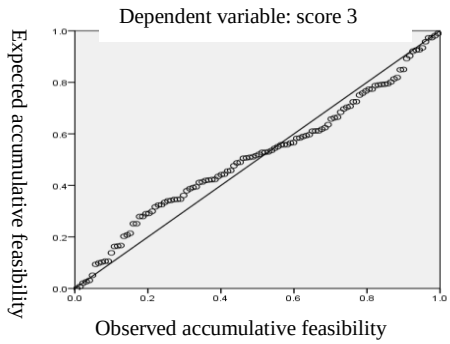

Fig. 1. Regression and standardized residual normal p-p plot.

As shown in "Fig. 1", there is a there is a linear relationship between the students' scores of Mathematical Analysis of this major in freshman year and in the later periods,

A. Modelling

Take students' midsemester scores of a subject in the first semester and the second semester as subjects, and establish a linear regression equation for predicting the students' scores of the subject in the third semester.

$$y_i = \alpha_{i1}x_{i1} + \alpha_{i2}x_{i2} + c_i + \varepsilon \quad (1)$$

Where, y_i represents the predicted score of subject i in the third semester; x_{i1} and x_{i2} stand for the scores of subject i in the first and in the second semesters respectively; are the coefficients corresponding to the scores of subject i in the first and in the second semesters respectively; c_i is a constant, $\varepsilon \sim N(0,1)$.

B. Model Solution and Verification

Use SPSS software to make regression analysis of the study subjects. The estimated values of α_{i1}, α_{i2} and c_i are shown in "Table II".

TABLE II. The Estimated Values of α_{i1}, α_{i2} and c_i

Value	α_{i1}	α_{i2}	c_i
1	0.157	0.533	21.842
2	-0.076	0.435	39.011
3	0.164	0.631	14.686
4	0.527	0.425	33.406

The linear regression equations for predicting the scores of courses in the third semester respectively are:

$$\begin{aligned} y_1 &= 0.157x_{11} + 0.533x_{12} + 21.842 \\ y_2 &= -0.076x_{21} + 0.435x_{22} + 39.011 \\ y_3 &= 0.164x_{31} + 0.631x_{32} + 14.686 \\ y_4 &= 0.527x_{41} + 0.425x_{42} + 33.406 \end{aligned} \quad (2)$$

Take the mathematical analysis statistical table as the example and use SPSS software to draw the residual statistical table in "Table III". According to the residual statistical table, the standard error of predicted value of Mathematical Analysis scores falls on interval [1.121, 3.889] and Mahalanobis distance (D) is in interval [0.060, 11.663]; the standard error of predicted value of Algebra course falls on interval [1.153, 4.700] and Mahalanobis distance (D) is in interval [0.001, 15.489]; the standard error of predicted value of English course falls on interval [0.469, 1.555] and Mahalanobis distance (D) is in interval [0.008, 9.998]; the standard error of predicted value of Computer Language course falls on interval [1.399, 4.762] and Mahalanobis distance (D) is in interval [0.080, 11.325]. The above data table shows the model established is feasible and can be adopted as a model of regression analysis of students' scores of a subject in the freshman year and in later periods.

TABLE III. MATHEMATICAL ANALYSIS RESIDUAL STATISTICAL DATA

	Minimum Value	Maximum Value (X)	Average Value	Standard Deviation	Value
Predicted value	39.5895	93.2362	73.7328	11.53480	116
Normal predicted value	-2.960	1.691	0.000	1.000	116
Standard deviation of the predicted value	1.121	3.889	1.811	0.525	116
Predicted value after adjustment	40.2806	93.1291	73.7261	11.54073	116
Residual error	-63.02269	27.08079	0.00000	11.61999	116
Standardized residual	-5.376	2.310	0.000	0.991	116
Studentized residual	-5.422	2.401	0.000	1.004	116
Deleted residual	-64.09528	29.25678	.00664	11.91177	116
Studentized deleted residual	-6.275	2.454	-.007	1.051	116
Mahalanobis distance (D)	0.060	11.663	1.983	1.887	116
Cook's distance	0.000	0.167	0.008	0.023	116
Leverage value	0.001	0.101	0.017	0.016	116

III. REGRESSION ANALYSIS OF STUDENTS' SCORES OF SIMILAR COURSES

Mathematics is a tool-based curriculum. Students' learning level of basic knowledge of mathematics will directly affect their academic scores of major courses. According to the surveys, mathematical analysis and advanced algebra are basic courses for mathematics and applied mathematics and information computing science of mathematics and physics of a university, which are correlated to academic scores of students in later periods. That is to say, the quality of academic scores of basic courses will, to some extent, directly affect the academic scores of major courses in later periods. In order to study the correlation of mathematical analysis, advanced algebra and major course study, improve students' academic scores and strengthen teaching quality management, this thesis takes the academic scores of students of Grade 11 as study objects, and conducted numerical analysis for basic specialized courses of mathematical analysis, advanced algebra and specialized courses in later periods, and made regression analysis for students' scores of probability theory and data statistics.

A. Model Analysis and Establishment

Take academic scores of mathematical analysis, advanced algebra and numerical analysis of students of 116 of Grade 11 of this major as subjects, use SPSS software to make regression and standardized residual normal p-p plot, as shown in "Fig. 2":

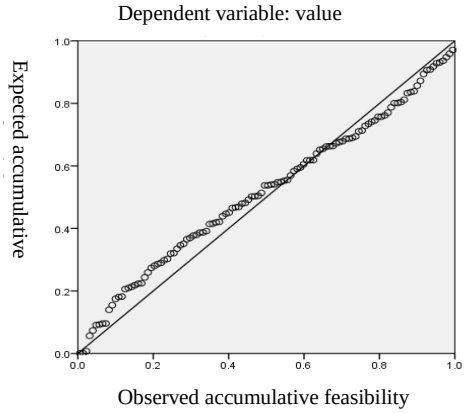

Fig. 2. Regression and standardized residual normal p-p plot.

As shown in the p-p plot above, students' scores of mathematical analysis and advanced algebra have a linear correlation to those of numerical analysis in later periods. From this, we can infer that ordinary differential equation is also linearly correlated, and make the following correlation analysis.

Take students' scores of statistical analysis and algebra in the first semester, the second semester and the third semester as the independent variables, and those of numerical analysis and ordinary differential equation as the dependent variables, establish a linear regression equation for predicting the academic scores of numerical analysis and ordinary differential equation:

$$z_j = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 y_1 + \alpha_5 y_2 + \alpha_6 y_3 + c_j + \varepsilon \qquad (3)$$

Where, $z_j (j=1、2)$ represents the predicted scores of numerical analysis and ordinary differential equation respectively; x_1、x_2、x_3、y_1、y_2、y_3 represents the students' scores of statistical analysis and algebra in the first semester, the second semester and the third semester; c_i is a constant, $\varepsilon \sim N(0,1)$.

B. Model Solution and Verification

Use SPSS software to conduct regression analysis of subjects, the predicted values of $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ are as shown in "Table IV".

TABLE IV. THE PREDICTED VALUES OF $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$

	α_1	α_2	α_3	α_4	α_5	α_6	c
y_1	0.030	0.017	0.136	0.134	0	0.355	26.414
y_2	0.294	-0.046	0.541	0.093	0	0	11.648

The linear regression equations for predicting the students' scores of numerical analysis in later period and ordinary differential equation are:

$$\begin{aligned} z_1 &= 0.030x_1 + 0.017x_2 + 0.136x_3 + 0.134y_1 + 0.355y_3 + 26.414 \\ z_2 &= 0.294x_1 - 0.046x_2 + 0.541x_3 + 0.093y_1 + 11.648 \end{aligned} \qquad (4)$$

TABLE V. STATISTICAL DATA FOR NUMERICAL ANALYSIS RESIDUALS

	Min. Value	Max. Value (X)	Average Value	Standard Deviation	Value
Predicted value	54.2105	91.6697	75.6034	8.98384	116
Normal predicted value	-2.381	1.788	0.000	1.000	116
Standard deviation of the predicted value	1.325	6.205	2.545	0.755	116
Predicted value after adjustment	51.4191	91.4510	75.5678	9.07571	116
Residual error	-63.01842	22.09747	.00000	11.40945	116
Standardized residual	-5.402	1.894	0.000	0.978	116
Studentized residual	-5.664	2.051	0.001	1.012	116
Deleted residual	-69.26913	25.91529	.03567	12.23323	116
Studentized deleted residual	-6.698	2.082	-0.009	1.074	116
Mahalanobis distance (D)	0.493	31.542	4.957	4.211	116
Cook's distance	0.000	0.530	0.012	0.052	116
Leverage value	0.004	0.274	0.043	0.037	116

TABLE VI. STATISTICAL DATA FOR ORDINARY DIFFERENTIAL EQUATION RESIDUALS

	Min. Value	Max. Value (X)	Average Value	Standard Deviation	Value
Predicted value	32.4430	97.5163	75.5345	12.77009	116
Normal predicted value	-3.374	1.721	0.000	1.000	116
Standard deviation of the predicted value	1.052	4.923	2.019	0.599	116
Predicted value after adjustment	23.9387	97.9203	75.4993	12.95772	116
Residual error	-26.77475	21.55705	0.00000	9.05308	116
Standardized residual	-2.893	2.329	0.000	0.978	116
Studentized residual	-3.022	2.750	0.002	1.015	116
Deleted residual	-29.22299	30.06129	0.03523	9.77772	116
Studentized deleted residual	-3.141	2.837	0.001	1.027	116
Mahalanobis distance (D)	0.493	31.542	4.957	4.211	116
Cook's distance	0.000	0.497	0.014	0.049	116
Leverage value	0.004	0.274	0.043	0.037	116

Draw the residual statistical table of the students' scores of numerical analysis and ordinary differential equation by using SPSS software, as shown in "Table V", "Table VI" respectively.

By observing the residual statistical table, we can conclude that the standard error interval of predicted values of mathematical analysis scores is [1.325, 6.205], and the standard residual error interval is [-5.402, 1.894]; the standard error interval of predicted values of ordinary differential equation scores is [1.052, 4.923], and the standard residual error interval is [-2.893, 2.329]. The above data shows that the established model is feasible and can be used as a correlation analysis model for the students' scores of mathematical analysis, algebra and those of numerical analysis in later periods and those of ordinary differential equation of this major.

IV. CONCLUSION

Through the regression analysis of students' scores of a subject in the freshman year and in later periods and those of similar courses, we can obtain Model (2) of students' scores in the freshman year versus those in later period and predictive Model (4) of students' basic courses scores versus students' similar courses scores in later periods. By observing the correlation coefficient Table 2 and Table 4, it is not difficult for us to find out that the academic scores of freshman of a subject has a significant influence on academic scores of this major in later periods, which indicates that only the foundation of a course is good, then the academic scores of the major can be further improved; The influence of basic courses in the freshman year on students' scores of similar courses shows that major foundation plays a crucial role in learning major courses in later periods. Therefore, in the teaching management of colleges and universities, it is very important to do a good job in students' study in the first year of university.

College students' academic performance is an important indicator reflecting the quality of teaching in colleges and universities. By analyzing the indicators like college students' academic performance data and make-up exams and so on, we intend to seek out unreasonable matters of the main factors affecting teaching quality in colleges and universities. Through case analysis, the following conclusions can be drawn: First, the learning status of college students in the first school year has a significant influence on their study in later periods, especially school courses in the second semester of the first school year. Second, the existing factors in students' learning style, such as: learning attitude, learning methods, etc., are the most important factors influencing the quality of teaching in colleges and universities. Third, students' knowledge, learning attitude, etc., are influenced by the curriculum system in some way. Fourth, teaching attitude and teaching method also have a certain influence on students' learning attitude and learning method.

The factors of students themselves are the most important factors affecting the teaching quality management in colleges and universities. However, the influence of a rational and scientific curriculum system, teaching attitudes and teaching methods on students' learning status can not be neglected as well, for they directly affect the college's talent training quality and management level. Therefore, if we want to improve the quality of teaching in colleges and universities, we must start from the main influencing factors and propose targeted and effective suggestions and countermeasures.

REFERENCES

[1] Backhaus. Multivariate Statistical Analysis [M], Shanghai: Gezhi Press. 2009.

[2] Li Chunlin, Chen Xuhong. Applied Multivariate Statistical Analysis [M], Beijing: Tsinghua University Press, 2013.

[3] Zhang Wentong. SPSS Statistical Analysis-Advanced Course, Beijing: Higher Education Press, 2017.

[4] Si Shoukui, Sun Xijing. Mathematical Modeling [M], Beijing: National Defence Industry Press, 2011.

[5] He Xiaoqun. Multivariate Statistical Analysis [M], Beijing: China Renmin University Press, 2004.

[6] Feng Zhaojun. Identification of Influencing Factors of China CPI Based on Multivariate Regression Analysis-Statistics and Decision [M], 24(492),2017:125-127.

[7] Li Xingxing. Study on College Students Learning Problems based on Multiple Regression Model [J], Journal of Higher Education, 2016 (18): 51-52.

[8] Deng Rong. Multimodal Teaching Principles Based on Multiple Regression Analysis [J], The Guide of Science & Education, 2015 (5): 35-36.