Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):1910-1917

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Research of statistical index of men's basketball on the basis of data character analysis

Wenhui Lv

Chuzhou University Institute of Physical Culture, Chuzhou, Anhui, China

---

ABSTRACT

Concerning that technical statistical indicators in a basketball match have some specific features, it will thus be more scientific and effective to analyze a basketball match from the perspective of indicators' features and offer pertinent trainings to basketball teams. Through studying technical statistical indicators in the World Championships in 2006 and 2010, this paper aims to explore the trend of indicators' features so as to find an effective way to carry out pertinent trainings for the World Championship. To begin with, 12 technical statistical indicators in every match among 24 teams respectively in 2006 and 2010 have been collected on the official website of World Championship, thus laying the basis of data analysis of technical statistical indicators of basketball. Then, this paper theoretically refers to the principle and steps of the principal factor analysis and cluster analysis to analyze data features. Moreover, SPSS 19.0 is adopted to analyze technical statistical indicators in World Championships of 2006 and 2010 and the results of the principal factor analysis and cluster analysis is achieved. Eigenvalue, square, square and load, rotating square and load are presented in either a table or a figure, a dendrogram is drawn by distance, and technical statistical indicators are classified into five categories and five principal factors are extracted. Finally, according to the constitution of five principal factors and the percentage of the variance of initial eigenvalue, the data of 2006 is compared with that of 2010. After analyzing the data feature differences, it can be concluded that five principal factors of the former World Championship can be referred to for the training of the next one and it is highly credible.

Key words: Principal factor analysis; cluster analysis; SPSS 19.0 statistical software; technical statistical indicators of the basketball match

---

INTRODUCTION

There are many official technical statistical indictors of basketball match, which make it complicated for scholars and coaches to analyze the match. To reduce the number of indicators while covering the general technical characteristics at the same time, this paper makes a principal factor analysis and a cluster analysis of 12 technical statistical indicators and indexes of 24 teams with a view to laying the theoretical basis for an objective evaluation of basketball match and the changes of principal factors with the change of session.

Many scholars have made efforts to explore the characteristics of technical statistical indicators of basketball match: Li Huilin (2009) made a principal factor analysis and a cluster analysis of technical statistical indicators of Tokyo World Championship in 2006 and put forward his own opinions according to the influence of all the factors on the whole variables and the characteristics of Chinese men's basketball[1]; Lou Bigang (2013) made a horizontal and longitudinal comparison of 18 indicators between the Chinese team and top four teams in the 16th World Championship and drew a conclusion afterwards[2]; Yang Lei analyzed assist conditions of the Chinese Men's Basketball team in the tournament of the 30th Olympic Games and offered his suggestions on the improvement of the collaboration among team members as well as the shooting rate[3].

On the basis of precedent research, this paper studies the technical statistical indicators of World Championships in 2006 and 2010 with an aim of exploring the trend of data features of these two sessions as well as the principal factor characteristics of match analysis.

Tab. 1: Technical statistical indicators of World Championship

symbol	indicator name	symbol	indicator name	symbol	indicator name
I-1	Mean score	I-5	free throw times	I-9	assists
I-2	three-point shot times	I-6	free throw percentage	I-10	steals
I-3	three-point shots	I-7	backboard	I-11	turn over
I-4	three-point shot percentage	I-8	block shots	I-12	fouls

Tab. 2: Technical statistical indicators of 2006 World Championship

nation	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10	I-11	I-12
Spain	88.6	22.0	8.2	37.4	28.0	71.0	35.1	2.9	14.3	9.8	14.1	16.7
Greece	80.0	18.9	6.2	32.9	26.7	70.0	27.0	2.4	12.2	10.1	13.9	20.0
America	103.6	24.7	9.1	36.9	30.3	66.7	36.3	4.9	18.8	10.8	10.8	19.8
Argentina	86.8	25.4	9.0	35.4	20.2	73.6	37.8	2.1	18.1	7.1	13.3	20.0
France	68.4	19.1	5.2	27.3	24.7	63.5	38.3	3.7	10.8	7.2	14.2	19.6
Turkey	74.3	22.0	7.7	34.8	23.3	63.8	29.7	3.4	11.9	7.3	16.1	25.0
Republic of Lithuania	79.1	20.2	6.2	30.8	24.7	64.0	36.8	2.6	15.3	9.0	20.7	23.7
Germany	77.7	23.9	8.2	34.4	20.7	79.0	34.1	2.0	12.7	5.0	16.6	21.1
Australia	73.8	25.5	9.8	38.6	18.8	65.5	30.7	1.5	15.2	8.7	20.7	20.8
Slovenia	86.3	19.0	8.5	44.7	23.3	67.9	35.5	2.7	14.8	7.3	16.2	23.3
Angola	85.5	28.0	10.3	36.9	19.3	73.3	35.5	2.5	14.3	8.3	12.0	23.3
Serbia	80.7	18.2	7.3	40.4	18.2	75.2	33.0	5.0	12.8	7.3	13.3	23.0
Italy	75.7	24.8	8.8	35.6	20.3	66.4	31.8	2.0	14.7	7.5	12.0	25.0
Nigeria	74.7	18.7	5.5	29.5	23.3	59.3	34.8	2.3	10.0	9.0	11.7	20.3
New Zealand	67.8	28.7	8.2	28.5	20.0	65.0	28.5	1.0	14.3	8.5	16.2	23.7
China	81.3	21.8	8.2	37.4	22.5	80.7	31.7	4.3	13.3	3.0	17.0	22.3
Lebanon	71.4	24.8	7.0	28.2	18.4	73.9	31.6	2.6	9.2	6.6	19.2	18.8
Senegal	71.0	15.4	5.4	35.1	19.0	69.5	34.0	2.6	12.2	7.8	16.2	22.8
Puerto Rico	86.4	19.2	9.8	51.0	27.6	70.3	30.2	2.0	12.2	6.4	14.4	23.2
Venezuela	67.2	18.8	5.4	28.7	19.6	60.2	37.0	3.0	11.0	6.6	18.4	19.4
Panama	65.2	17.4	4.2	24.1	23.8	54.6	33.8	2.0	8.4	6.6	18.8	21.2
Japan	64.4	23.4	7.4	31.6	16.4	67.1	24.6	1.4	10.4	7.0	14.0	23.0
Qatar	62.0	25.8	9.2	35.7	0.0	58.2	29.8	1.6	13.0	7.8	24.6	20.2
Brazil	79.8	20.8	6.0	28.8	26.8	61.9	28.8	2.0	14.0	10.2	15.0	22.6

Tab. 3: Technical statistical indicators of 2010 World Championship

nation	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10	I-11	I-12
Spain	84.1	18.2	6.5	35.8	17.6	71.5	35.2	4.4	18.5	7.5	13.1	21.3
Greece	85.1	14.3	4.8	33.1	10.5	65.3	35.9	2.9	14.9	8.3	11.6	22.4
America	92.7	19.9	7.7	38.5	14.7	7.3	38.9	4.0	18.2	10.4	11.8	18.9
Argentina	83.2	15.9	6.2	38.7	17.4	72.9	30.5	1.3	15.4	7.3	10.0	19.1
France	71.3	9.9	3.8	37.8	10.6	74.6	28.8	3.0	15.6	7.0	15.1	20.8
Turkey	81.0	16.9	7.3	42.9	18.3	59.6	32.9	3.4	16.5	8.1	10.5	17.9
Republic of Lithuania	82.8	17.6	7.0	39.8	14.9	72.8	34.1	3.2	15.2	5.3	13.0	19.7
Germany	75.6	10.5	4.3	40.5	6.5	76.2	30.6	2.8	13.6	5.6	14.8	21.8
Australia	73.1	12.3	3.7	29.9	9.3	72.7	32.1	2.3	13.0	7.0	12.5	18.8
Slovenia	78.3	18.0	6.4	35.6	16.0	74.4	30.1	1.3	13.2	6.8	12.8	23.3
Angola	72.1	13.9	4.3	31.1	8.8	63.4	30.8	2.0	12.5	6.3	13.0	20.8
Serbia	90.9	18.0	7.3	40.3	17.3	74.1	34.5	1.8	17.0	7.5	13.3	21.0
Croatia	77.7	11.8	4.3	36.2	7.7	66.7	31.0	1.6	12.5	6.1	11.8	23.8
Canada	66.0	9.0	2.8	31.5	7.9	67.1	30.2	3.0	10.6	7.6	11.2	20.6
New Zealand	79.9	13.8	4.6	33.1	11.7	73.3	29.6	1.0	16.0	7.3	12.3	24.6
China	80.3	9.9	3.7	37.0	10.6	73.0	33.9	4.0	11.9	6.6	15.1	20.6
Lebanon	67.8	8.3	2.8	33.3	10.0	63.4	27.8	1.2	10.6	7.8	15.6	16.2
Russia	73.9	15.7	5.4	34.6	14.7	79.7	32.3	3.8	16.0	5.7	13.7	21.4
Puerto Rico	77.2	9.8	3.2	32.2	7.7	69.0	35.6	3.4	15.8	4.0	12.4	22.0
Côte d'Ivoire	66.8	9.7	2.9	30.2	7.9	66.5	33.4	4.8	10.6	8.4	13.8	20.6
Tunisia	60.0	9.1	2.6	28.4	7.7	68.1	31.0	3.2	8.4	6.8	13.2	16.0
Iran	60.2	7.6	1.8	24.1	6.8	70.2	30.0	3.2	7.8	7.6	17.0	16.6
Jordan	72.2	11.7	3.9	33.6	7.1	68.0	32.6	1.0	12.2	4.8	14.2	19.2
Brazil	80.9	10.8	4.3	39.5	10.6	73.7	28.3	1.6	14.3	8.2	11.3	20.3

2. OBJECT OF STUDY AND RESEARCH METHOD

2.1 Object of study

This paper studies 20 technical statistical indicators of the teams in two World Championships in 2006 and 2010 respectively. These 20 indicators are listed in the following Table 1, and 12 technical indicators of 24 teams are listed in Table 2 and 3 as follows:

2.2 Research method

Literature consultation: To lay theoretical basis for the data feature analysis of statistical indicators of basketball, 6 papers about technical indicators of basketball have been collected from relevant journals, and 10 papers about principal factor analysis of sporting index, 8 papers about SPSS applied into indicator analysis and 12 papers about the cluster analysis principle have also been collected.

Mathematics statistics: Official statistical data is processed to produce 20 indicators as Table 1 shows. Excel and SPSS are used to carry out principal factor and cluster analyses of these indicators to extract 20 typical factors among them according to these data features so as to simplify match analysis.

Principal factor analysis: To conduct a comprehensive, scientific and short-cut analysis of matches, it is necessary to consider the correlation among these indicators and reduce the number of factors. Therefore, the principal factor analysis is adopted in this paper.

Cluster analysis: To classify these data by category, this paper makes a cluster analysis of them through calculating their correlation coefficients and cluster coefficients according to the similarity principle of their properties and attributes.

3. PRINCIPLES OF PRINCIPAL FACTOR ANALYSIS AND CLUSTER ANALYSIS

3.1 The principle of factor analysis

In research, there are many factors which are relevant to the research objective and classified into either common factor or only factor. The former refers to the one that is common to all the original variables and can explain the correlation among them; whereas the latter refers to the one that is specific to initial variables and cannot be explained by the common factor. In analyzing initial variables and factors, it is necessary to extract the factor load represented by common factor correlation, thus the most common model of factor analysis is as Expression (1） illustrates:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + U_j\left(j = 1,2,\cdots,n\right) \tag{1}$$

In this expression, Z_j stands for the standard score of the j th variable, F_i refers to the common factor, m represents the number of common factors of all the variables, and U_j stands for the only factor of Z_j and a_{ji} represents the factor load.

Factors in Expression (1） can be understood as m coordinate axes in the high-dimensional space. a_{ji} refers to the factor load, which is the load of the j th initial variable in the i th factor. If Z_j is regarded as the standard regression coefficient of the m -dimensional factor space and U as the special factor which stands for the part of initial variable that can't be explained by factor and whose mean value is 0, therefore U can be regarded as the residual of the multivariable linear regression model.

In order to state the mathematical models of factor analysis more conveniently, this section is elaborated with two common elements extracted from three variables being the example. These three variables are represented by Z_1, Z_2, Z_3 and two common elements represented by F_1 and F_2. Then three variables can be presented by the linear combination represented by two common elements, which is shown in formula (2）:

$$\begin{cases} Z_1 = a_{11}F_1 + a_{12}F_2 + U_1 \\ Z_2 = a_{21}F_1 + a_{22}F_2 + U_2 \\ Z_3 = a_{31}F_1 + a_{32}F_2 + U_3 \end{cases} \tag{2}$$

Factor analysis aims to condense original variables and extract core variables. If factor analysis were to be used,

whether observation data is suitable for factor analysis would have to be determined in the first place. Then the common factor is extracted before the factor scores of individual samples are calculated.

STEP1. Using the four statistics provided by SPSS software can determine whether the observation data is suitable for factor analysis.

STEP2. Common factors are extracted.

STEP3. After common factors are obtained, each factor is analyzed so as to achieve the aim of research. In this step, three aspects including screening factor, sample size and the number of factors need to be taken into account to get the screen plot, its analysis result and ultimately the result report of factor analysis.

3.3 Principle of cluster analysis

Also referred to as group analysis, cluster analysis is a multivariate statistical method that classifies samples or indexes, that is, categorizing similar factors. Depending on the classification objects, it can be divided into sample cluster and variable cluster.

Sample cluster is called Q - type cluster in statistics and in SPSS terminology, it clusters observed quantity. Based on various characteristics of observed objects, it classifies the value of each variable that reflects the characteristics of observed objects. Variable cluster is called R-type cluster in statistics. As there are plenty of variables that reflect the characteristics of the same thing, it chooses some variables to study its certain aspect based on the research question. Since people's understanding of objective things is limited, it is often difficult to find out representative variables independent from one another, which affects the further knowledge and research of problems. Therefore, in general, variable cluster needs to be conducted before finding out independent and representative variables so that most information is not lost.

To classify samples and indexes in a scientific way, it is of necessity to study the relationship between samples. At the present stage, two most widely used methods are similarity coefficient and spatial distance. The larger the similarity coefficient, the closer to samples the property. Then similar samples are categorized into one group. The larger the spatial distance, the farther to samples the property. The samples can be categorized into one group when the distance is minimal. Frequently used distances consist of Minkowski distance, Mahalanobis distance and Canberra distance, as represented by formulas (3）, (4） and (5） respectively. Common similarity coefficients include cosine of the angle and correlation coefficient as presented by formulas (6） and (7） respectively.

$$d_{ij}(q)=\left(\sum_{a=1}^{p}\left|x_{ia}-x_{ja}\right|^{q}\right)^{\frac{1}{q}} \tag{3}$$

$$d_{ij}^{2}(M)=\left(X_i-X_j\right)'\sum{}^{-1}\left(X_i-X_j\right) \tag{4}$$

$\sum$ in formula (3） stands for the covariance matrix of indexes.

$$d_{ij}(L)=\frac{1}{p}\sum_{a=1}^{p}\frac{\left|x_{ia}-x_{ja}\right|}{x_{ia}+x_{ja}},i,\ j=1,2,\cdots,n \tag{5}$$

$$\cos\theta_{ij}=\sum_{a=1}^{p}x_{ia}x_{ja}X_i\Big/\sqrt{\sum_{a=1}^{p}x_{ia}^{2}\cdot\sum_{a=1}^{p}x_{ja}^{2}}\ ,-1\le\cos\theta_{ij}\le 1 \tag{6}$$

$$r_{ij}=\sum_{a=1}^{p}\left(x_{ia}-\overline{x}_i\right)\left(x_{ja}-\overline{x}_j\right)\Big/\sqrt{\sum_{a=1}^{p}\left(x_{ia}-\overline{x}_i\right)^{2}\cdot\sum_{a=1}^{p}\left(x_{ja}-\overline{x}_j\right)^{2}}\ ,-1\le r_{ij}\le 1 \tag{7}$$

This paper mainly takes advantage of the special case of $q=2$ in Minkowski distance, i.e. Squared Euclidean distance presented by formula 8.

$$d(x,y)=\sum_{i}\left(x_i-y_i\right)^{2} \tag{8}$$

4. ANALYSIS OF THE RESEARCH RESULTS OF DATA CHARACTERISTCIS IN BASKETBALL STATISTICAL INDEXES

4.1 Result of principle factor analysis

Principal factor dimensionality reduction analysis is conducted in SPSS19.0 software on 12 indexes involving 24 teams participating in World Basketball Championships of 2006 and 2010. The principal factor analysis table shown in figure 1 in which 5 main components are extracted based on the principle that the cumulative percentage of factor in total variance is greater than 80%.

Tab. 1: Principal factor analysis of technical indexes in 2010 World Basketball Championship

	Initial eigen-value			Extraction of the square and load			Rotation of the square and load
Component	Total	the percentage of variance	Accumulation %	Total	the percentage of variance	Accumulation%	Total
1	5.658	47.147	47.147	5.658	47.147	47.147	5.494
2	2.030	16.918	64.065	2.030	16.918	64.065	1.835
3	1.276	10.632	74.698	1.276	10.632	74.698	2.449
4	.932	7.691	82.388	.932	7.691	82.388	2.411
5	.583	4.856	87.244	.583	4.856	87.244	1.821
6	.564	4.701	91.946				
7	.493	4.105	96.051				
8	.212	1.769	97.820				
9	.187	1.561	99.381				
10	.050	.415	99.796				
11	.022	.184	99.980				
12	.002	.020	100.000				

The screen plot and rotating space composition diagram resulting from the number of components and engine-value in 2010 are shown in figure 2.

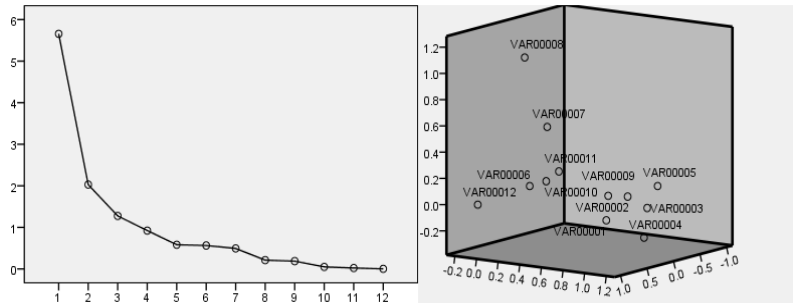

Fig. 2: The screen plot and rotating space composition diagram of the principal factor analysis

Being the principal factor analysis table of 2006, figure 3 extracts 5 main components based on the principle that the cumulative percentage of factor in total variance is greater than 80%.

Tab. 3: Principal factor analysis of technical indexes in 2006 World Basketball Championship

	Initial eigen-value			Extraction of the square and load			Rotation of the square and load
Component	Total	the percentage of variance;	Accumulation %	Total	the percentage of variance;	Accumulation%	Total
1	3.601	30.006	30.006	3.601	30.006	30.006	3.173
2	2.537	21.142	51.148	2.537	21.142	51.148	2.827
3	1.643	13.694	64.842	1.643	13.694	64.842	2.415
4	1.297	10.810	75.652	1.297	10.810	75.652	1.742
5	.853	7.111	82.763	.853	7.111	82.763	1.448
6	.699	5.826	88.589				
7	.505	4.211	92.800				
8							
9	.271	2.261	98.463				
10	.128	1.065	99.528				
11	.54	.446	99.975				
12	.003	.025	100.000				

The screen plot and rotating space composition diagram resulting from the number of components and engine-value in 2006 are shown in figure3.

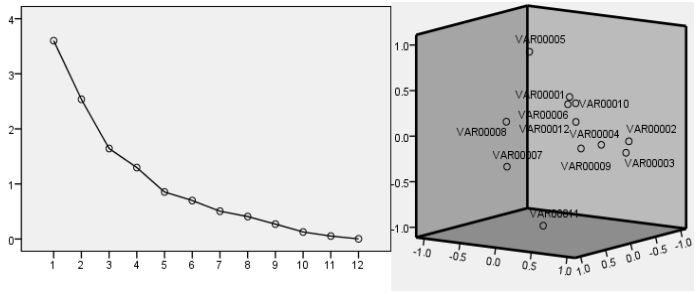

Fig. 3: The screen plot and rotating space composition diagram of the principal factor analysis

I-i represented by VAR0000i in figure 2 and 4 corresponds to the statistical indexes shown in table 1.

4.2 Result of cluster analysis

The SPSS19.0 software is used to conduct cluster analysis on 12 indexes of 2006 World Basketball Championship. Resulting cluster members and the tree diagram that uses average connection are shown in figure 5.

I-i represented by VAR0000i in figure 5 corresponds to the statistical indexes shown in table 1. 12 indexes of statistical indicators in basketball games are divided into five types. The first type is single index I-1 and the second type includes I-2, I-5, I-9, I-11and I-12. The third type is comprised of I-3, I-8 and I-10 that stand for the number of three-point shots, block shot and steal respectively. The fourth type consists of I-4 and I-7, representing three-point shot percentage and backboard respectively. The fifth type is also single index I-6 that refers to free throw percentage.

The SPSS19.0 software is used to conduct cluster analysis on 12 indexes of 2010 World Basketball Championship. Resulting cluster members and the tree diagram that uses average connection are shown in figure 6.

I-i represented by VAR0000i in figure 6 corresponds to the statistical indexes shown in table 1. 12 indexes of statistical indicators in basketball games are divided into five types. The first type is single index I-1 and the second type includes I-2, I-5, I-9, I-11and I-12. The third type is comprised of I-3, I-8 and I-10. The fourth type consists of I-4 and I-7. The fifth type is also single index I-6.

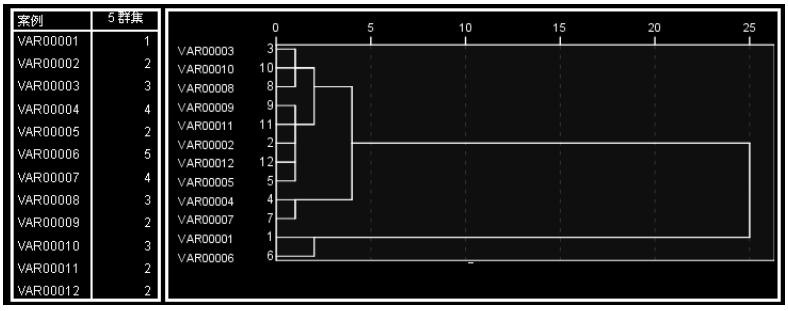

Fig. 5: Cluster members and the tree diagram presented by the cluster analysis result in 2006

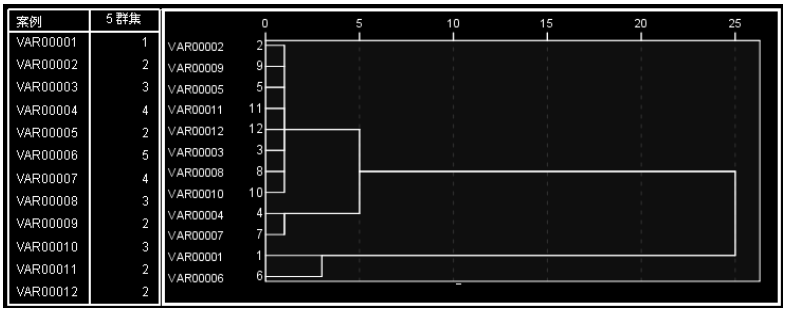

Fig. 6: Cluster members and the tree diagram presented by the cluster analysis result in 2010

In summary, the cluster analysis result and divided clusters of 2010 are completely the same with those of 2006. Cluster results are shown in table 4.

Tab. 4: Cluster analysis results

The first type	single indexI-1: points per game	The second type	I-2: the number of three-point shots I-5: free throw attempts I-9: assist I-11: fault I-12: foul	The third type	I-3: the number of three-point shots I-8: block shot I-10: steal	The fourth type	I-4: three-point shot percentage I-7: backboard	The fifth type	single index I-6:free throw percentage

4.3 Analysis of principle factor characteristics

Based on the principle that R^2 is the maximum in the same group as shown in formula (9), the factor with the maximum R^2 in the same group is taken as the principal factor representing the variables of this group.

$$R_j^2 = \frac{\sum r^2}{m_j} - 1 \qquad (9)$$

Then we can get the following results. Principal factor 1 is I-1 and refers to points per game; principal factor 2 is I-11 and refers to fault; principal factor 3 is I-3 and refers to the number of three-point shots; principal factor 4 is I-7 and refers to backboard; and principal factor 5 is I-6 and refers to free throw percentage. Thus, the above five factors can be regarded as the main factors of basketball technical index analysis. The percentages of variance in 2010 and 2006 are represented by [] and 【】respectively. Then points per game factor [47.147%]【30.006%】, fault factor [16.918%]【21.142%】, three-point shot factor [10.632%]【13.694%】, backboard[7.691%]【10.810%】and free throw factor [4.856%]【7.111%】are named respectively.

Seen from the perspective of the percentage of variance, the first principal factor of 2010 has increased more significantly than of 2006 while the proportion of other factors has gradually decreased. If five main factors summarized in 2006 were used to conduct targeted training on the 2010 World Basketball Championship, significant achievements could be achieved.

In the light of the data characteristics of two consecutive world basketball championships, this paper summarizes the category and division based on data as well as the characteristics of main factors. Using the data characteristics of previous basketball game can serve to analyze those of the next basketball match in a more objective way, thereby providing more accurate approaches for targeted training.

CONCLUSION AND DISCUSSION

This paper firstly retrieves data through the official website of World Basketball Championships and extracts 12 technical indexes displayed in each game by 24 participating teams to provide a basis for data characteristic analysis of basketball technical indexes.

Next, principles and analysis steps of main factor analysis and cluster analysis are shown in this paper to lay the theoretical foundation for data characteristic analysis.

Then, statistical software SPSS19.0 is used to analyze the technical statistical data in 2006 and 2010 World Basketball Championships, after which the results of main factor analysis and cluster analysis are obtained. Then this paper employs graphs to present the eigen-value of each factor, square and load and rotating square and load and obtains the classification tree diagram based on distance. The result analysis divides technical statistical indexes into five categories and extracts five main factors.

Ultimately, on the basis of the constitution of five main factors and the variance percentage of the initial eigen-value, the differences of data characteristics between 2006 and 2010 are analyzed. Five main factors extracted from the previous World Basketball Championships are obtained. They can be taken as the factor references for the targeted training of the next World Basketball Championship and have higher reliability.

REFERENCES

[1] Li Huilin. *Journal of Xuchang University*. **2009**.28(2）:73-79.
[2] Lou Bigang. *Comparative Study on Offensive and Defensive Capabilities between Chinese Men's Basketball Team and Internationally Leading Teams – Taking the 16th World Championship and the 30th Olympic Games for Example* [D]. Chongqing: Southwest University. **2013**.29(8）:161-163.
[3] Yang Lei. *Journal of Chifeng University*. **2013**.29(8）:161-163.
[4] Wang Zhiwei, et al. *Journal of Mudanjiang Normal University*. **2012**.81:57-59.