

2014

# BioTechnology

*An Indian Journal*

FULL PAPER

BTAIJ, 10(9), 2014 [3017-3025]

## Grey correlation degree-based CBA basketball game techniques influence factors study

Chunlei Zhang

Physical Education Department, Changchun University of Science and Technology,  
Changchun, 130000, JiLin, (CHINA)

### ABSTRACT

In recent years, basketball games are constantly popular; various people participate in basketball games, so the paper goes deeper research in basketball games. Based on documents literature, regression analysis, grey correlation analysis, the paper makes quantitative analysis of recent five sessions' CBA league matches champions teams, on this basis, it constructs indicators grey correlation sequence and regression model. By researching, it analyzes champions' teams winning influence factors that include field-goal percentage, free throw percentage; three-point shot percentage, assist, defense and others nine kinds of factors. And then by analyzing the nine kinds of influence factors, it gets games scores influential relative significant indicators, finally it makes summary and provides suggestions for teams training.

### KEYWORDS

CBA; Technical indicator; Grey correlation; Regression analysis; Champion team.



## INTRODUCTION

Basketball was originated in America in 1891, it was invented by America physical education teacher Dr. James Naismith. In the beginning, Naismith respectively installed two baskets that used to put pears in guard bar of both sides in corridor of second floor in school gym, basketball upper side was 3.05m higher from ground, used football as games tool, played games in the way of throwing into basket. Now basketball event is hot, National Basketball Association (NBA) is of great concerns of every basketball fan. Chinese basketball games are also unprecedented hot, CBA games become indispensable talk when every youth that is fond of basketball at leisure. Basketball is a kind of technological event, basketball techniques are essential conditions to decide games performances, the paper carries on research and analysis of basketball games each indicator technique by grey correlation method; by stepwise regression model, it solves scores correlated significant indicators and verifies.

## RESEARCH OBJECTS AND METHODS

### Research objects

CBA league matches' recent five years' champions teams that season 2009-2010, season 2010-2011, season 2011-2012, season 2012-2013 and season 2013-2014 champions teams. Utilize Chinese basketball association official website obtained CBA league matches' recent five years' champions teams all games data as research objects, and meanwhile carry on quantitative analysis<sup>[1]</sup>.

### Research methods

Documents literature: On the basis of consulting numerous relative basketball sports techniques, basketball tactics, basketball defense techniques' time factors, correlation test and other aspects literature information, the paper focuses on researching relative information of them; by detailed searching and consulting teams information, it gets acknowledge CBA league matches' champion team.

Video observational method: Watch games videos by China internet TV station, and observe applied sports techniques, tactics statistical method by video, make statistics of champions' teams' field goal percentage, rebound, fault and other technical indicators.

Data statistics method:

(1) By handling with CBA league matches recent five seasons all teams data, make statistics of every season champion team scores status, count each item technical indicator and game scores correlations, use EXCEL software to make statistics of basketball technical indicators, and make classification of indicators and draw<sup>[2]</sup>.

(2) The core of grey correlation analysis is utilizing sequence curve geometric shape approximate degree to judge whether its connections are compact or not, curve geometric shape gets more similar, then corresponding sequences correlation degree will be bigger, on the contrary, it gets smaller. It is well known that all time sequence data, horizontal sequence data and indicator sequence data can be used to make correlation analysis. Due to CBA champion team attack and defense data and winning rate are one kind of horizontal sequence data, therefore, the paper using correlation analysis is reasonable and feasible.

(3) Grey correlation analysis application range<sup>[3]</sup>. When solve correlation degree, it uses indicators every point correlation coefficient average value. When correlation coefficient distribution is discrete, then it will produce one inclination that point with big correlation coefficient numeric value decides overall correlation degree. Therefore, in case sub sequence gets closer to main sequence corresponding elements, results obtained by grey correlation degree is obtained by little discrete degree correlation coefficient average value, now average value can reflect correlation coefficient overall commentary, therefore it is reliable.

(4) Calculate basketball game's all technical indicators (that are influence factors) and game result correlation degree, concrete calculation process is 1) firstly define reference sequence and comparison sequence; 2) make original data transformation of sequence, and then solve absolute difference sequence;

3)calculate correlation coefficient, correlation degree; 4) rank correlation sequenced, get correlation matrix and make advantage analysis.

(5)Utilize regression analysis method to handle with research objects recent five years' participated all games (total 241 games) data, and then construct regression model test<sup>[4]</sup>.

## RESULT AND ANALYSIS

### Data analysis

Make table and sort out data net's data as TABLE 1 shows, it is clear that champion teams games performance status in the five years show unstable tendency as first stable and then declining, teams performances are gradually declining after 2009 season, which shows teams final scores levels have declination. In recent five seasons period, except for Zhejiang team's every champion team field goal percentage hasn't had fiercely fluctuation that among 50%-52%. It indicates every champion team except for Zhejiang team,its field-goal percentage keeps stable state, fluctuation is no more than two percentage, of course exclude 2013-2014 season. From the table, it is clear that recent two years season's free throw percentages are highest; it shows teams have improved in free throw aspect. Three-point shot percentage also always keeps among 0.32-0.37, it is relative stable but in the declining trend. 2013-2014 seasons Zhejiang team free throw shot percentage is the highest, 2009-2010 and 2010-2011 these two seasons champions are Guangdong team, while Shanxi team is the champion in 2011-2012and 2012-2013 the two seasons. And 2013-2014 season champions are Zhejiang team, which indicates in 2013 season, both Guangdong team and Shanxi team performance decline. But Guangdong team rebound number keeps unchanged, it shows Guangdong delegation is relative skillful and has stable basis in the technical indicator.

By TABLE 1 data, it is clear that amount of steals are relative stable in the five seasons period, no big changes, from which steal amount is largest in 2009-2010 season; fault amount changes are larger,which represents first reduce and then increase, and2009-2010 season fault amounts are still largest, therefore the paper studied teams focus on fault reduction, give their normal performance into play; free throw percentage and scores have larger correlations, so research objects should more strict with free throw training, improve their hit rates; amount of block shot changes are larger in the five seasons, it presents unstable tendency, and achieve highest records in 2012-2013 season ;amount of foul changes get closer to increasing function, during the five seasons, it shows increasing by season, which indicates research objects should put emphasis on competition rules, and understand competition rules in their hearts, they should take more cultivation and training in key techniques at ordinary times, so that can get good results in competitions;Amount of assists are relative stable, in these five seasons, no excessive increases or decreases appear, but assists are least in 2011-2012 season.

**TABLE 1 : Recent five seasons' champions' teams influence factors and competition scores status**

Season	Champion team	Score	Field goal percentage	Free throw percentage	Three-point shot percentage	Block shot	Total rebounds	Assist	Foul	Steal	Fault
2009-2010	Guangdong	111.30	0.52	0.68	0.37	3.50	45.80	17.30	25.10	11.10	16.70
2010-2011	Guangdong	109.80	0.50	0.72	0.37	4.50	45.80	15.90	26.40	11.00	16.00
2011-2012	Shanxi	110.40	0.52	0.74	0.37	3.50	43.30	14.50	26.30	10.00	14.30
2012-2013	Shanxi	110.40	0.51	0.71	0.33	4.90	44.80	16.90	25.30	9.60	14.90
2013-2014	Zhejiang	101.42	0.46	0.78	0.32	3.40	45.30	16.20	27.10	9.60	15.40

### Correlation analysis

By correl function,make correlation coefficient calculation of TABLE 1 every influence factor and competition score, it gets TABLE 2.

TABLE 2 : Correlation coefficient

Field goal percentage	Three-point shot percentage	Free throw percentage	Total rebounds	Assist	Steal	Block shot	Fault	Foul
0.97	0.80	-0.86	-0.14	0.05	0.52	0.38	0.09	-0.78

From TABLE 2, it is clear that competition score maximum correlation factor is field goal percentage. Among them, free throw percentage, three-point shot percentage, foul and steal have significant correlations by comparing to other technical indicators, and have bigger differences with other technical indicators correlation coefficients. It shows that if one team has obvious advantages in above aspects, then team scores in competitions are higher, so finally probability that wins the competitions similarly will be bigger. Therefore, CBA league matches' all coaches, athletes deserve to focus on these technical indicators aspects training, cultivation and improvements, so that possess initiative in competition, and have best chance of winning. But it doesn't mean the rest technical indicators are not important, every influence factors in basketball field cannot be ignored. It always occurs that strong team is defeated by weak team. In addition, coaches, athletes should also strengthen psychological qualities training and cultivation, which is beneficial to give their best performance into play.

### Grey correlation degree analysis

As a development and changing system, correlation analysis actually is quantization comparative analysis of dynamic process development trend. So-called development trend comparison, is also system each period relative statistical data's geometric relationship comparison. By grey system theory, it is clear that in fact every champion team game score is a grey system project. According to every champion team influence factors and games scores grey correlation degree analysis, reflect every score influential technical indicators and games scores influence and effects, and learn their correlations.

To guarantee established model accuracy and scientificity, the paper makes transformation and handling with original data, let it to eliminate dimension and has comparability.

Data transformation technique:

Definition 1: Set it has sequences:

$$x = (x(1), x(2), \dots, x(n)),$$

Then map:

$$f : x \rightarrow y$$

$$f(x(k)) = y(k), \quad k = 1, 2, \dots, n$$

It is series  $x$  to series  $y$  data transformation.

1)When:

$$f(x(k)) = \frac{x(k)}{x(1)} = y(k), \quad x(1) \neq 0$$

It is called  $f$  is initial value transformation.

2)When

$$f(x(k)) = \frac{x(k)}{\bar{x}} = y(k), \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x(k)$$

It is called  $f$  is average value transformation.

3)When:

$$f(x(k)) = \frac{x(k)}{x_0} = y(k)$$

From which  $x_0$  is some value that above 0, it is called  $f$  is normalization transformation.

4)When:

$$f(x(k)) = \frac{x(k) - \min_k x(k)}{\max_k x(k)} = y(k)$$

It is called  $f$  is range maximum value transformation.

5)When:

$$f(x(k)) = \frac{x(k) - \min_k x(k)}{\max_k x(k) - \min_k x(k)} = y(k)$$

It is called  $f$  is interval value transformation.

In the paper basketball competition technical indicators analysis, it makes following transformation:

$$\bar{x} = (1, \frac{x(2)}{x(1)}, \dots, \frac{x(n)}{x(1)})$$

It is initialization sequence of original sequence  $X$ .

So that we can make initialization processing with TABLE 1 nine sequences except for competition performances. It is worth noting that for previous seven sequences, increasing of numeric values means sports levels advancement, while for the post two sequences, reduction of numeric values means sports levels advancement. Therefore, when do initialization processing with sequence  $x_8$  and  $x_{10}$ , adopt following formula:

$$x_i = (1, \frac{x_i(1)}{x_i(2)}, \frac{x_i(1)}{x_i(3)}, \frac{x_i(1)}{x_i(4)}, \frac{x_i(1)}{x_i(5)}), i = 8,9$$

Correlation analysis:

Definition 2: Select reference sequence:

$$x_0 = \{x_0(k) | k = 1, 2, \dots, n\} = (x_0(1), x_0(2), \dots, x_0(n))$$

Among them, time is represented by  $k$ . Assume that the number of comparison series is  $m$ :

$$x_i = \{x_i(k) | k = 1, 2, \dots, n\} = (x_i(1), x_i(2), \dots, x_i(n)), i = 1, 2, \dots, m$$

Then:

$$\xi_i = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \tag{1}$$

Formula (1) is comparison sequence  $x_i$  to reference sequence  $x_0$  in the  $k$  moment correlation coefficient, from which resolution coefficient is using  $\rho \in [0,1]$  to express, two-level minimum difference and two-level maximum difference respectively use  $\min_s \min_t |x_0(t) - x_s(t)|$ ,  $\rho \max_s \max_t |x_0(t) - x_s(t)|$  to express.

In general, resolution rate and resolution coefficient have same changes.

Due to correlation coefficient is a kind of indicator that describes comparison sequence and reference sequence correlation degree at one moment, but every moment has a corresponding incidence number, information is not so concentrated and is bad for comparison. Therefore, we make following definition.

Definition 3: Call:

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (2)$$

It is sequence  $x_i$  to reference sequence  $x_0$  correlation degree.

Then utilize MATLAB to calculate, it gets correlation degree as TABLE 3.

TABLE 3 : Each indicator and scores correlation degree table

Indicator	Field goal percentage	Free throw percentage	Three-point shot percentage	Total rebounds	Assist	Foul	Steal	Block shot	Fault
Normal grey correlation degree	0.85	0.603	0.604	0.791	0.787	0.80	0.74 2	0.678	0.84
Rank	1	9	8	4	5	3	6	7	2

Handle with team winning influential each kind of technical indicators data, and by grey correlation analysis calculating, it gets influence factors and competition scores correlation degree sizes relations: field goal percentage >fault>foul>total rebounds>assist>steal>block shot>three-point shot percentage>free throw percentage. By the relationship, it gets the first three influence factors obviously have larger correlation degrees with scores, relationship last two items have least correlation degrees with scores, but the two items are also a kind of effective scoring indicator, so it cannot simply ignore them. If two teams strength have no bigger differences, which team is well trained in the two aspects, it can grasp the initiative in competitions, and winning odds will be larger.

### Team scores ability influence (stepwise) linear regression analysis

Stepwise regression is a common method in multiple regressions that is used to select independent variables. The basic thought of the method is: introduce independent variables into equation one by one, condition of introduction is the independent variable partial regression square sum should be the largest one in unselected independent variables (unselected quantity), and it shows significance by F testing. On the other hand, when introduce a new variable every time, it should stepwise carry on F test on previous selected variables in equation (selected quantity), and remove variables that partial regression squares sum is the smallest and have no significance from equation, till independent variables beyond equation cannot be introduces, and independent variables in equation cannot be removed. Another one is "backward method", its basic thought is: firstly establish regression equations that include all independent variables, and then stepwise remove variables, firstly do F (or t) test on every independent variable, remove insignificant variables' independent variable with minimum partial regression squares sum, reconstruct equation. Then, make stepwise F test on independent variable out of equation, introduce variable with maximum partial regression squares sum and has significance into equation. Repeat above process, till equation's all independent variables have significances and

independent variables out of equation have no significance. The method is available in case independent variables not so many, especially insignificant variables are not so many. Compare to general multiple regressions, regression equation that is solved by stepwise regression method has following advantages: it included independent variables numbers are little and easy to apply; its rest standard deviation is also small, equation stability is better: due to every step will be tested; it ensures equation’s all independent variables to have significances.

Let  $Y$  to represent champion team every game score, it is related to variable field goal percentage, free throw percentage, three-point shot percentage, rebound, steal, foul, assist, block shot and fault nine influence factors, from which nine factors respectively use  $X_1, X_2, \dots, X_9$  to express. Assume random factor influence exists, even  $X_1, X_2, \dots, X_9$  are relative fixed, teams rankings are also not all the same. Therefore on the condition affirm that  $Y$  has linear correlations with  $X_1, X_2, \dots, X_9$ , the paper analyzes the linear relationship by two parts. Multiple linear regression model is:

$$\begin{cases} Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_9 X_9 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \tag{3}$$

Among them,  $\varepsilon$  is supposed to a random variable that conforms to normal distribution,  $\beta_0$  is constant, each factor  $Y$  regression coefficient is using  $\beta(1-9)$  to express, if its numeric values get bigger, then it shows impacts are more significant. Utilize independent variables that have significant impacts on  $Y$ , construct regression model. By data, it carries on stepwise regression analysis, gets variables of significant impacts, and according to correlation sizes, so that obtained decisive factors by previous correlation analysis is more convincing and significant. Input data into formula (3), it can be expressed as:

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_n) \end{cases} \tag{4}$$

Among them,  $E_n$  is  $n$  order unit matrix.

Parameters estimation, model’s parameters  $\beta_0, \beta_1, \dots, \beta_m$  still use least square estimation method, that it should select estimation value  $\hat{\beta}_j$ , let when  $\beta_j = \hat{\beta}_j, j = 0, 1, 2, \dots, m$ , error sum of squares:

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_m x_{im}) \tag{5}$$

Arrive at minimum. Thereupon, let:

$$\frac{\partial Q}{\partial \beta_j} = 0, j = 1, 2, \dots, n \tag{6}$$

It gets equation sets, by sorting, it gets:

$$\begin{cases} \beta_0 n + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \dots + \beta_m \sum_{i=1}^n x_{im} = \sum_{i=1}^n y_i \\ \beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \beta_m \sum_{i=1}^n x_{i1} x_{im} = \sum_{i=1}^n x_{i1} y_i \\ \beta_0 \sum_{i=1}^n x_{im} + \beta_1 \sum_{i=1}^n x_{im} x_{i1} + \beta_2 \sum_{i=1}^n x_{im} x_{i2} + \dots + \beta_m \sum_{i=1}^n x_{im}^2 = \sum_{i=1}^n x_{im} y_i \end{cases} \tag{7}$$

Matrix equation set’s matrix form is:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y} \quad (8)$$

When matrix  $X$  column is in full rank,  $X^T X$  is reversible matrix, formula (8) solution is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (9)$$

Substitute  $\hat{\boldsymbol{\beta}}$  back to original model then can get  $Y$  estimation value, as TABLE 4.

**TABLE 4 : Regression model calculation result**

Parameter	Parameter estimation value	Parameter confidence interval
$\beta_0$	-41.5513	[-56.146 -26.958 0]
$\beta_1$	114.711	[101.522 9 127.909 0]
$\beta_2$	77.5698	[67.787 2 87.3424]
$\beta_4$	22.0227	[13.211 9 30.8334]
$\beta_5$	1.3713	[1.737 6 1.605 0]
$\beta_6$	0.6647	[0.4816 0.8478]
$\beta_8$	-0.6947	[-0.9146 -0.4748]
$\beta_9$	0.2714	[0.0699 0.4738]

$$\mathbf{R}^2 = 0.7338 \quad \mathbf{F} = 75.9565 \quad \mathbf{P} = 0$$

By stepwise regression analysis, it gets two indicators are removed from nine influence indicators that are respectively steal and assist. Teams' rankings relative important influence factors are field goal percentage, free throw percentage, rebounds, three-point shot percentage and scores. Obviously each influence factor has very remarkable significance.

Required to test is original hypothesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9$$

If  $H_0$  doesn't deny, then it shows usage of linear regression model is inappropriate.

(1)  $F$  test method

Model obtained statistical quantity  $F = 75.9565$ , by consulting  $F$  distribution numerical table  $F_{0.95}(10,230) = 1.8$ , obviously  $F > 1.8$ , by  $F$  test criterion, it is clear in confidence significance level 0.05, it refuses  $H_0$  that thought  $X_1, X_4, X_8, X_9$  and  $Y$  linear relationship is significant.

(2) Regression model correlation coefficient evaluation

Negative correlation coefficient between  $Y$  and independent variable after elimination is  $R^2 = 0.6671$ , judgment coefficient  $r = 0.8358$ . It shows the paper constructed mathematical model fitness extent belongs to good level, linear correlation is very strong, and meanwhile in confidence significance level 0.05, it refuses  $H_0$ .

(3)  $P$  Value test

Obtained  $P = 0.00$  obviously meets  $P < \alpha = 0.05$ , it indicates regression variable  $X$  and dependent variable  $Y$  linear relationship is significant.

Above three statistical inference method inferred results are consistent, all think after elimination regression independent variable  $X$  and dependent variable  $Y$  linear relationship is significant, it shows regression model can basically reflect  $X$  and  $Y$  relationship. Therefore, coaches can get competition



performance influence factors and sizes according to the model, find out respective advantages and disadvantages, and purposely carry on technique and tactics layout and adjustment.

## CONCLUSION

By statistics and researches on the five seasons data, it finds that champions teams field goal percentage, three-point shot percentage and free throw percentage the three technical indicators performance are very stable since recent five seasons, the fluctuation is not big. By each influence

Factors indicators and scores correlation analysis, it gets that field goal percentage has largest correlation with scores, secondly is free throw percentage, three-point shot percentage, total rebounds, foul and steal. By data processing and researching, it gets free throw percentage and three-point shot percentage have least impacts on competition scores, however in general, the two influence factors indicators are common thought to be most important. It indicates that people should correct their recognition mistakes.

## REFERENCES

- [1] Chen Jian, Yao Song-ping; Application of Connected Analysis to Analysis of Statistical Data of Basketball Technique. *Journal of Shanghai Physical Education Institute*, **33(5)**, (2009).
- [2] Wang Luning et al.; Development of a Computerized System for On-the-spot Technical Statistics in Basketball Matches. *China Sport Science*, **19(3)**, 37-40 (1999).
- [3] Mao Jie, Shan Shuguang; Basketball Technology Statistical Software Development. *Journal of Wuhan Institute of Physical Education*, **46(2)**, 70-73, 87 (2012).
- [4] Zhang Lei; Analysis on Foul in Man's Basketball Final Phase in 10th National Games. *China Sport Science and Technology*, **42(1)**, 50-52 (2006).
- [5] Fu Fan-fei, Yu Zhen-feng, Zhang Quan-ning; Analysis of offensive turnover in period of men's basketball final matches at 10th National Games. *Shandong Sports Science & Technology*, **28(2)**, 24-25 (2006).
- [6] Wang Luning et al.; Development of a Computerized System for On-the-spot Technical Statistics in Basketball Matches. *China Sport Science*, **19(3)**, 37-40 (1999).
- [7] Li Ji-hui; Research on the Current Situation and Developing Strategy of Chinese Track and Field Coaches. *Journal of Shenyang Sport University*, **25(2)**, 63-66 (2006).
- [8] Yang Yue-qing, Rao Han-fu, Len Ji-lan; A Study on the Improvement of Computer Statistic Software for Basketball Games. *Journal of Hubei Sports Science*, **22(2)**, 204-205 (2003).