Based on Principal Component Analysis (Pca) of the influence factors of sports dance development research

Xiaoyun Shi[a]

Department of P.E., Guangdong University of Foreign Studies, Guangzhou 510006, China

ABSTRACT. In the rapid development of material civilization and spiritual civilization of human beings today, sports dance has become a hot sport for people. Due to China's vast territory, large population, the development of sports dance under the influence of many factors, such as the limited economic level, people's acceptance level is not enough and so on a series of problems. Some problems about the development of the sports dance in research, these factors can cause too much inconvenience. In this paper, the development of the sports dance as the research object, to teachers, referees, athletes based on the questionnaire investigation of the three kinds of people, in view of the economic level, competition system, the scientific research level ten three influence factors such as principal component analysis. Analysis results show that there are two principal components can replacement ten three influence factors. The two principal components for 10 linear combination of the three factors, the difference of both, the weight of each factor is different.

1 Introduction

In recent years, sports dance in many colleges and universities, has become many college students one of the most popular sport. Today, the development of sports dance is still not balance. This problem has sparked the factors for restricting sports dance development thinking [1-3].

In hebei province in 2010, into the sea in the study of sports dance development present situation investigation and the development countermeasures, for carry out the current situation of sports dance in hebei province were analyzed, and the results show that in terms of sports dance, male athletes in hebei province is less, the cultural level of the coaches and referees need to improve. Competition organization chaos, poor management, these factors severely limited the development of sports dances in hebei province [4-7]. In 2004, Schwartz rs front and others in the research on developing countermeasures of sports dance in colleges and universities, based on the current development situation of sports dance in colleges and universities has carried on the comprehensive analysis. Results show that the varying degrees of universities to carry out the sports dance, student’s interest and motivation for sports dance project are relatively high, but the overall performance for backward theoretical knowledge and scientific research. To this, the author puts forward the improving measures. For sports dance to better development, the author points out that should promote the sports dance option lessons and the club. In 2012, feibi "the research of the university sports dance education project", and other research methods, using the teaching experiment method in view of the institutions of higher learning sports dance teaching to sports dance education problems are analyzed, and the results show that the effect of sports dance education not only is to make students master the basic action of dance sport, more can let the student's quality was improved. The teaching of sports dance project conforms to the trend of the Times is worth promoting various colleges and universities. 2003, Zhao Li in the our country sports dance development present situation investigation and the countermeasure research, using various research methods, studies on influence factors of sports dance, the results show that the lack of sports dance in China's own teaching material system, the scientific research level is low, the referee's rating system is not standard, the referee's cultural level is generally low [5-8].

This article will promote situation of sports dance in colleges and universities as the research object, in view of the teachers, referees, athletes situations are analyzed, and the conclusion.

2 Model establishment

The model study influencing factors of sports dance, with the percentage accounted for the various factors of teachers, the percentage accounted for the referee and athletes of the percent as the basis, will be reduced to a few of influential factors of influencing factors, to facilitate such problems later.

[a] Corresponding author: XiaoyunShi55@sina.com

Table 1. Original data table.

Influence factor	Teacher	Percentage %	Referee	Percentage %	Athlete	Percentage %	Rank
Economic level constraints	132	80	22	43	630	61	8
Competition and referee system	140	84	37	73	932	90	3
Teachers level	154	93	51	100	1000	97	1
Funding issue	93	56	43	84	403	39	11
Related to non-Olympic Games events	103	62	41	80	726	70	4
Fewer international exchange	149	90	47	92	955	92	2
Public concept	57	34	30	59	745	72	6
Scientific research level	73	44	25	49	352	34	12
Sports level	62	37	32	63	601	58	9
Mass media influence	88	53	30	59	722	70	5
Disordered organizational management	77	46	24	47	561	54	10
Field facilities	96	58	17	33	677	65	7
Others	26	16	6	12	135	13	13

Table 1 data is from "Chinese sports dance development status investigation and countermeasure research".

Main thought of principal component analysis is variable's dimension reduction. It is a statistical analysis method that transforms multiple variables into fewer main variables. It generally is used to data compression, system evaluation, regression analysis and weighted analysis so on.

2.1 Principal component analysis method

Main way of principal component analysis is reducing dimension of variables, which is recombining original many variables with correlation into a group of uncorrelated variables to replace original variables. Therefore, we can pay attention to every time observation's variables that have maximum variation, to every time observation's small changed variables that can be used as constant to process and get rid of them, so that it reduces variables number in problem that needs to be considered [9,10].

Assume that there is m pieces of original indicators to do principal component analysis, which are recorded as $x_1, x_2, \cdots, x_m$, now it has n pieces of samples, corresponding observation value is $x_{ik}(i=1,2,\cdots,n)$, and $k=1,2,\cdots,m$ takes standardization transformation, and then transform x_k into x_k^*, that:

$$x_k^* = \frac{x_k - \overline{x_k}}{s_k}, \quad k=1,2,\cdots,m \tag{1}$$

Among them, $\overline{x_k}$ and s_k are respectively x_k average number and standard deviation, x_k^* average number is 0, standard deviation is 1.

According to each sample original indicator observation value x_{ik} or after standardization observation value x_{ik}^*, it solves coefficient b_{kj}, establish indicator x_k^* that is transformed through standardization to express comprehensive indicator z_j equation $z_j = \sum_k b_{kj} x_k^*$, which can also establish equation that uses original indicator x_k to express comprehensive indicator z_j:

$$z_j = \sum_k \tilde{b}_{kj}\, x_k^* + a_j \tag{2}$$

There are two requirements on defining b_{kj}:

(1) Comprehensive indicators are mutual independent from each other or uncorrelated.

(2) Every comprehensive indicator reflected each sample gross information content is equal to corresponding feature vector (comprehensive indicator coefficient) feature values. In general, it is required that selected comprehensive indicator feature vales contribution ratios sum to be above 80%.

2.2 Principal component analysis general steps

(1) According to observed data, calculate $\overline{x_k}$ and $s_k(k, j=1,2,\cdots,m)$.

(2) By correlation coefficient matrix R, it can get feature value $\lambda_j\ (j=1,2,\cdots,m)$ and each principal component variance contribution、 contribution ratio and accumulative contribution ratio, and define principal component reserved number p with accumulative contribution ratio as evidence.

(3) m pieces of basic equations are as following:

$$\begin{cases} r_{11}x_1^{(j)} + r_{12}x_2^{(j)} + \cdots + r_{1m}x_m^{(j)} = \lambda_j x_1^{(j)} \\ r_{21}x_1^{(j)} + r_{22}x_2^{(j)} + \cdots + r_{2m}x_m^{(j)} = \lambda_j x_2^{(j)} \\ \cdots \\ r_{m1}x_1^{(j)} + r_{m2}x_2^{(j)} + \cdots + r_{mm}x_m^{(j)} = \lambda_j x_m^{(j)} \end{cases} \tag{3}$$

Among them, $j=1,2,\cdots,m$.

Proceed with Schmidt orthogonalization, for every λ_i, solve its basic equations solution $x_1^{(j)}$, $x_2^{(j)}$, …, $x_m^{(j)}$ $(j=1,2,\cdots,m)$, and then let:

$$b_{kj}=\frac{x_k^{(j)}}{\sqrt{\sum_k\left(x_k^{(j)}\right)^2}} \tag{4}$$

It can get expressed by x_1^*, x_2^*, …, x_m^* principal component $z_j=\sum_k b_{kj}x_k^*$, or input $x_k^*=\frac{x_k-\overline{x_k}}{s_k}$ and then get expressed by x_1, x_2, …, x_m principal component $z_j=\sum_k \tilde{b}_{kj}x_k^*+a_j$.

(4) Input x_1, x_2, …, x_m observed values into principal component expressions, calculate each component value.

(5) Calculate original indicator and principal component correlation coefficient that is also factor loading that use it to explain principal component significances.

Table 2. Variables communalities table.

	Initial	Extract
Economic level constraints	1.000	1.000
Competition and referee system	1.000	1.000
Teachers level	1.000	1.000
Funding	1.000	1.000
With non-Olympic Games events	1.000	1.000
International exchange	1.000	1.000
Public concept	1.000	1.000
Scientific research level	1.000	1.000
Sports level	1.000	1.000
Mass media influence	1.000	1.000
Disordered organizational management	1.000	1.000
Field facilities	1.000	1.000
Others	1.000	1.000

Extract method: principal component analysis.

Table 2 represents every variable communalities result. Table 2's left side represents every variable explainable variance from all factors, while the right side represents variable communalities. From table data, we can see that variable communalities are 1 that are very high, which shows most information in variables can be extracted by factors; it shows the analysis is valid.

Table 3. Factor contribution ratio table.

Component	Initial feature value			Extract squares sum and load in			Rotate squares sum and load in		
	Total	Variance %	Accumulation %	Total	Variance %	Accumulation %	Total	Variance %	Accumulation %
1	7.568	58.212	58.21	7.568	58.212	58.212	7.411	57.010	57.01
2	5.432	41.788	100.00	5.432	41.788	100.00	5.589	42.990	100.00
3	5.919E-16	4.553E-15	100.00						
4	2.462E-16	1.894E-15	100.00						
5	2.419E-16	1.861E-15	100.00						
6	1.329E-16	1.022E-15	100.00						
7	9.404E-17	7.234E-16	100.00						
8	-1.660E-17	-1.277E-16	100.00						
9	-1.103E-16	-8.484E-16	100.00						
10	-1.479E-16	-1.138E-15	100.00						
11	-1.711E-16	-1.316E-15	100.00						
12	-2.965E-16	-2.281E-15	100.00						
13	-3.367E-16	-2.590E-15	100.00						

Extract method: principal component analysis.

In Table 3, accumulation items' data indicates percentage of total feature values. From table data, it can easily see that factor 1 and factor 2 feature values are above 1, and the two factors' feature values sum are 100% of total feature values. Therefore, we use factor 1 and factor 2 as main factors.

Data in Table 4 indicates factor loading value after using Kaiser standard orthogonal rotation. By such rotating, every factor's significance is relative clear. From the table, it can see that two main factors are extracted. From Figure 1, it can more intuitive indicate.

Table 4. Rotational factor loading table.

	Component	
	1	2
x1	-.939	.345
x2	-.318	.948
x3	.960	-.282
x4	.292	-.956
x5	.908	-.418
x6	.988	.155
x7	.878	.479
x8	-.034	-.999
x9	1.000	-.026
x10	.662	.750
x11	.464	.886
x12	-.453	.892
x13	-.996	.087

Extract method: principal component.
Rotational method: Orthogonal rotation method with Kaiser standardization.

a. Convergent after three times' iterating of rotation.

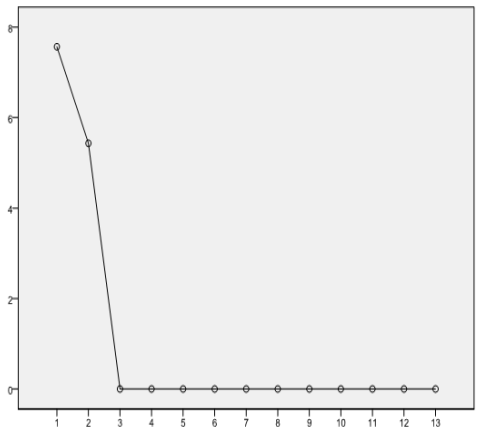

Figure 1. Scree plot.

Figure 1 is feature values' scree plot. In general, the figure shows big factor steep slope and surplus factor gentle tail has obvious interruption. Generally selected main factors are in the very steeply slope, and factors lie in gentle slope have insignificant effects on total. From Figure 1, it is clear that the former two factors are in the relative steeply slope, and starts from the third factor, the slope turns to be gentle, while starts from the third factor, the slope is nearly zero, therefore select two factors as comprehensive factors.

From Figure 2, it is clear that principal component analysis totally extracts two main factors this time, x_8, x_9 that get closer to coordinate axis have big factor loading and explanatory ability is relative strong.

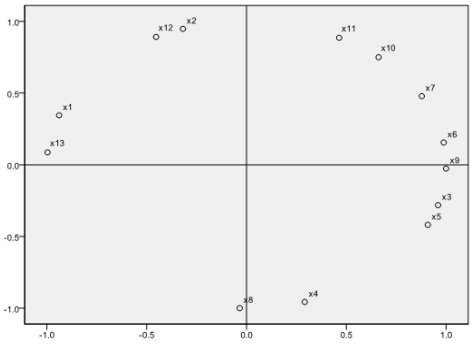

Figure 2. The view of the rotating components of the space.

After defining main factors amount, it should calculate feature vectors, feature vectors amount is the same as main factors amount. Feature vector matrix is as Table 5 show.

F1	F2
-1.58831	-0.42853
-1.01902	0.98213
1.03123	-0.32397
0.42783	-1.62071
1.01548	-0.5518
0.91186	0.3439
0.65869	0.78239
0.03037	-1.82201
0.99056	0.0755
0.28878	1.09916
-0.00898	1.21932
-1.16855	0.84169
-1.56994	-0.59705

Table 5: Feature vector matrix

By Table 5 feature vectors, it can get principal component computational formula:

$$z_1 = -1.58831x_1 - 1.01902x_2 + 1.03123x_3 + 0.42783x_4 + 1.01548x_5 + 0.91186x_6 + 0.65869x_7 + 0.03037x_8 + 0.99056x_9 + 0.28878x_{10} - 0.00898x_{11} - 1.16855x_{12} - 1.56994x_{13} \quad (5)$$

$$z_2 = -0.42853x_1 + 0.98213x_2 - 0.32397x_3 - 1.62071x_4 - 0.5518x_5 + 0.3439x_6 + 0.78239x_7 - 1.82201x_8 + 0.0755x_9 + 1.09916x_{10} + 1.21932x_{11} + 0.84169x_{12} - 0.59705x_{13} \quad (6)$$

Respectively input original data into formula (5) (6), it can get data as Table 6.

Table 6. Main components variables.

z_1	z_2
0.97268	0.268955
0.853204	0.2337
0.24219	1.49121

Table 6 is two main components variables after factor analysis. z_1, z_2 are economic restriction level and other thirteen factors' linear combinations. That is principal component analysis reducing original thirteen factors into two factors so that is easy to research, but physical significances after factor dimensions reducing is hard to define.

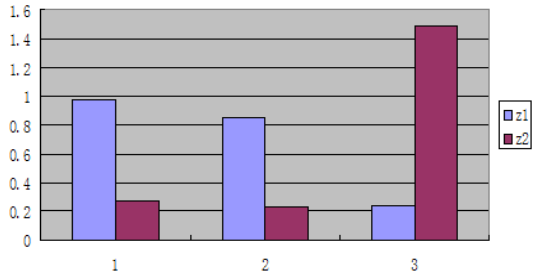

Figure 4. Principal component comparison chart.

Figure 4 is principal component comparison chart. From Figure 4, it is clear that to teachers and referees, principal component 1 occupied weight is bigger than that of principal component 2, to athletes; principal component 1 occupied weight is smaller than that of principal component 2.

3 Conclusion

Using the ideas of dimension reduction with a few variables instead of the original multiple variables, these a few variables can reflect most of the original data information. In addition, the model is more focused on the comprehensive evaluation of information. The method also has certain drawbacks, for example, when the principal component factor loading of plus or minus signs exist at the same time, the meaning of the evaluation function is not clear, named low clarity, involves only a group of related variables. The main component of the model is composed of linear combination of the original factors, so the practical significance of the principal component is difficult to determine, just have played an important role in dimension reduction. Principal component analysis of applications is very extensive, for example, "regional water resources carrying capacity problem", "urban land evaluation", "3 g network comprehensive performance evaluation", and other aspects of the problem analysis. In this paper, applying principal component analysis (pca) to study the limiting factor of sports dance development, successfully reduced ten three influence factors to the two principal components, convenient for later research of this problem.

References

1. X.G. Wang, Research on Development **5** (2007).
2. N.Y. Chen, Journal of Yunnan Finance and Trade Institute **21** (2005).
3. C.M. Liu, B. Guan, Journal of Tianjin University (Social Sciences) **11** (2009).
4. L.W. Fani, J.Y. Liang, Value Engineering **25** (2006).
5. Y.K. Sun, Z. Lao, Journal of Anhui University (Philosophy & Social Sciences) **30** (2006).
6. J.H. Lu, Journal of Anhui University (Philosophy & Social Sciences) **30** (2006).
7. X.Y. Xu, Journal of Anhui University (Philosophy & Social Sciences) **30** (2006).
8. B.M. Su, Z.X. Liu, M. Chang, Journal of Jinan University **12** (2002).
9. K. Rokni, A. Ahmad, K. Solaimani, Journal of Coastal Research 32 (2016).
10. H. Huang, Z.L. Liu, H.B. Liu, Ebm 2010: International Conference on Engineering And Business Management **1-8** (2010).