



Research Article

ISSN : 0975-7384  
CODEN(USA) : JCPRC5

## Chinese sports development status research based on SPSS principal component analysis and GM model

Xiaona Zhu and Qicheng Wei

*Institute of Physical Education, Zhoukou Normal University, Zhoukou, Henan, China*

### ABSTRACT

*Sports development is one kind of reflection about one country's social harmony and overall strength. In order to easier government plan future sports development, from sports development multiple influence factors, adopt principal component analysis to get main factors that are respectively sports funding, sports structure staff arrangement as well as champions quantity, make prediction of future years' sports development status with curve fitting and grey model; Meanwhile adopt time series method for comparing, utilize grey prediction the results would be more precise. That is government sports funding in 2013, 2014 are respectively 484.21 billion Yuan, 542.96 billion Yuan, athletes that achieve world champion numbers and items maintain 180 people, 24 champion items.*

**Key words:** Sports development, Principal Component Analysis (PCA), Residual analysis, Grey Model (GM)

### INTRODUCTION

With national economic development, mass sports fitness requests are growing. To speed up sports development, give sports function into play, government should make better plans for sports future development.

Grey model is one indefinite factor method prediction system, all random process, as a certain range time relative grey process change, and in model fewer data is required, less calculation gets more widely application [1]. Principal component analysis was introduced on non-random variable by Pearson in 1901. The main purpose of principal component analysis is hope that can use fewer variables to explain original information most part deformation, convert our higher correlation variables into independent or uncorrelated variables from each other [2-4]. Normally, it selects some new variables that numbers are less than original variables and can explain most part of information deformation as so-called principal components, and used them to explain information comprehensive indicators. From this, it is clear that principal component analysis is a kind of dimensional reduction [5, 6]. Advantages of principal component analysis are that it can eliminate evaluation indicators correlation influence, reduce indicator selection work, when there are more evaluation indicators, it can also use fewer comprehensive indicator to substitute original indicators to carry out analysis on the condition that most part information kept [7]. In comprehensive evaluation functions, each principal component weight is its contribution rate; it reflects the principal component included original data information weight accounts for whole information, in this way the defined weight is objective and reasonable, which overcomes some evaluation methods thought defined weight shortcomings. Practical application of principal component analysis is universal, such as census, quantitative geography, and molecular dynamics simulation as well as other disciplines all have such applications, is a kind of common multivariate analysis method [3-5].

Therefore this paper applies principal component analysis to carry out dimensional reduction on it, finds out its main influence factors. Then make prediction on future years' sports development status with its main factors through curve fitting and grey prediction model.

**CHINESE SPORTS DEVELOPMENT STATUS PRINCIPAL COMPONENT ANALYSIS**

Table 1 from China Statistical Yearbook (<http://www.stats.gov.cn/tjsj/ndsj/>) gets Chinese sports development status main indicators statistics data (2008-2012). Through SPSS software, make principal component analysis of it; make use of each variable correlation to get principal component factors.

**Table 1: Chinese sports development status main indicators (2008-2012)**

| Years | Sports funding (a hundred million Yuan) | Sports system structure staff(person) | Numbers of athletes get world champion | Numbers of Grade coaches | Numbers of athletes creating world record items |
|-------|---|---------------------------------------|--|--------------------------|---|
| 2008  | 1589.9                                  | 150575                                | 151                                    | 669                      | 14  |
| 2009  | 2383.4                                  | 153398                                | 223                                    | 781                      | 8   |
| 2010  | 2959.4                                  | 155527                                | 180                                    | 1451                     | 15  |
| 2011  | 3162                                    | 157333                                | 198                                    | 1054                     | 22  |
| 2012  | 4271.3                                  | 159762                                | 140                                    | 767                      | 16  |

Through SPSS analysis, result as following Table 2:

**Table 2: Correlation matrix**

| Indicator   |   | Sports funding( a hundred million Yuan) | Sports system structure staff(person) | Numbers of athletes get world champion | Numbers of Grade coaches | Numbers of athletes creating world record items |
|-------------|---|---|---------------------------------------|--|--------------------------|---|
| Correlation | Sports funding( a hundred million Yuan)         | 1.000                                   | .987                                  | -.257                                  | .207                     | .410  |
|             | Sports system structure staff(person)           | .987                                    | 1.000                                 | -.180                                  | .258                     | .519  |
|             | Numbers of athletes get world champion          | -.257                                   | -.180                                 | 1.000                                  | .233                     | -.274   |
|             | Numbers of Grade coaches                        | .207                                    | .258                                  | .233                                   | 1.000                    | .317  |
|             | Numbers of athletes creating world record items | .410                                    | .519                                  | -.274                                  | .317                     | 1.000   |

From Table 2 correlation matrix, it is clear that each variable has a certain correlation, from which some variables has larger correlations that close to 1, if sports funding and sports system staff correlation arrive at 0.987, therefore it is proper to make principal component analysis.

**Table 3: Variance accumulation contribution ratio**

| Explanatory total variance |                       |            |                |                              |            |                |
|----------------------------|-----------------------|------------|----------------|------------------------------|------------|----------------|
| Element                    | Initial feature value |            |                | Extract squares sum and load |            |                |
|                            | Total                 | Variance % | Accumulation % | Total                        | Variance % | Accumulation % |
| 1                          | 2.517                 | 50.342     | 50.342         | 2.517                        | 50.342     | 50.342         |
| 2                          | 1.237                 | 24.735     | 75.077         | 1.237                        | 24.735     | 75.077         |
| 3                          | .791                  | 15.824     | 90.901         |                              |            |                |
| 4                          | .455                  | 9.099      | 100.000        |                              |            |                |
| 5                          | -7.831E-17            | -1.566E-15 | 100.000        |                              |            |                |

Extract method: Principal component analysis.

From Table 3 variance accumulation contribution ratio, it is known that only previous two feature values beyond 1, therefore SPSS only select previous two principal components. The previous three principal components accumulation contribution ratio arrives at 90.9%, from which the previous two principal components variances account for 75% of total variance, so select previous two principal components to substitute original variables.

Table 4: Component matrix

| Component matrixa                               |           |       |
|---|-----------|-------|
| \   | Component |       |
|   | 1         | 2     |
| Sports funding( a hundred million Yuan)         | .920      | -.070 |
| Sports system structure staff(person)           | .948      | .015  |
| Numbers of athletes get world champion          | -.339     | .793  |
| Numbers of Grade coaches                        | .384      | .775  |
| Numbers of athletes creating world record items | .714      | .030  |
| Extract method: Principal component analysis.   |           |       |
| a. Already extracted 2 components.              |           |       |

From Table 4, it outputs principal component coefficient matrix, indicates each principal component load in each variable, it can get principal component expression:

$F1 = 0.920 * \text{sports funding} + 0.948 * \text{sports system structure staff} - 0.339 * \text{numbers of athletes get world champion} + 0.384 * \text{numbers of Grade coaches} + 0.714 * \text{numbers of athletes creating world record items}$

$F2 = -0.070 * \text{sports funding} + 0.015 * \text{sports system structure staff} + 0.793 * \text{numbers of athletes get world champion} + 0.775 * \text{numbers of Grade coaches} + 0.030 * \text{numbers of athletes creating world record items}$

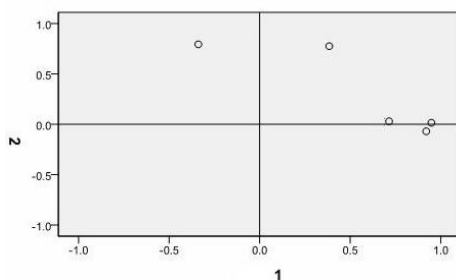


Figure 1: Component figure

Combine Table 4 with Figure 1, it is clear that in first principal component, sports funding and sports system structure staff account for large proportion of component coefficients, which can be regarded as comprehensive indicators reflect such variables; in the second principal component, numbers of athletes get world champion and numbers of Grade coaches account for large proportion of component coefficients.

## SPORTS DEVELOPMENT PREDICTION MODEL

### Carry out grey model analysis of sports

To easier government make better planning on future sports, select principal component analysis results' larger proportion principal component coefficients to carry out prediction of future two years' sports funding and numbers of athletes get world champion respectively by curve fitting and grey model. To make prediction correctly, following Table 5 is consulted statistics data (2013-2012) from China Statistical Yearbook.

Table 5: Sports funding and Athletes' champions' status

| Year | Sports funding( a hundred million Yuan) | Numbers of athletes get world champion | Numbers of athletes creating world record items |
|------|---|--|---|
| 2003 | 531.5                                   | 94                                     | 17  |
| 2004 | 773.4                                   | 175                                    | 27  |
| 2005 | 857                                     | 159                                    | 22  |
| 2006 | 955.4                                   | 169                                    | 24  |
| 2007 | 1243.4                                  | 217                                    | 22  |
| 2008 | 1589.9                                  | 151                                    | 24  |
| 2009 | 2383.4                                  | 223                                    | 30  |
| 2010 | 2959.4                                  | 180                                    | 22  |
| 2011 | 3162                                    | 198                                    | 24  |
| 2012 | 4271.3                                  | 140                                    | 24  |

Utilize MATLAB making fitting figure and residual analysis figure for sports funding as following Figure 2 shows:

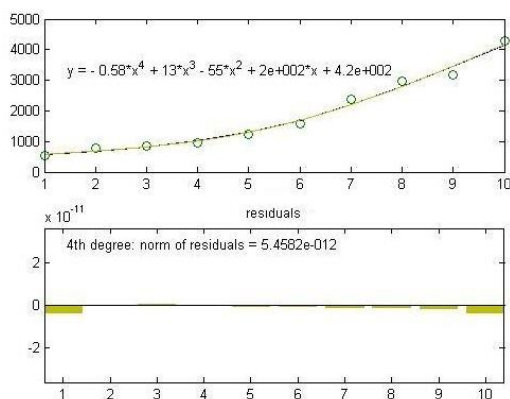


Figure 2: Fitting figure and residual analysis figure

From Figure 2 residual analysis figure, it is clear that residual is very small, fitting precise is higher. Received fitting curve as formula (1):

$$y = -0.58x^4 + 13x^3 - 55x^2 + 200x + 420 \quad (1)$$

Input  $x=11$ ,  $x=12$  into formula (1), solve:

$$x=11, z = 4842.1$$

$$x=12, z = 5429.6$$

That is as following Table 6 shows:

Table 6: National sports funding predicted value

| Year  | 2013   | 2014   |
|---|--------|--------|
| Funding predicted value(a hundred million Yuan) | 4842.1 | 5429.6 |

From above Table 6, it is clear that recently two years government sports funding is gradually increasing. Adopt grey model to predict numbers of athletes get world champion:

Step 1: Extract data from Table 5, form into original data column formula (2):

$$X^0 = [94, 175, 159, 169, 217, 151, 223, 180, 198, 140] \quad (2)$$

Step 2: Make first order accumulation of data  $X^0$  generation, it gets formula (3):

$$X^1 = [94, 269, 428, 597, 814, 965, 1188, 1368, 1566, 1706] \quad (3)$$

Step 3: Construct matrix  $B$ ,  $Y_n$  as formula (4) (5):

$$B = \begin{bmatrix} -\frac{1}{2} * [Y^1_{(1)} + Y^1_{(2)}] & 1 \\ -\frac{1}{2} * [Y^1_{(2)} + Y^1_{(3)}] & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ -\frac{1}{2} * [Y^1_{(9)} + Y^1_{(10)}] & 1 \end{bmatrix} = \begin{bmatrix} -181.5 & 1 \\ -348.5 & 1 \\ -512.5 & 1 \\ -705.5 & 1 \\ -889.5 & 1 \\ -1076.5 & 1 \\ -1278.5 & 1 \\ -1467 & 1 \\ -1636 & 1 \end{bmatrix} \quad (4)$$

$$Y_{10} = [X^0_{(2)}, \dots, X^0_{(n)}]^T = [175, 159, 169, 217, 151, 223, 180, 198, 140]^T \quad (5)$$

Step 4: Solve constant coefficients  $a$ 、 $b$  formula (6):

$$\begin{pmatrix} a \\ b \end{pmatrix} = (B^T * B)^{-1} * B^T * Y_n = \begin{pmatrix} -0.0004 \\ 178.7174 \end{pmatrix} \quad (6)$$

Step 5: Get Table expression formula (7):

$$Y^1_{(k+1)} = \left( X^1_{(k)} - \frac{a}{b} \right) * e^{-a*k} + \frac{a}{b} = 446889.5e^{-0.0452k} + 446795.5 \quad (7)$$

Let:

$$Y^0_{(k)} = Y^1_{(k+1)} - Y^1_{(k)} \quad (8)$$

$Y^0_{(k)}$  represents values after restoration.

From formula (1)-(8), it can get following result Table 7.

**Table 7: Prediction on numbers of athletes' world champions**

| Year            | 2003 | 2004  | 2005  | 2006  | 2007 | 2008  | 2009   | 2010   | 2011  | 2012  |
|-----------------|------|-------|-------|-------|------|-------|--------|--------|-------|-------|
| Reduction value | 94   | 178.7 | 178.8 | 178.9 | 179  | 179.1 | 179.19 | 179.27 | 179.3 | 179.4 |
| Actual value    | 94   | 175   | 159   | 169   | 217  | 151   | 223    | 180    | 198   | 140   |
| Error           | 0    | 3.7   | 19.8  | 9.9   | -38  | 28.1  | -43.81 | -0.73  | -18.7 | 39.4  |
| Relative error% | 0    | 2.11  | 12.4  | 5.8   | 17.5 | 18.6  | 19.6   | -0.4   | -9.4  | 28.1  |

Utilize formula (7) (8), prediction results for 2013-2017 numbers of champions can refer to Table 8.

**Table 8: Numbers of champions predicted value**

| Year            | 2013  | 2014   | 2015   | 2016   | 2017   |
|-----------------|-------|--------|--------|--------|--------|
| Predicted value | 179.5 | 179.58 | 179.66 | 179.74 | 179.82 |

Similarly, it can get athletes champion items predicted values as Table 9.

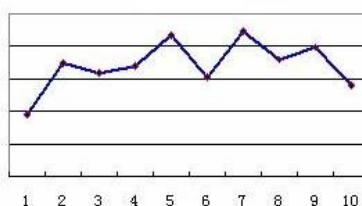
**Table 9: Champion item predicted value**

| Year            | 2013  | 2014  | 2015 | 2016  | 2017 |
|-----------------|-------|-------|------|-------|------|
| Predicted value | 24.16 | 24.13 | 24.1 | 24.03 | 24   |

From Table 7, it is clear, the maximum relative error is 28.1%, so grey prediction method is not quite correct, in order to get relative correct results, adopt time series to make comparison.

### Time series prediction model analysis of sports

From Table 5 athletes win world champion numbers data as Figure 3 shows, though random fluctuation exists, time series still shows stable with time development. Select exponential smoothing to predict world champion numbers.



**Figure 3: Numbers of athletes winning world champion**

Exponential smoothing using past time series value weighted average as next term prediction, mathematical expression as (9).

$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t \quad (9)$$

From which,  $F_{t+1}$  = t+1 term time series predicted value,  $Y_t$  = t term time series actual value,  $F_t$  = t term time series prediction,  $\alpha$  is smoothing constant. Here takes  $\alpha = 0.3$ , use the model getting Table 10.

Table 10: ( $\alpha = 0.3$ ) numbers of champion predicted value and error

| Year | Numbers of champion actual value | Predicted value | Predicted error | Relative error |
|------|----------------------------------|-----------------|-----------------|----------------|
| 2003 | \                                | \               | \               | \              |
| 2004 | \                                | \               | \               | \              |
| 2005 | 159                              | 175             | -16             | -0.1006        |
| 2006 | 169                              | 170.2           | -1.2            | -0.0071        |
| 2007 | 217                              | 169.8           | 47.2            | 0.21751        |
| 2008 | 151                              | 183             | -32             | -0.2119        |
| 2009 | 223                              | 173.4           | 49.6            | 0.22242        |
| 2010 | 180                              | 188.3           | -8.3            | -0.0461        |
| 2011 | 198                              | 186             | 12              | 0.06061        |
| 2012 | 140                              | 190             | -50             | -0.3571        |

Error that received is larger than grey model by comparison, so adopt grey model prediction result. That is post five years' numbers of athletes win champion are fixed on 180 people, items of champion are fixed on 24 items.

## CONCLUSION

Based on above quantitative analysis results, it is clear that government improving sports development can start with following main aspects strengthen sports funding, increase sports system structure staff. Of course, develop sports industry should mainly rely on social each aspect strength, it cannot more rely on national funding, it should insist on principle "who invests, who possesses, who benefits", break through regions, departments and ownership constraints, encourage all sectors of society invest and hold sports industry, fully provide positive favorable policies. Adhere to popularization and improvement combination; insist mass sport and competitive sport coordinate development. According to development law of mass sport and competitive sport, comprehensive improve our country sports overall levels. Numbers of athletes win world championship can reflect one country sports level, it can add some body building equipments so that improve people physical quality, enlarge athletes' physical training, select excellent coaches, so that increase national sports levels, realize great China dream.

## Acknowledgment

Philosophy and social science papers get Henan planning project support, project number: 2011FTY011. Project name: Concept and athletic ability, approach - avoidance goal orientation related to cognitive evaluation model.

## REFERENCES

- [1] Ding Lin-mei. *Journal of Shenyang Sport University*, **2004**, 23(6), 753-755.
- [2] Liu Ya-yun; LI Yan-ling; Zhong Li-ping; Fan Cheng-wen. *Journal of Beijing Sport University*, **2012**, (2), 120-124, 129.
- [3] Xu Mao-wei, Guan Wen-tao. *Journal of Wuhan Institute of Physical Education*, **2011**, 45(12):14-18.
- [4] Chen Po; Qin Zhong-Mei; Yin Ying; Xia Chong-De. *Journal of Beijing Sport University*, **2007**, 30(12), 1610-1613.
- [5] He Qiang. *Journal of Wuhan Institute of Physical Education*, **2011**, 45(7), 16-20.
- [6] Wang Luyao, Yan Ping. *Journal of Wuhan Institute of Physical Education*, **2012**, 46(7), 26-30.
- [7] Zhang B.; Zhang S.; Lu G. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(9), 256-262.