Factor Analysis Algorithm and Precision Teaching

Yuxiong Chen[1], *Jianhua Zhang[2]

[1]*School of information, Guangdong Nanfang Institute of Technology, Jiangmen, China*
[2]*School of information, Guangdong Nanfang Institute of Technology, Jiangmen, China*
*35524677@qq.com, *2727611989@qq.com*

Abstract:
After the talent training plan is determined, how to arrange the teaching plan, teaching plan and teaching sequence of various courses in different periods of time is a very complex mathematical programming problem of "multifactor and multi-objective". This paper describes how to use "factor analysis", "principal component analysis", and other algorithms, use IBM SPSS as the data analysis tool, and take the actual teaching data as the sample to conduct a comprehensive quantitative analysis on the students' grades of various subjects at the end of the semester, and take this as the basis to guide the process of teaching work improvement. This kind of teaching research based on advanced computing tools, scientific quantification and accurate data is a necessary measure for digitization and standardization of the current teaching reform. In this paper, the basic conditions of using a factor analysis algorithm, the process of establishing a mathematical model, and the detailed operation steps of using IBM SPSS as a data analysis tool to solve problems are explained first, and then the calculation results are analyzed. And, the characteristics of the algorithm, the practical application, and even the taboo of the application are introduced. Finally, it is also mentioned that different algorithms have different characteristics and different applications. Only by flexibly applying various algorithms to deal with different problems can we improve work efficiency. This is our basic working principle

Keywords: *Factor analysis, correlation analysis, application of IBM SPSS, precision teaching.*

1. INTRODUCTION

In recent years, there are two hot topics in the teaching reform of colleges and Universities: one is precision teaching. Different teaching programs are implemented according to different groups of students, to give full play to the talents of each student and make them gain more learning during their college years. The second is to use advanced data analysis tools in teaching reform to improve and process daily teaching management based on accurate data. This also includes timely revision of teaching plans and teaching plans and reasonable arrangement of teaching sequence of each course during the implementation of talent training programs. The relationship between the two can also be understood as follows: the latter is a specific work process, and the former is the goal of educational reform.

After the college talent training plan is determined, how to arrange the learning sequence of multiple courses reasonably, effectively and realistically is one of the important steps for teaching management to realize and achieve the teaching objectives. Due to the differences in students' learning bases, for example, new college students from different regions must have different learning base. The teaching management department should plan and arrange their courses more carefully. For another example, in different teaching stages, under the same teaching scheme, the differences between the characteristics of students and the degree of accepting the course content will also have different needs for the teaching scheme, teaching plan and even the arrangement order of multiple courses in this period. These are the teaching reform issues that we should explore and study. The teaching plan, teaching plan and teaching sequence of multiple courses at the same time are the mathematical programming problems of "multifactor and multi-objective". How to achieve the optimal or relatively optimal effect and make it relatively reasonable and adaptive, we must adopt targeted and reasonable algorithms and advanced calculation tools to solve the problem, that is, based on the accurate data after calculation, revise, implement and arrange new teaching programs, teaching plans, and even the teaching sequence of multiple courses.

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 797-803, 2023.
https://doi.org/10.2991/978-94-6463-040-4_121

Moreover, because the students who enter the school each year and their situations are different, the students who are educated in each learning period in the established talent training program are different groups, and their teaching in the current learning period is certainly related to their teaching in the previous learning period, this teaching research not only has its complexity, but also is a work that should be insisted and explored all the year round. If we can find out the dynamic change law of this teaching research, this teaching research can even be said to be more complex system engineering. Moreover, because the students who enter the school each year and their situations are different, the students who are educated in each learning period in the established talent training program are different groups, and their teaching in the current learning period is certainly related to their teaching in the previous learning period, this teaching research not only has its complexity, but also is a work that should be insisted and explored all the year round. If we can find out the dynamic change law of this teaching research, this teaching research can even be said to be more complex system engineering.

In order to test the teaching effect of a certain teaching period, it involves which method, algorithm and calculation tool we can use to accurately obtain reliable data after the implementation of the teaching plan, and rely on these data to improve the next teaching plan, teaching plan, and even the teaching sequence of multiple courses. It is indeed quite difficult to solve this problem, because in the same period of teaching, the correlation and degree of correlation between multiple courses are different, and their correlation is divided into different course groups according to whether their teaching effects are relevant. The teaching effects of each course in different course groups are related to the number of class hours of the course, it is also related to the teaching effect of other closely related courses in the same course group. Therefore, this is a mathematical problem of "multifactor analysis" and, of course, a mathematical programming problem of "multifactor and multi-objective", which further reflects the complexity of solving this problem. Only by abandoning the old and outdated working method of "relying on experience" and implementing scientific, digital and standardized teaching management can we realize the modernization of teaching management. Of course, this is the perfect solution and implementation process to achieve precision teaching in teaching management [1].

2. ALGORITHM

In practical problems, it is often encountered to study multiple indicators. In most cases, there is a certain correlation between different indicators. Because there are many indicators, and there is a certain correlation between indicators, this is bound to increase the complexity of the analysis. Generally speaking, we pay attention to and carry out principal component analysis, factor analysis and other algorithms to solve such problems. In fact, in more occasions, we often combine multiple algorithms to better solve practical problems.

Principal component analysis, through linear combination, synthesizes the original variables into several principal components, and replaces the original more indicators (variables) with fewer comprehensive indicators. In multivariate analysis, some variables are often correlated. What is the reason for the correlation between variables? Is there a common factor that cannot be directly observed but affects the change of observable variables? Factor analysis is a model analysis method to find these common factors. It constructs several common factors with clear meaning on the basis of principal components, decomposes the original variables with them as a framework, and inspects the relationship and difference between the original variables.

The main purpose of factor analysis: when there is a high correlation between variables, we use fewer factors to summarize their information. After transforming the original variables into factor scores, factor scores are used for other analyses, such as cluster analysis, regression analysis, etc. And calculate the comprehensive score through each factor score to comprehensively evaluate the analysis object.

The basic idea of factor analysis is to classify the observed variables and classify the variables with high correlation, that is, the closely related variables into the same category, while the correlation between variables of different classes is low. Then, each kind of variable actually represents a basic structure, namely, common factor. The problem we study is to try to use the sum of the least number of linear functions of un-measurable common factors and special factors to describe each component of the observation.

It should be noted that whether a group of data is suitable for the factor analysis method needs to be determined first. We use KMO (Kaiser Meyer Olkin) sampling suitability test, which is one of the methods of suitability test. KMO sampling fitness test statistic is an index used to compare the square sum of correlation coefficient and the square sum of partial correlation coefficient between observation variables. Partial correlation refers to the degree of correlation between two variables when they are related to the third variable at the same time and the influence of the third variable is excluded. It can be seen that the greater the absolute value of partial correlation coefficient, the less likely there is a common factor between the two variables, indicating that factor analysis may not be suitable. If the absolute value of correlation coefficient between variables is large and the absolute value of partial correlation coefficient is small, it indicates that the high correlation between variables may be related to the third variable, and there is a high possibility of common factors, so factor analysis is suitable. Calculation formula of KMO statistics:

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}^{2}}{\sum\sum_{i \neq j} r_{ij}^{2} + \sum\sum_{i \neq j} p_{ij}^{2}}$$

Where: r_{ij} is the correlation coefficient between variables; p_{ij} is the partial correlation coefficient between variables. According to the formula, the value of KMO statistics is between 0 and 1. According to experience: the common kmo metrics for judging whether variables are suitable for factor analysis are:

KMO>0.9 means very suitable;

0.9> KMO>0.8 means suitable

0.8> KMO>0.7 means normal

0.7> KMO>0.6 means not suitable

KMO<0.5 means extremely unsuitable

The purpose of using factor analysis in the experiment is: "through factor analysis, the original variables will be transformed into new factors. The correlation between these factors is low, while the correlation of variables within the factors is high". In this way, we can judge the correlation (high or low) between the courses arranged in a certain teaching period. Using advanced methods to analyze problems, accurate data are obtained, which is of great help to the subsequent teaching work: Based on this data, it can be used as a reference for revising, implementing and arranging new teaching programs, teaching plans, and even the teaching sequence of multiple courses [2].

3. IMPLEMENTATION PLAN

We randomly selected one teaching class as the data collection object, and took their scores of each subject in the first semester of the 2021-2022 academic year as the data sample. These courses are military courses, college physical education, computer mathematics, C programming language, Fundamentals of computer application, the localization of Marxism and the mission of young students entering education, Fundamentals of ideology, morality and law, etc. To simplify the table, we replace the names of these courses with symbols A, B, C, D, E, F, G, H, I, etc. The sample data collected is in Table 1.

Table 1: Data by Class 1.

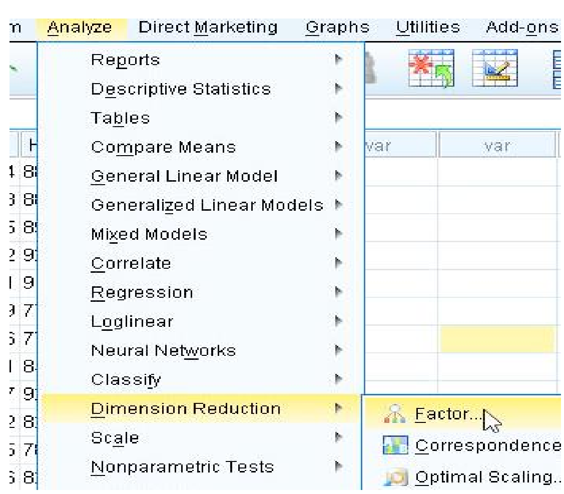
No.	A	B	D	E	F	H	I
10301	89	66	57	76	64	88	72
10302	85	65	77	72	68	88	71
10303	80	91	73	93	65	89	76
10304	85	92	68	76	72	92	73
10305	83	64	64	80	61	91	80
10306	83	66	70	76	69	77	71
10307	83	70	59	82	66	77	74
10308	71	67	66	93	61	84	80
10309	85	70	66	63	67	92	74
10310	80	66	62	85	62	82	67
10311	78	72	57	78	65	78	65
10312	83	74	66	69	76	82	72
10313	80	75	62	78	62	81	69
10314	73	67	64	71	71	66	73
10315	83	66	71	63	61	92	72
10316	82	92	74	95	80	95	67
10317	85	69	63	82	60	81	73
10318	79	68	63	72	61	73	71
10319	83	68	65	91	70	91	76
10320	83	93	83	63	61	80	73
10321	85	93	67	90	85	88	81
10322	71	64	81	76	60	92	70
10323	78	71	65	83	68	82	70
10324	77	64	69	72	72	78	65
10325	77	95	76	81	62	82	82
10326	78	86	78	91	83	93	83
10401	78	66	85	64	63	80	69
10402	71	92	81	73	60	88	63
10403	73	66	71	78	82	75	72

Of course, the data in the above data table has been cleaned and normalized.

3.1 Data Analysis Process

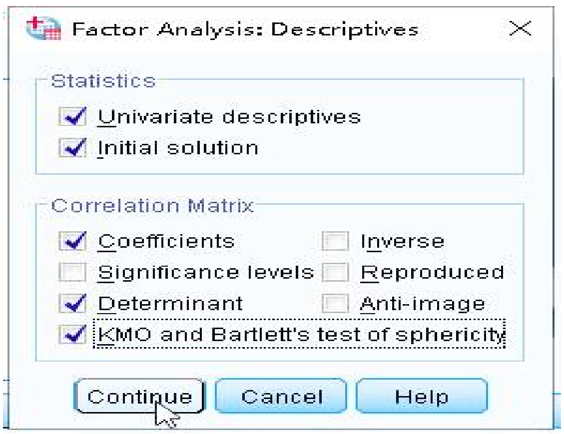
Firstly, import the data of Table 1 in the IBM SPSS work interface, and then perform dimension reduction.

Table 2: Perform dimension reduction.



As we know, factor analysis is a statistical analysis method to classify variables by studying the correlation matrix between variables, reducing the complex relationship between these variables to a few comprehensive factors. Because the number of factors summed up is less than the number of original variables, but they also contain the information of original variables, this analysis process is also called dimension reduction.

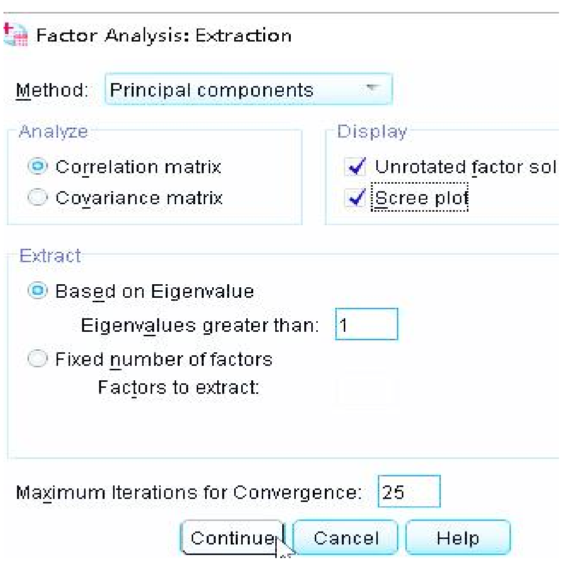
Table 3: Set parameters of descriptive statistics.



Descriptive statistical analysis can make us clearly understand the relationship between variables, and also make necessary preparations for inference and prediction analysis.

Next, set relevant parameters of data extraction.

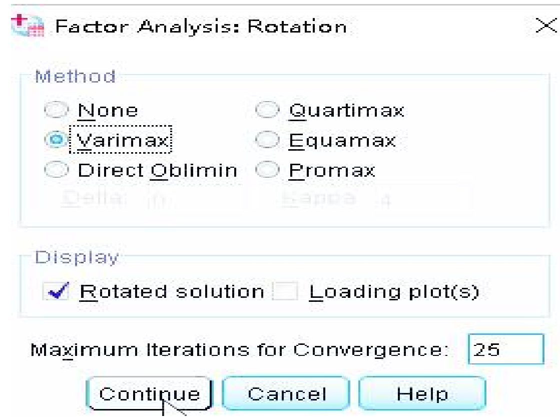
Table 4: Set relevant of data extraction.



In factor analysis, the eigenvalue of a factor should reach a certain standard. The eigenvalue reflects the contribution of a factor's variation to the variance of all variables. The larger the eigenvalue, the greater the explanatory power of the factor to all original variables, and the smaller the eigenvalue, the smaller the explanatory power of the factor to all original variables. Generally, the value is greater than or equal to 1, and the characteristic value less than 1 is generally not used as a common factor.

Next:

Table 5: Rotate the factor load matrix.



Next:

Table 6: Set the factor score parameters.

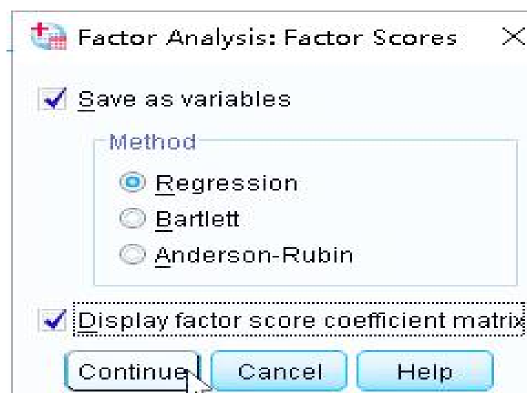
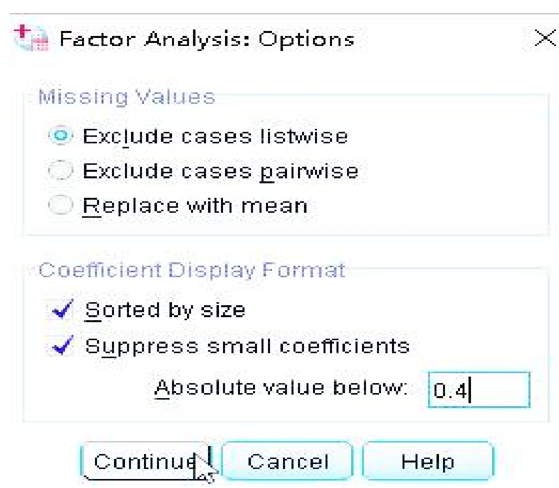


Table 7: Set the parameters of factor analysis options.



All operations completed. Let's analyze the calculation results.

3.2 Operation Result Analysis

Table 8: Descriptive Statistics.

	Mean	Std. Deviation	Analysis N
A	80.06	6.277	50
B	74.02	10.508	50
D	70.44	8.598	50
E	79.54	10.156	50
F	67.96	8.850	50
H	82.98	8.703	50
I	73.94	6.300	50

It can be seen from the table that the highest average score of item H is 82.98; the standard deviation is 8.703.

Table 9: Correlation Matrix.

		A	B	D	E	F	H	I
Correlation	A	1.000	.202	.197	.297	.215	.432	.470
	B	.202	1.000	.316	.193	.325	.321	.249
	D	.197	.316	1.000	.264	.142	.247	.382
	E	.297	.193	.264	1.000	.308	.410	.466
	F	.215	.325	.142	.308	1.000	.248	.367
	H	.432	.321	.247	.410	.248	1.000	.377
	I	.470	.249	.382	.466	.367	.377	1.000

a. Determinant = .237

Analysis: it can be seen from the above table that the correlation coefficient between item a and item I is 0.47, and the correlation in this table is the largest. The correlation coefficient between E and I is 0.466, which is second only to the former.

Table 10: KMO and Bartlett's Test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.773
Bartlett's Test of Sphericity	Approx. Chi-Square	65.988
	df	21
	Sig.	.000

The measurement value of KMO shown in this table is 0.773, which indicates the partial correlation between variables. This value shows that our analysis method of the sample data (the data in Table 1) using the factor analysis algorithm is basically consistent and acceptable [3] [4].

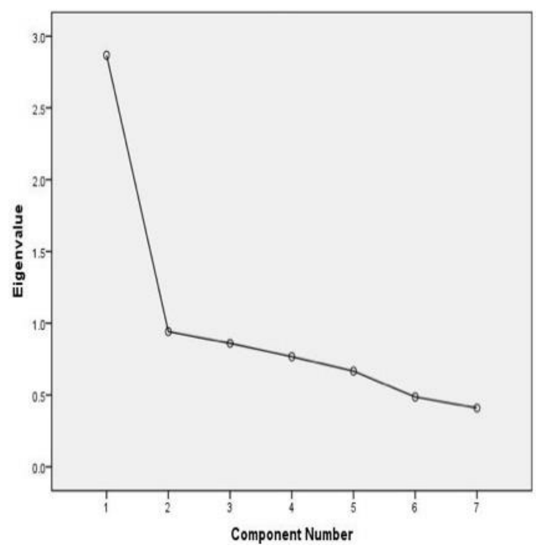
Table 11: Total Variance Explained.

	Initial	Extraction
A	1.000	.414
B	1.000	.305
D	1.000	.301
E	1.000	.455
F	1.000	.320
H	1.000	.482
I	1.000	.589

Extraction Method: Principal Component Analysis.

This table gives the common degree of each initial variable in the second column; and the extracted quantity confirmed by the system in the third column.

Table 12: Scree Plot.



The horizontal axis in this table represents the serial number of the factor; the sequence numbers of factors are arranged according to the size of their eigenvalues. The factor of the maximum eigenvalue is arranged on the far left. The data of the ordinate is the characteristic value. The characteristic curve suddenly becomes relatively flat at a certain inflection point (the eigenvalue of a factor), which means that the eigenvalues of the factors arranged behind are relatively close, and have little effect in the process of simplifying variables. Therefore, the number of factors corresponding to the inflection point can be used as a reference for the number of factor extraction. Of course, the above specific factor extraction criteria should be comprehensively considered in practical application according to the research purpose, relevant experience and theoretical assumptions [5] [6].

4. CONCLUSIONS

Firstly, as mentioned above, factor analysis has many advantages for multifactor data analysis, but not all data sets are suitable for the analysis algorithm. In practical problems, it is often encountered to study multiple indicators. In most cases, there is a certain correlation between different indicators. Because there are many indicators, and there is a certain correlation between indicators, this is bound to increase the complexity of the analysis. Generally speaking, we pay attention to and carry out principal component analysis, factor analysis and other algorithms to solve such problems. In fact, in more occasions, we often combine multiple algorithms to better solve practical problems.

Secondly, by analyzing their academic achievements, we preliminarily identify the influencing factors that can be divided into small course groups, to provide a meaningful basis for later teaching reform program. Of course, different students in different periods of time and in different groups will inevitably increase the complexity of analyzing problems. In addition, more data samples with more universal significance should be selected to participate in the calculation.

ACKNOWLEDGEMENTS

This work was supported by the second batch of high-level specialty group construction projects of Guangdong Higher Vocational Colleges under grant no. GSPZYQ2021052.

REFERENCES

[1] *HaiBin Wang*, 2021. *BASE AND APPLICATION OF ARTIFICIAL INTERLLIGENCE*, China Electronic Industry Press, ISBN: 987-7-121-41296-7

[2] JunJun Cen 2021. *CHINA BIG DATA APPLICATION DEVELOPMENT REPORT NO. 5 (2021)*, China Social Sciences Literature Press. ISBN: 9787520191531

[3] Li Lu, 2021. *MATHEMATICAL FOUNDATIONS OF DATA SCIENCE,* China People's Posts and Telecommunications Press, ISBN: 978-7-115-55288-1

[4] Phil Simon, 2014. *BIG DATA APPLICATION*, China People's Posts and Telecommunications Press, ISBN: 9787115365262

[5] YiHua Huang, 2014. *UNDERSTANDING BIG DATA: BIG DATA PROCESSING AND PROGRAMMING*, China Machine Press, ISBN: 9787111473251

[6] ZhiHua Zhou, 2016. *MACHINE LEARNING*, China QingHua University Press, ISBN: 9787302423287