ISSN : 0974 - 7435 Volume 8 Issue 6

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 8(6), 2013 [738-743]

The analysis of factors affecting the development of tennis based on principal component analysis

FengLi[1]*, Xiaoxuan Jia[2], Geng Du[3], AbdolGhaffar Ebadi[4]

[1]Institute of PhysicalEducation,HuanggangNormalUniversity,Huangzhou 438000 (CHINA)

[1]E-mail: tiyuxi@qq.com

[2]Physical Education Department,Agricultural University of Hebei, Baoding 071001 (China)

[2]Email: hebeinongyedaxuewy@163.com

[3]Department of Sports Training, Wuhan Institute of Physical Education, Wuhan 430079 (China)

[3]Email: Dugeng@wipe.edu.cn

[4]Department of Agriculture, Jouybar branch, Islamic Azad University, Jouybar (IRAN)

ABSTRACT

Tennis has been widely recognized and welcomed by the world public, but the development of tennis is very imbalanced, and to find the most important factors affecting the development of tennis has become a research hotspot. This paper uses the questionnaire method, principal component analysis and other statistical analysis to analyze the obtained data using Excel and SPSS. The results showed that only put several factors that the general publics think will impact tennis development into statistical software SPSS, conduct frequency analysis and the standard deviation of several observe parameters are around 40, which is not ideal. Principal component analysis method well solves this problem, and the study found: F2 (income and consumption attitudes), F4 (professional level of coaches) and F6 (less training institutions) are the main factors limiting the development of tennis in some areas.

KEYWORDS

Principal component analysis;
Tennis promotion;
Influencing factors;
Factor analysis.

INTRODUCTION

Throughout the development of the tennis sport in recent years, China's tennis development trend is imbalanced, which is reflected in the region of the economically developed provinces it is carried out in full swing, while in the economically underdeveloped regions and remote areas, development situation of tennis is relatively slow and lagged. For various reasons, there is a big gap between East and West, the development level of the society has a difference about a decade comparing with the developed areas. So the development of national fitness industry including tennis is also severely constrained.

In this paper, through document literature, survey, mathematical statistics, principal component analysis and other research methods, it uses the obtained data the by questionnaire, through statistical software Excel and SPSS, conducts analysis on the education level of tennis participants and other factors, focuses on analyzing several factors limiting the development of tennis, and hopes to further provide reference and decision basis for further promoting the popularity and increasing of tennis development.

TABLE 1 : Educational level of participants

Primary school	Secondary school	Undergraduate course	Graduate student	Other
15	105	98	50	29

TABLE 2 : Occupational structure of the involved investigating officers

Unit staff	School student	Office leader	teachers	citizens	other
70	100	35	20	34	38

THE DATA ACQUISITION AND CONDUCT PRELIMINARY ANALYSIS

In this paper, through survey and research methods, it conducts a survey of tennis players in the community, schools and government agencies. Then compare them respectively according to educational level, occupation, gender and age structure, and use statistical software EXCEL for analysis.

Tennis is cultural consumption with a high level. From the participants' education degree, this presents the characteristic of a high education degree. Most of them are above university degree, secondary or higher education degree accounts for 75% . These people accept higher levels of education with more open vision, and more prefer to receive some new things.

Analysis from the occupational structure, occupational structure is relatively simple, focused almost entirely on the students in school and unit staff. The main cause is that school students have more time after school, rich hobby after-school and less life pressure. As for the unit staff, their work is relatively easy, so after work they have more ample time for leisure activities.

TABLE 3 : Respondents' gender

Male	Female
197	100

In the survey of athletes and tennis enthusiasts, the proportion of males is significantly higher than that of females. And there are very few female athletes. From the characteristics of tennis, outdoor wind and sun, especially in Gansu such climatic conditions, ultra-intense ultraviolet radiation, are the difficulties that women choose this favorite sport with courage and stick with this sport. This has clearly affected the attitude of women actively participate in tennis. Women's weakness in sports consumption and fitness conscious, difficulties to learn tennis technique as well as requirements on the physical quality all these make women subjected to a lot of adverse effects taking part in tennis. In the survey of athletes and tennis enthusiasts, the proportion of males is significantly higher than that of females. From the characteristics of tennis, outdoor wind and sun makes many girls keep respectfully aloof from the sport.

TABLE 4 : The age structure of the respondents

Teenager	Young	Prime	Aged
54	125	88	30

The age distribution of tennis players is more concentrated, mainly in youth and middle-aged, and the two are essentially fair. Young and middle-aged tennis players are higher than the prime tennis players. Such a survey result has some differences comparing with the more developed cities. As can be seen from the table, the proportion of young and aged participants is rare. Due to various reasons, such as entering a higher school, family environment, it makes the young people's understanding of tennis is not high or have no opportunity to participate in this sport.

THE STATISTICAL ANALYSIS OF FACTOR LIMITING TENNIS DEVELOPMENT

In the second part of this article it has conducted a preliminary analysis of the obtained survey data. The following we will focus on the analysis on limiting factors that impact tennis development. Preliminarily clear

TABLE 5 : The statistical tables of main factors influencing tennis

Influencing factors	Athletes and enthusiasts	Coaches and teachers	Leader of Sports Bureau
F1	103	59	15
F2	105	64	15
F3	98	58	14
F4	101	59	12
F5	89	61	11
F6	91	52	12
F7	87	60	15

F1: Poor weather conditions; F2: Income and consumer attitudes; F3: Inadequate social participation; F4: Professional level of coaches; F5: Insufficient Tennis promotion; F6: Less training institutions; F7: Government support is not enough

up the obtained questionnaire and get TABLE 5.

In TABLE 5, we list several major factors that the public think impact tennis development, such as climatic conditions, income and consumer attitudes, government support capacity, social participation and so on. Put TABLE 5 into the statistical software SPSS conduct frequency analysis and obtain the statistical data in TABLE 6.

TABLE 6 : Statistical data

Statistics		F1	F2	F3	F4	F5	F6	F7
N	Valid	3	3	3	3	3	3	3
	Missing	0	0	0	0	0	0	0
Std. Deviation		44.000	45.059	42.016	44.523	39.514	39.501	36.373
Variance		1936.000	2030.333	1765.333	1982.333	1561.333	1560.333	1323.000
Minimum		15	15	14	12	11	12	15
Maximum		103	105	98	101	89	91	87

From TABLE 6, we can see that: the standard deviation of a few investigated parameters is in 40 or so, which is the square root of the arithmetic average of the deviation square sum between the overall units' standard values and the mean value. The formula is:

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2} \qquad (1)$$

Analyze factors limiting the development of tennis using principal component analysis

In many cases, there is a certain correlation relationship between the various characteristics. when there is a certain correlation relationship between the two characteristics, and it can be interpreted as these two variables reflects the sample information with a certain overlap. Principal component analysis is for all the original suggested characteristics, create new characteristics as little as possible, making these new variables are irrelevant pair-wise. And these new variables maintain the original information in aspect of reflecting the objects' information as far as possible.

Principal component analysis is a statistical analysis method that designates the original more variables as a few comprehensive indicators. From a mathematical point of view, this is a dimension reduction technique.

A study object is often multi-factor complex system. Too many variables will undoubtedly increase the difficulty and complexity of analyzing problems, use the relationship between the original variables, use fewer new variables to replace the original more variable, and make these few variables to retain the reaction information of original more variables as much as possible, so the problem is simplified.

Principle: Assume there are n samples, each sample has p variables, and form a data matrix of order $n \times p$:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ . & . & & . \\ x_{n1} & x_{n2} & \cdots & x_{no} \end{bmatrix} \qquad (2)$$

When p is large, it is difficult to observe problem in p-dimensional space. To overcome this difficulty, we need to reduce the dimension, which uses relatively few comprehensive indexes instead of the original more variable indicators, furthermore these few comprehensive indexes can not only reflect the information that the original more variable indicators can reflect as much as possible, but also they are independent of each other.

Take the original variable indicators as $x_1, x_2, \ldots, x_p$, suppose the composite indicators after the dimension reduction process i.e. the new variables are $z_1, z_2, z_3, \ldots, z_m (m \le p)$, then:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ \cdots\cdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{cases} \qquad (3)$$

Determination principles of coefficient l_{ij}:

1. z_i and z_j $(i \neq j; i, j = 1, 2, \ldots, m)$ are independent of each other;
2. z_1 is the greatest variance of all linear combinations $x_1, x_2, \ldots, x_p$, z_2 is the greatest variance of all linear combinations $x_1, x_2, \ldots, x_p$ not relevant with z_1; z_m is the greatest variance of all linear combinations $x_1, x_2, \ldots, x_p$ not relevant with $z_1, z_2, \ldots, z_{m-1}$.

New variable indicators $z_1, z_2, \ldots, z_m$ respectively are the $1, 2, \ldots, m$ principle factors of the original variable index $x_1, x_2, \ldots, x_p$.

As can be seen from the above analysis, principal component analysis is to determine the loads $x_1, x_2, \ldots, x_p$ of the original variables $x_j (j = 1, 2, \ldots, p)$ on the various principal components $z_i (i = 1, 2, \ldots, m)$.

Can be proved mathematically, they respectively are the corresponding eigenvector to the large eigenvalue of the correlation matrix.

The calculation steps of principal component analysis

(1) Calculating the correlation coefficient matrix:

$$R = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1p} \\ r_{21} & r_{22} & \ldots & r_{2p} \\ \ldots & \ldots & & \ldots \\ r_{p1} & r_{p2} & \ldots & r_{pp} \end{bmatrix} \quad (4)$$

$r_{ij} (i, j = 1, 2, \ldots, p)$ is the correlation coefficient of the original variables x_i and x_j, $r_{ij} = r_{ji}$, the calculation formula is:

$$r_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)^2 \sum_{k=1}^{n} (x_{kj} - \overline{x}_j)^2}} \quad (5)$$

(2) Calculating the eigenvalues $|\lambda I - R| = 0$ and eigenvectors: Solve the characteristic equation, commonly use Jacobi (Jacobi) method to calculate the eigenvalues and rank in the order of size $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$;

Obtain the eigenvector $e_i (i = 1, 2, \cdots, p)$ corresponding to the eigenvalue λ_i, require $\|e_i\| = 1$, i.e. $\sum_{j=1}^{p} e_{ij}^2 = 1$, where e_{ij} represents the j-th component of the vector e_i.

(3) Calculate the principal component contribution rate and accumulative contribution rate

Contribution rate:

$$\frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p) \quad (6)$$

Cumulative contribution rate:

$$\frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p) \quad (7)$$

Generally we take the eigenvalue when the cumulative contribution rate reaches 85% -95%, the 1, 2…… $m (m \leq p)$ principle factors corresponding to $m (m \leq p)$.

(4) Calculate the principal component load:

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1, 2, \cdots, p) \quad (8)$$

(5) The various principal component scores:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} \quad (9)$$

Import TABLE 1 into SPSS, and conduct factor analysis. The basic purpose of factor analysis is to use a few factors to describe the link between many indicators or factors, namely group several more closely related variables in the same class; each class variable becomes a factor. It is called a factor, because it is unobservable, that is, not specific variables. Use several less factors to reflect most information of the original

FULL PAPER

data and obtain the correlation matrix as in TABLE 7:

From TABLE 7 we can see that, the less correlation coefficients of climatic conditions, income and consumer attitudes, professional level of coaches and less training institutions are all above 0.9, indicating that the four are closely linked and have a very significant relationship. The direct correlation between various variables is relatively strong. They have overlap in information.

TABLE 7 : Correlation Matrix

Correlation Matrix							
	F1	F2	F3	F4	F5	F6	F7
F1	1.000	.999	1.000	.999	.987	1.000	.990
F2	.999	1.000	1.000	1.000	.994	.999	.996
F3	1.000	1.000	1.000	1.000	.991	1.000	.993
F4	.999	1.000	1.000	1.000	.992	1.000	.994
F5	.987	.994	.991	.992	1.000	.988	1.000
F6	1.000	.999	1.000	1.000	.988	1.000	.991
F7	.990	.996	.993	.994	1.000	.991	1.000

From TABLE 8, we can see that: calculate the eigenvalues, the contribution rate and the cumulative contribution rate of each principal component by the correlation coefficient matrix. The cumulative contribution rate of first principal component reaches up to 99.637% (greater than 85%); consequently it simply requires solving the first and second principal component.

PCA extraction principle is the first x principal components that the corresponding eigenvalues are greater than 1. Eigenvalues to some extent can be seen as a index to represent the size of component influencing degree. If the eigenvalue is less than 1, it indicates that the explanatory power of the principal component is not greater than the mean explanatory power of introducing an original variable. Thus we often take the eigenvalues greater than 1 as the inclusion criteria. TABLE 9 shows the two extracted principal components F2 and F4, namely $x = 2$.

TABLE 8 : Extraction analysis tables of principal component by variance decomposition

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative %
1	6.975	99.637	99.637	6.975	99.637	99.637
2	.025	.363	100.000			
3	2.497E-16	3.568E-15	100.000			
4	1.042E-16	1.488E-15	100.000			
5	8.340E-18	1.191E-16	100.000			
6	-3.883E-17	-5.548E-16	100.000			
7	-4.275E-16	-6.108E-15	100.000			

Extraction Method: Principal Component Analysis.

Using principal component analysis we obtain two factors, the bigger the correlation coefficient absolute value of the variables and certain factor is, the closer the relationship between the factor and the variables is. Factor matrix can also be used as a measure of factor contribution, and the bigger its absolute value is, the greater the contribution is. Can be seen from TABLE 9, F2 (income and consumption attitudes), F4 (professional level of coaches), and F6 (less training institutions) have a higher load in the first principal component.

TABLE 9 : Component matrix

Component Matrixa	
F1	.998
F2	1.000
F3	.999
F4	1.000
F5	.995
F6	.999
F7	.997

components extracted

CONCLUSIONS

By the preliminary statistics analysis of EXCEL, the majority of the tennis population is concentrated in adults and youth groups. People of 30-50 years old mostly have stable income, life style and rhythm is also relatively stable, and can arrange a more stable time to participate in tennis movement. In particular the proportion of the elderly population is small. From the age characteristics of those participants, the development of young people participating in our province tennis is relatively slow. How to increase the popularity of carrying out tennis in adolescents in the future should be a key emphasis in work.

However, only put several factors that the general public think will impact tennis development into statistical software SPSS, conduct frequency analysis and the standard deviation of several observe parameters are around 40, which is difficult to get the desired conclusion. Principal component analysis method well solves this problem, and the study found: F2 (income and consumption attitudes), F4 (professional level of coaches) and F6 (less training institutions) are the main factors limiting the development of tennis in some areas.

REFERENCES

[1] Bing Zhang,Hui Yue; Bio-mechanical Mathematical Model Analysis for Race Walking Technique. International Journal of Applied Mathematics and Statistics, **40(14)**, 469-476 **(2013)**.

[2] Cai Cui; Application of Mathematical Model for Simulation of 100-Meter Race. International Journal of Applied Mathematics and Statistics, **42(12)**, 309-316 **(2013)**.

[3] Haibin Wang, Shuye Yang; An Analysis of Hurdle Performance Prediction Based On Mechanical Analysis and Gray Prediction Model. International Journal of Applied Mathematics and Statistics, **39(9)**, 243-250 **(2013)**.

[4] Hongwei Yang; Evaluation Model of Physical Fitness of Young Tennis Athletes Based On AHP-TOPSIS Comprehensive Evaluation. International Journal of Applied Mathematics and Statistics, **39(9)**, 188-195 **(2013)**.

[5] Wang Nan; Research on the influence factors of goal involvement and training methods of teenager's tennis players. Journal of Harbin Institute of Physical Education, **25(5)**, 124-126 **(2007)**.

[6] Wang Ya-bing, Shao Li-yan; Investigation and countermeasures on the current condition of junior amateur tennis training in Liaoning province. Journal of Shenyang Sport University, **25(03)**, 110-112 **(2006)**.

[7] Yi Liu; The Establishment of Hierarchical Model for Basketball Defensive Quality. International Journal of Applied Mathematics and Statistics, **44(14)**, 245-252 **(2013)**.

[8] Yong Fan; Statistical Analysis Based On Gray System Theory Basketball Team Scores Its Technical Indicators Associated. International Journal of Applied Mathematics and Statistics, **44(14)**, 185-192 **(2013)**.