Frontiers in Sport Research
ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

Application of Principal Component Analysis in Teaching Evaluation

MengYi

Business School/Lingnan Normal University, Zhanjiang 524048, China

ABSTRACT. *Principal components analysis (PCA) is a multivariate data analysis technique whose main purpose is to reduce the dimension of the observations and thus simplify the analysis and interpretation of data, as well as facilitate the construction of predictive models. In this paper, principal component analysis is used to analyze the indicators of teachers' teaching ability, calculate the scores of each teacher in each principal component and comprehensive scores by using SPSS. It realizes the comprehensive ranking of teachers' teaching evaluation, thus providing quantitative basis for each link of teachers' teaching evaluation.*

KEYWORDS: *Principal Component; Eigen Values; Variance; Factor Matrix; Teaching Evaluation*

1. Introduction

As a key link and important mechanism of modern education and teaching management, teaching level evaluation is an important means to ensure the development of teaching. It has many functions, such as guidance, improvement, judgment and incentive, and plays an irreplaceable role. The difficulty of teacher' s teaching evaluation lies in the design of evaluation index and the process of quantification. On the one hand, in order to avoid missing important information and consider as many indicators as possible, on the other hand, considering the increase of indicators increases the complexity of the problem. At the same time, because each indicator is a reflection of the same thing, it inevitably causes a large number of overlapping information, which sometimes even obliterates the real characteristics and internal laws of things. Moreover, the evaluation process is vulnerable to the intervention of the evaluators, which makes the evaluation results deviate.

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

Principal Component Analysis (PCA) is commonly used in economic research, as well as in other fields of activity. When faced with the complexity of economic and financial processes, researchers have to analyze a large number of variables (or indicators), fact which often proves to be troublesome because it is difficult to collect such a large amount of data and perform calculations on it. In addition, there is a good chance that the initial data is powerfully correlated; therefore, the signification of variables is seriously diminished and it is virtually impossible to establish causal relationships between variables. Researchers thus require a simple, yet powerful analytical tool to solve these problems and perform a coherent and conclusive analysis. This tool is PCA.

PCA is a method of dimension reduction, which can transform multiple indicators into several comprehensive indicators on the premise of losing little information. Generally, the synthetic index generated by transformation is called principal component, in which each principal component is a linear combination of the original variables, and each principal component is not correlated with each other, which makes the principal component have better performance than the original variable. The weights of principal components are automatically generated through PCA, which not only reduces the workload, but also reduces the interference of human factors in the evaluation process. Therefore, the principal component analysis of the index data reflecting teachers' teaching level can be used to establish an objective evaluation model of teachers' teaching as a basis for teachers' teaching evaluation.

2. mathematical model

2.1 Principal Component Definition

Principal component analysis is a statistical multivariate technique that uses orthogonal transformation to convert several correlated observed variables into a smaller number of linearly uncorrelated variables known as principal components [1]. The first principal component accounts for the highest variation in data and the subsequent component has the highest variance possible, as long as it is orthogonal to the preceding component. The number of p original features is reduced into a few unobserved variables, k known as principal components. The principal components (k) account for the maximum variance such that k ≤ p [2]. when the following exist:

$$\begin{cases} Y_1 = u_{11}X_1 + u_{21}X_2 + \cdots + u_{m1}X_p \\ Y_2 = u_{12}X_1 + u_{22}X_2 + \cdots + u_{m2}X_p \\ Y_3 = u_{13}X_1 + u_{23}X_2 + \cdots + u_{m3}X_p \\ \cdots\cdots\cdots\cdots\cdots \\ Y_m = u_{1p}X_1 + u_{2p}X_2 + \cdots + u_{mp}X_p \end{cases} \quad (1)$$

Then

$$\mathrm{Y} = \begin{pmatrix} u_{11} & \cdots & u_{m1} \\ \vdots & \ddots & \vdots \\ u_{1p} & \cdots & u_{mp} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = AX \quad (2)$$

2.2 Steps of PCA

(1) Selection of p-dimensional random vectors and n samples based on research questions $(n < p)$,then standardization of raw data, using equation (3).

$$x_{ij}^{*} = \frac{(x_{ij} - \overline{x_j})}{\sqrt{\sigma_{ii}}}, i = 1,\cdots,n; j = 1,\cdots,p \quad (3)$$

Where x_{ij} is the original data of the j-th index of the sample i-th.$\overline{x_j}$ is the average value of the original data for the j-th index of all samples; ;σ_{ii} is the variance of the original data of the j-th index of all samples. The normalized data matrix Z of the original data matrix is obtained.

(2) Calculating correlation coefficient matrix for standardized matrix Z, using equation

$$\mathrm{R} = [r_{ij}]_p xp = \frac{Z^T Z}{n-1} \quad (4)$$

Where r_{ij} is the correlation coefficient of the original variable x_i and x_j.

(3) Computing the characteristic equation $|R - \lambda I_p| = 0$ of sample correlation matrix R to get p characteristic roots.

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

(4) Selection of Principal Components

Determining principal components by cumulative contribution rate of variance. Usually the cumulative contribution rate is more than 85%,as equation (5).

$$\sum_{i=1}^{m} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i \geq 85\% \tag{5}$$

At this point, it is considered that the information of the original variable can be adequately reflected, and the corresponding M is the first m principal component extracted.

(5) Comprehensive Evaluation of m Principal Components

The final evaluation value is obtained by weighted summation of M principal components. Weight is the variance contribution rate for each principal component.

3. An example of application

In this paper,we will use the PCA evaluation of teaching quality of 12 mathematics teachers of a senior high school in the first term of 2017-2018.We have selected a number of nine(9) relevant variables: moral integrity and rigorous scholarship(X_1), content science and rigorous system(X_2),scientific research and teaching reformation(X_3),fostering ability and improving quality(X_4),inspire innovation(X_5),harmonious relationship with students(X_6),caring for students(X_7),open classroom teaching(X_8),language teacher humor(X_9).The original data is presented in Table 1.

Table 1 original data table

	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	16.00	16.00	13.00	18.00	16.00	15.00	14.00	16.00	16.00
2	18.00	19.00	15.00	16.00	18.00	18.00	18.00	17.00	19.00
3	17.00	17.00	14.00	14.00	17.00	17.00	20.00	14.00	15.00
4	17.00	17.00	17.00	16.00	18.00	18.00	16.00	20.00	14.00

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

5	16.00	15.00	17.00	17.00	18.00	18.00	19.00	16.00	19.00
6	15.00	17.00	16.00	17.00	18.00	18.00	15.00	19.00	16.00
7	17.00	16.00	16.00	18.00	18.00	18.00	17.00	15.00	18.00
8	20.00	18.00	16.00	20.00	15.00	15.00	19.00	14.00	17.00
9	14.00	16.00	18.00	17.00	19.00	19.00	18.00	17.00	18.00
10	16.00	16.00	15.00	19.00	18.00	18.00	18.00	15.00	14.00
11	18.00	19.00	16.00	14.00	14.00	14.00	17.00	16.00	13.00
12	19.00	15.00	15.00	18.00	16.00	16.00	18.00	19.00	17.00

In order to remove the unit limitation of data and transform it into dimensionless pure values, so as to facilitate the comparison and weighting of indicators in different units or scales, scoring needs to be standardized.

The high levels of standard deviations can be explained taking into account the powerful correlations between the original variables.It is presented in Table 2.

Table2 correlation matrix of standardized data

	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1.000	0.390	-0.251	0.120	-0.682	-0.412	0.327	-0.178	-0.045
X2	0.390	1.000	-0.049	-0.407	-0.390	-0.338	0.010	-0.119	-0.266
X3	-0.251	-0.049	1.000	0.000	0.367	0.446	0.179	0.302	0.208
X4	0.120	-0.407	0.000	1.000	0.067	-0.041	-0.087	-0.127	0.324
X5	-0.682	-0.390	0.367	0.067	1.000	0.800	0.020	0.260	0.410
X6	-0.412	-0.338	0.446	-0.041	0.800	1.000	0.385	0.241	0.206
X7	0.327	0.010	0.179	-0.087	0.020	0.385	1.000	-0.492	0.243
X	-0.17	-0.119	0.302	-0.12	0.260	0.241	-0.49	1.000	-0.04

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

8	8			7			2		6
X 9	-0.04 5	-0.26 6	0.208	0.324	0.410	0.206	0.243	-0.04 6	1.000

Now face the problem of information redundancy due to powerful correlations between the variables. Therefore, a primary purpose of PCA is to eliminate information redundancy, along with dimensionality reduction.

Table 3 the eigenvalues greater than one of the correlation matrix

component	total	% of variance	Cumulative %
1	3.005	33.392	33.392
2	0.748	19.419	52.811
3	1.459	16.216	69.028
4	0.975	10.834	79.862
5	0.677	7.525	87.387
6	0.588	6.535	93.922
7	0.437	4.853	98.776
8	0.069	0.771	99.546
9	0.041	0.454	100.000

Table 3 shows that when the cumulative contribution rate of the first five principal components$y_1, y_2, y_3, y_4, y_5$is 87.387% or more than 85%, the first five principal components replace the original nine indicators and play a role in dimensionality reduction.

Table 4 rotated factor matrix

	factor 1	factor 2	factor 3	factor 4	factor 5
X_1	-0.689	0.412	0.049	0.449	-0.218
X_2	-0.574	-0.114	0.551	0.217	0.386
X_3	0.560	0.030	0.367	0.528	-0.043
X_4	0.135	0.361	-0.775	0.251	-0.163
X_5	0.922	-0.022	0.045	-0.149	0.173
X_6	0.829	0.158	0.347	-0.085	-0.251

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

X_7	0.062	0.832	0.480	-0.077	-0.193
X_8	0.364	-0.665	0.013	0.516	-0.172
X_9	0.444	0.524	-0.255	0.289	0.542

Divide each element of Column I of Component Matrix by $\sqrt{\lambda_i}$, the coefficient of principal component i is obtained. The linear combinations of y_1, y_2, y_3, y_4, y_5 are obtained as follows:

$$y_1 = -0.397463x_1^* - 0.331123x_2^* + 0.323047x_3^* + 0.077877x_4^* + 0.531874x_5^* + 0.478225x_6^* + 0.035766x_7^* + 0.209981x_8^* + 0.25613x_9^*$$

$$y_2 = 0.311621x_1^* - 0.086225x_2^* + 0.022691x_3^* + 0.273046x_4^* - 0.01664x_5^* + 0.119505x_6^* + 0.629292x_7^* - 0.50298x_8^* + 0.396333x_9^*$$

$$y_3 = 0.040567x_1^* + 0.456167x_2^* + 0.303835x_3^* - 0.641614x_4^* + 0.037255x_5^* + 0.287278x_6^* + 0.397387x_7^* + 0.010763x_8^* - 0.186275x_9^*$$

$$y_4 = 0.45472x_1^* + 0.219764x_2^* + 0.534726x_3^* + 0.254198x_4^* - 0.150898x_5^* - 0.086083x_6^* - 0.077981x_7^* + 0.522573x_8^* + 0.292682x_9^*$$

$$y_5 = -0.264949x_1^* + 0.469129x_2^* - 0.052261x_3^* - 0.198104x_4^* + 0.210258x_5^* - 0.305056x_6^* - 0.234565x_7^* - 0.209042x_8^* + 0.658726x_9^*$$

Calculate the principal components of each teacher. Then, the proportion of variance contribution rate of each principal component in the cumulative contribution rate of five principal components is taken as the weight to sum up, and the comprehensive score Y of each teacher is obtained.

$$Y = (0.334y_1 + 0.194y_2 + 0.162y_3 + 0.108y_4 + 0.075y_5)/0.87387$$

According to the principal component synthesis model, the comprehensive principal component values can be calculated and sorted, and then the comprehensive evaluation and comparison of the school teachers can be made. The results are shown in Table 5

Table 5 the factor scores

Teacher number	Y	rank

1	-1.521	12
2	0.467	3
3	-0.101	9
4	0.034	6
5	0.955	2
6	0.016	7
7	0.389	4
8	-0.368	10
9	1.282	1
10	0.038	5
11	-1.099	11
12	-0.091	8

From table 5, we can see that the teachers with higher scores are No. 9, No. 5 and No. 2. The school can give proper encouragement to these teachers and let other teachers watch their classroom teaching to improve the teaching quality of the school. From the table, we can also see that the score of the fifth principal component is low. School leaders can help the group to form their unique teaching style by organizing training and other measures.

4. Conclusion

Based on the original teaching evaluation data,this paper obtains a comprehensive evaluation method of teachers' teaching quality by using principal component analysis. By calculating the scores of each principal component and making an objective evaluation of teachers' teaching quality, teachers can basically accept the evaluation results. By calculating the comprehensive score, each teacher can see his strengths and weaknesses, and can improve his teaching and improve the quality of teaching. School leaders can also reward excellent teachers and give some suggestions to teachers at the average level, so as to improve the overall level of teachers and make teaching evaluation play a practical role in improving the quality of teaching.

ISSN 2618-1576 Vol. 1, Issue 1: 34-42, DOI: 10.25236/FSST.080106

References

[1] Jamilah, M.; Zakaria, A., Md. Shakaff, A.Y., Idayu, N.; Hamid, H., Subari, N., Mohamad, J(2012). Principal Component Analysis-A Realization of Classification Success in Multi Sensor Data Fusion. In Principal Component Analysis-Engineering Applications; InTech: Philadelphia, PA, USA, 2012; pp. 1–25.

[2] Paul, L.C., Suman, A.A., Sultan, N(2013). Methodological Analysis of Principal Component Analysis (PCA) Method. Int. J. Comput. Eng. Manag, no.16, pp.32–38.

[3] Simar, L(2004). Applied Multivariate Statistical Anaylsis, Springer.

[4] Dan Armeanu, Leonard Lache(2006). Application of the Model of Principal Components analysis on Romanian insurance market. theoretical and applied economics, pp. 11-20.

[5] Ramsay, J. O. (2011), 'Functional data analysis', McGill University, Canada. Accessed 16th March.

[6] Shang, H. L. & Hyndman, R. J. (2011), 'Nonparametric time series forecasting with dynamic updating', Mathematics and Computers in Simulation, vol. 81, no.7, 1310–1324.

[7] Maindonald and Braun(2007). Data Analysis and Graphics Using R: An Example-based Approach. Cambridge University Press.