STATISTICAL ANALYSIS OF THE RELATIONSHIP BETWEEN THE SCORES OF BASIC COURSES AND PROFESSIONAL COURSES IN COLLEGES AND UNIVERSITIES -- TAKING SCIENCE AND ENGINEERING AS AN EXAMPLE

Zhiwen Zhu, Wenting Xu, Shixin Wang

Shangdong University of Technology

Abstract: Based on the latest data of students' scores in a university, this paper studies the relationship between the scores of basic courses and professional courses in Colleges and universities. Firstly, canonical correlation analysis is used to qualitatively understand the correlation between basic courses and professional courses, and find out which courses have correlation; Secondly, the common factors of basic courses are extracted by factor analysis to achieve the purpose of dimension reduction and simplification; Finally, through regression analysis, find out the specific correlation between basic courses and professional courses. The results show that there is a close correlation between basic courses and professional courses in Colleges and universities. Among them, the basic course of mathematics has the greatest impact on science and engineering.

Key words: university education curriculum relationship research canonical correlation analysis factor analysis

1. Introduction

With China's reform and opening up and the further development of modern education system, higher education has been gradually improved, but there are still some problems, such as unreasonable curriculum arrangement and lack of pertinence in college students' education and learning. Therefore, it is particularly important to find the curriculum characteristics between different disciplines and design the curriculum arrangement according to the relationship between basic courses and professional courses.

At present, there have been a lot of comprehensive analysis on College Students' examination results, but there are few cases of comparative study on the relationship between basic courses and professional courses for different majors. In view of this, studying the internal relationship between professional courses according to the nature of different disciplines, constructing a reasonable achievement model, and reasonably setting up the curriculum arrangement of colleges and universities according to the relationship between basic courses and professional courses is not only of great significance to the development of the school, but also plays a vital role in students' more solid mastery of professional knowledge.

2. Empirical Analysis

2.1 Data Selection and Processing

The data in this paper comes from the summary of 2017's total scores of colleges and universities of a university, which is true, reliable and comprehensive. In this paper, the test results are used to represent the learning ability of students in each course. Combined with the nature of the course and the actual situation of the school, the representative science and engineering major vehicle engineering of the school is selected.

The selection of professional research courses is summarized from two aspects: basic courses and professional courses. For the selection of basic courses, it mainly selects highly representative public courses such as advanced mathematics, line generation, probability, English and big things that all majors need to learn, so as to make a comparative analysis of the impact of the same courses of all majors on this basis. In addition, targeted basic courses in all majors are added. For the selection of professional courses, combined with the national professional development direction and the importance of courses, several courses are selectively selected, which not only simplifies the analysis of data, but also reduces the impact of errors.

It can be seen from Fig. 1 and Fig. 2 that their distributions are similar and obey normal distribution. Firstly, the missing value is filled by linear fitting prediction method; Secondly, the repeated value is processed by weight adjustment method; Finally, standardization is carried out to eliminate the dimension between variables and make the data comparable.

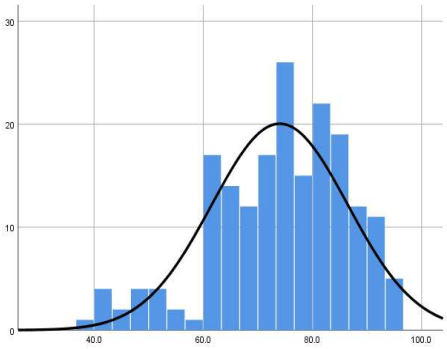

Figure 1 higher mathematics 2 score distribution

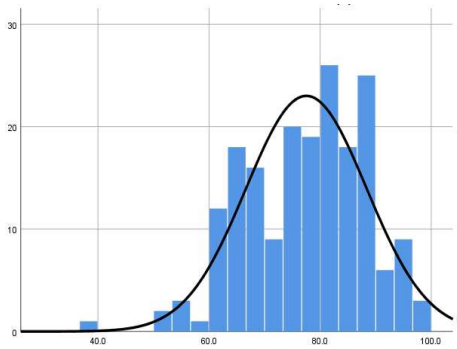

Figure 2 distribution of results in probability theory and mathematical statistics

2.2 Data Analysis of Vehicle Engineering in Science and Engineering

This section analyzes the scores of basic courses and professional courses of students majoring in vehicle engineering. Firstly, through descriptive statistical analysis, we have a general grasp of the score distribution of the whole major; Then use canonical correlation analysis to qualitatively understand the correlation between basic courses and professional courses, and find out which courses have correlation;Then conduct factor analysis and regression analysis to find out the specific correlation formula between the courses with correlation.

2.2.1 Definition of Variables

Firstly,the symbolic representation of each variable is given, as shown in Table1:

Table 1 definition of variables

Independent variable (score)	Course name	Dependent variable (grade)	Course name
x1	C language	y1	Automobilestructure (A)
x2	College Physics	y2	Automobile theory
x3	College English	y3	Automotive design (A)
x4	Advanced mathematics	y4	Mechanical design (A)
x5	Descriptive geometry and engineering drawing	y5	Mechanical principle (A)
x6	linear algebra	y6	Mechanics of materials (B)
x7	Probability theory and mathematicalstatistics	y7	Electronic Technology (A)
		y8	Electrical Technology (A)
		y9	Theoretical mechanics (traffic) (A)

2.2.2 Correlation Analysis

Because analyzing the correlation between various subjects alone can not draw the conclusion we want, we can directly study the overall correlation between the two groups of variables, that is, using canonical correlation analysis. According to the correlation coefficients between the subjects, the correlation coefficients between Yi and X are very close, indicating that the information covered by Yi overlaps, and it is necessary to extract typical variables to represent this correlation. Only the canonical correlation coefficient of the first pair of canonical variables is high, and the probability P-value is less than the significance level of 0.05; The significance p-values of the canonical correlation coefficients of the second to seventh pairs of typical variables were greater than 0.05, which were 0.221, 0.842, 0.962, 0.955, 0.949 and 0.993 respectively, indicating that a pair of typical variables were extracted from the two groups of variables.

Write the expression of typical variables according to the standardized typical coefficient of each typical variable and each variable in each variable group:

$$U = -0.051x_1 - 0.311x_2 - 0.059x_3 - 0.317x_4 - 0.152x_5 - 0.073x_6 - 0.33x_7$$
$$V = 0.111y_1 - 0.132y_2 - 0.107y_3 - 0.081y_4 - 0.221y_5 - 0.255y_6 - 0.06y_7 - 0.399y_8 - 0.155y_9$$

V reflects the learning ability of professional courses of students majoring in vehicle engineering in University A, and u reflects the learning ability of basic courses of students majoring in vehicle engineering in University a. Because the canonical correlation coefficient is as high as 0.825 and has passed the significance test, it is considered that the learning ability of basic courses is also very important to the learning of professional courses, so we can conclude that there is an important correlation between the two kinds of indicators.

Table 2 proportional interpretation of variance

Canonical Variable	Set 1 by Self	Set 1 by Set 2	Set 2 by Self	Set 2 by Set 1
1	0.411	0.548	0.810	0.747
2	0.057	0.008	0.074	0.011
3	0.079	0.006	0.043	0.003
4	0.111	0.004	0.067	0.002
5	0.067	0.002	0.071	0.002
6	0.096	0.001	0.049	0.001
7	0.078	0.000	0.054	0.000

The proportion of variance explained by typical variables to each variable group is known from table 2. The variance proportion of dependent variable group that can be explained by typical variable U is as high as 0.747, which shows the explanatory ability of independent variable group to dependent variable group, that is, the basic course performance of students majoring in vehicle engineering can have a great impact on the professional course performance. The variance ratio of the independent variable group that can be explained by the typical variable V is 0.548. Compared with the influence of the independent variable group on the dependent variable group, the explanatory ability of the dependent variable group on the independent variable group is relatively limited, which fully proves the importance of learning basic courses first.

2.2.3 Factor Analysis

This paper makes factor analysis on the scores of basic courses of students majoring in vehicle engineering in a university, reduces the dimension of variables, and then uses the common factors of basic courses to study the relationship between basic courses and professional courses。

There is a strong collinearity between basic courses, and the correlation coefficients between variables are high, so factor analysis can be used to reduce dimensions. The observed value of Bartlett sphericity test statistic is 495.715, KMO value is 0.859, and the corresponding probability P-value is close to 0. It is considered that the correlation coefficient matrix is significantly different from the unit matrix, and the original variables are suitable for factor analysis。

When the number of common factors extracted is 4, the commonality of all variables is high, and the information loss of each variable is less. Therefore, the overall effect of factor extraction is ideal.

Table 3 interpretation of total variance

	Initial eigenvalue			Extract the sum of squares of the loads			Sum of squares of rotating loads		
component	total	Percentage variance	Cumulative %	total	Percentage variance	Cumulative %	total	Percentage variance	Cumulative %
1	3.707	52.961	52.961	3.707	52.961	52.961	2.425	34.639	34.639
2	0.847	12.107	65.068	0.847	12.107	65.068	1.323	18.895	53.534
3	0.778	11.112	76.180	0.778	11.112	76.180	1.098	15.679	69.213
4	0.545	7.792	83.972	0.545	7.792	83.972	1.033	14.759	83.972
5	0.441	6.300	90.272						
6	0.375	5.360	95.632						
7	0.306	4.368	100.000						

Extraction method: principal component analysis

The relevant information among the four factors can be seen from table 3. The variance contribution of the first factor was 3.707, which explained 52.961% of the total variance of the original seven variables; The variance contribution of the second factor is 0.847, which explains 12.107% of the total variance of the original seven variables, the variance contribution of the third factor is 0.778, which explains 11.112% of the total variance of the original seven variables, and the variance contribution of the fourth factor is 0.545, which explains 7.792% of the total variance of the original seven variables. Their cumulative variance contribution rate, that is, the proportion of these four factors in the impact of all basic course scores, is 83.972%, Furthermore, these four potential factors can explain about 83.972% of the total variance.

Table 4 composition matrix after rotation [a]

	component			
	1	2	3	4
Linear algebra (B)	0.822	-0.009	0.334	0.171
Probability theory and mathematical statistics (D)	0.795	0.247	-0.029	0.155
Advanced mathematics	0.766	0.285	0.246	0.107
College Physics	0.638	0.581	0.058	0.101
C language	0.201	0.869	0.311	0.117
Descriptive geometry and engineering drawing	0.201	0.273	0.898	0.141
College English	0.205	0.117	0.134	0.962

Extraction method: principal component analysis
Rotation method: Caesar normalization maximum variance method
a. The rotation has converged after 6 iterations

The loads of the first and second factors of the seven variables are very high, while the loads of the third and fourth factors are low. The actual meaning of the factor is vague, so factor rotation is carried out. After orthogonal rotation, the factor has naming interpretability. As can be seen from table 14, the first factor mainly explains the four variables of higher mathematics, linear algebra, probability theory and mathematical statistics, and college physics, which can be interpreted as the learning ability of logical geometry; C language has a high load on the second factor, which can be interpreted as the learning ability of programming language; The third factor can be explained as the learning ability of engineering drawing; The fourth factor mainly explains English learning ability.

The factor analysis model is established as follows:

$$\begin{cases} \text{Linear algebra} = 0.822F, -0.009F2 + 0.334\,F3 + 0.171F \\ \text{Probability theory and mathematical statistics} = 0.795F + 0.247F, -0.029F, + 0.155F \\ \text{Advanced mathematics} = 0.766F, + 0.285F, + 0.246F, + 0.107F \\ \text{College physics} = 0.638\,F + 0.581\,F, + 0.058\,F3 + 0.101\,F \\ \text{C I} = 0.201F + 0.869F, + 0.311F, + 0.117F \\ \text{Descriptive geometry and engineering drawing} = 0.201F, + 0.273F + 0.898F, + 0.141F \\ \text{College English} = 0.205F + 0.117F + 0.134F, + 0.962\,F4 \end{cases}$$

The factor score can be used to evaluate the basic course learning of each student. According to figure 3 and Figure 4, we can see that there is a certain gap in the learning ability of logic geometry and C language among students,

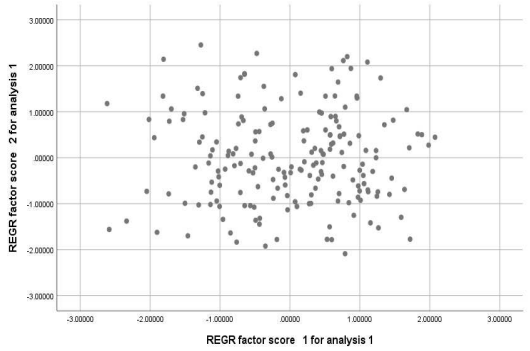

Fig. 3 Relationship between factors 1 and 2

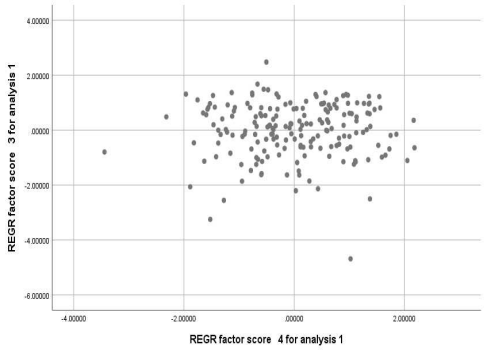

Fig. 4 Relationship between factors 1, 3 and 4

2.2.4 Regression Analysis

The extracted common factors of basic courses and each professional course are regressed in turn, and the four common factors F1, F2, F3 and F4 are used for linear regression with professional courses to find out which professional courses are closely related to the study of basic courses. Taking automobile structure as an example, the regression model is established by stepwise regression. According to table 5, the fitting effect of the third model is better, which shows that the learning ability of the professional course of automobile structure is related to the learning ability of the basic course studied.

Table 5 model summary

Model	R	R square	Adjusted R square	Error in standard estimati
1	0.366a	0.134	0.129	6.7685
2	0.780b	0.608	0.602	6.3989
3	0.899c	0.808	0.804	6.3364

Except for the fourth common factor, the probability p-values of other items are less than the significance level of 0.05, that is, factors F1, F2 and F3 are important variables, which have a significant impact on the learning of automobile structure. The regression model of basic course scores on automobile structure scores is as follows:

$$y_1 = 2.655 f_1 + 2.249 f_2 + 1.001 f_3 + 76.229$$

From the above analysis, it can be seen that the professional course of automobile structure is mainly affected by the learning ability of logical geometry, programming language and engineering drawing, among which the learning ability of logical geometry is the most influential. Therefore, when designing the course, the study of the professional course of automobile structure should be arranged after advanced mathematics, linear algebra, probability theory and mathematical statistics, C language, descriptive geometry and engineering drawing, and the study of mathematics should be effectively connected with the study of automobile structure. The time span should not be too long, so as to consolidate and apply important knowledge in time, Study the professional course of automobile structure more efficiently.

For the analysis of other professional courses, substitute the above analysis process, and draw a conclusion: the professional course of automobile theory is mainly affected by the learning ability of logical geometry and programming language, among which the learning ability of logical geometry is the most influential; The professional course of automobile design is mainly affected by the learning ability of logical geometry, programming language and engineering drawing, among which the learning ability of logical geometry is the most influential; Mechanical design and electronic technology are mainly influenced by the learning ability of logical geometry, programming language and English; Mechanical principle, material mechanics, electrical technology and theoretical mechanics are mainly affected by the learning ability of logical geometry, programming language, engineering drawing and English, among which the learning ability of logical geometry is the most influential.

3. Conclusions and Recommendations

3.1 The Influence Between Basic Courses and Professional Courses of Science and Engineering

For science and engineering majors, there is a significant correlation between the common factors extracted from basic courses and professional courses, which greatly affects the results of professional courses. Combined with the coefficient proportion of principal components, it is concluded that the most influential basic course is higher mathematics. There is a close relationship between basic courses and professional courses, and students' performance is that the higher the basic courses, the higher the professional courses, and the lower the basic courses, the lower the professional courses. Relevant conclusions are as follows:

Programming is a very important basic course, and its influence even goes beyond English learning. It plays a vital role in the establishment of learning ability and logical thinking. This conclusion is also in line with the development trend of today's society.

Mathematics has a great impact on professional courses, and science and engineering professional courses are more affected by mathematics. The study of science and engineering courses focuses more on the accumulation of science knowledge such as mathematics and physics, and less on the study of written courses such as English and Chinese.

3.2 Suggestions on Curriculum Design of Science and Engineering

In view of the relationship between basic courses and professional courses of science and engineering majors, combined with the conclusions of empirical analysis, the following suggestions are provided for the curriculum arrangement of science and engineering majors in Colleges and universities in China:

Strengthen the learning duration of basic mathematics courses and ensure their continuity with the curriculum arrangement of professional courses. For example, arrange to study advanced mathematics in the first semester, study linear algebra, probability theory and mathematical statistics in the second semester, start the study of professional courses in the third semester, and do a good job in the connection of professional courses on the premise of repeatedly consolidating the foundation of mathematics, In order to better apply the basic knowledge learned to the study of professional courses and improve learning efficiency.

The learning mode of English can be appropriately changed. Since its impact on the learning of professional courses cannot be reflected in the short term, the learning of English courses can be flexibly arranged on the premise of ensuring the reserve of basic knowledge.

Pay more attention to physics and computer programming courses and ensure that relevant courses are carried out before professional courses, so as to avoid the phenomenon that cannot be understood due to the knowledge of physics and programming in the learning process of some professional courses.

We should reasonably set up basic courses in this professional field. No matter what major, we should set up at least one professional basic course, rather than directly enter the deep-seated professional course learning. Because learning should be gradual, so should the curriculum, which helps to lead students to more effectively understand the research direction of their major.

BIBLIOGRAPHY

Wang Songgui , Chen Min, Chen Liping. Linear statistical model: linear regression and analysis of variance [M]. Higher Education Press: Beijing, 1999:55

Kang bodi, Yang ruxuan, Cheng Xiaorui, et al. Analysis of influencing factors of College Students' achievement based on regression model [J]. Education and teaching forum, 2020, (17): 94-95

Liao Yi jing. Analysis on Influencing Factors of College Students' academic performance [J]. Journal of Shanxi energy University, 2020, 33 (5): 45-47

Chen Zhongwei, Hui Shurong, Dong Jianguo. Application of principal component analysis in performance analysis of professional basic courses [J]. Higher agricultural education, 2011 (06): 42-45 + 50