ISSN : 0974 - 7435 Volume 10 Issue 2

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 10(2), 2014 [095-102]

NBA grand champion in 2013to 2014 seasons prediction regression analysis model research

Jijun Guo
Department of Physical Education, Shenyang Ligong University, Shenyang 110159, Liaoning, (CHINA)

ABSTRACT

In 2013 to 2014, NBA contest warm up again, it attracts whole world numerous basketball fan attentions, many people guess which team will get NBA grand champion. In multiple 30 NBA teams, it selects large possible winning teams each two teams from eastern and western conference as grand champion candidates to carry out technical analysis. At first, to define 9 factors significance on scores, apply multiple linear regressions, analyze scores technical significances with 2-point, 3-point percentage, rebounds and other 9 items, and establish scoring ability model accordingly. Utilize mathematical statistics knowledge comprehensive comparing 2013 to 2014 NBA regular seasons each team technical statistics, each team player, winning rate, age as well as experience four factors. Finally, synthesize above prediction, grand champion in 2013 to 2014 NBA is Heat.

KEYWORDS

NBA grand champion;
Regression analysis;
Statistic;
Technical statistics.

INTRODUCTION

Basketball is a kind of competitive sports events by applying each technique to get scores, since it was invented by America Massachusetts Field YMCA training school teacher Doctor James Naismith in 1892; it quickly became popular throughout the whole world. Thereupon, people also have many researches on it. In 1999, Fei Xue applied IBM formula predicting NBA grand champion in 1999. In 2007, Zong Zhen-Ji, Man Xiao-Xia adopted document literature, video observation, mathematical statistics and other methods carrying out comprehensive evaluation on 2005 to 2006 seasons NBA finals two teams' technique and tactics abilities[1]. In 2012, Jin Huang-Bin, Bai Yin-Long applied stepwise regression analysis and rank correlation analysis researching 2011 to 2012 seasons NBA playoffs each team scores regression analysis.

NBA officials will make detailed statistics on competition each kind of data, from which statistical items are playing time, shooting total times as well as hitting times and hitting rate, 2-point shooting total times as well as hitting times and hitting rate, 3 point shooting total times as well as hitting times and hitting rate, free throw times as well as hitting times and hitting rate, front court rebounds and back court rebounds, assists times, fault times, steal times, block shot times, foul times as well as score totals and other items, based on that officials will make historical data comparative analysis of players and teams, it can get each statistical item historical data maximum value, average value and minimum value, define which competition data that player

FULL PAPER

or team achieves it can win or fail under these conclusive statistics. For NBA game research and grey correlation analysis trial, lots of people have made efforts, formers' research results have detailed summary on NBA competition process techniques and provided more wonderful theoretical platform on the game arresting whole world attentions, and the later researches provide more widely trial range for grey correlation analysis applying. The paper applies regression analysis model analyzes 2013 to 2014 seasons NBA basketball champion team affiliation, in the hope of providing new research method and valuable theoretical guiding on basketball teaching guiding techniques.

REGRESSION ANALYSIS ALGORITHM

Multiple linear regression model[2]

Given dependent variable y and independent variable x_1, x_2, x_3, $\cdots$, x_9 relationships to be:

$$\mathbf{y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_9x_9} \quad \mathbf{(1)}$$

Among them, y is observable random variable, $b_0, b_1, b_2, \cdots, b_{16}$ is unknown parameter, ε is unobservable random error, it meets $E_\varepsilon = 0, D(\varepsilon) = \sigma^2 (\sigma^2 \text{unknown})$, but ε random error is very small, so it can be ignored with regard to actual0

By 16 groups of data, $(y_j, x_{1j}, x_{2j}, x_{3j}, \cdots, x_{9j})(j = 1,2,\cdots,16)$, from which x_{ij} is independent variable x_i the j value, y_j is dependent variable y the j value, input formula(1) and get model data structural formula:

$$\begin{cases} \mathbf{y_1 = b_0 + b_1x_{11} + b_2x_{21} + b_3x_{31} + \cdots + b_9x_{91}} \\ \mathbf{y_2 = b_0 + b_1x_{12} + b_2x_{22} + b_3x_{32} + \cdots + b_9x_{92}} \\ \vdots \\ \mathbf{y_n = b + b_0x_{1n} + b_1x_{2n} + b_3x_{3n} + \cdots + b_9x_{9n}} \end{cases} \quad \mathbf{(2)}$$

Above formula can use matrix to express the formula as :

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{91} \\ 1 & x_{12} & \cdots & x_{92} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{9n} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$\mathbf{Y = XB}$$

Unknown parameter estimation

$b_0 = -54.856$ $b_1 = 210.290$ $b_2 = 50.190$

$b_3 = 29.735$ $b_4 = 1.385$

$b_5 = -0.069$ $b_6 = -1.035$ $b_7 = 1.077$ $b_8 = -0.977$

$b_9 = -0.638$

Therefore, it solves multiple linear regression equation as:

$$\mathbf{y = -54.856 + 210.29x_1 + 50.190x_2 + 29.735x_3 + 1.385x_4 - 0.069x_5 - 1.035x_6 + 1.077x_7 - 0.977x_8 - 0.638x_9} \quad \mathbf{(3)}$$

Regression equation significance test

Multiple linear regression equation F test purpose is to test whether total regression equation is significant or not that is to test all regression coefficient is equal to 0 or not. The concrete steps are as following:

(1) Put forward null hypothesis and alternative hypothesis.

$\mathbf{H_0 : b_{ij} = 0, (i = 1,2,\cdots,9; j = 1,2,\cdots.16)}$

$H_1 : b_{ij}$ Not all of $0, , (i = 1,2,\cdots,9; j = 1,2,\cdots,16)$

(2) According to variance analysis (ANOVA) TABLE 1, it is known F statistical value is 0; corresponding Sig. is F value actual significance probability that is F value actual significance probability that is p value, Sig.=0.025 here. If given α =0.05, obviously p< α, therefore refuses H_0, it is thought regression equation linear relations are significant.

TABLE 1 : Variance analysis table

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	538.788	9	59.865	5.509	.025[a]
	Residual	65.196	6	10.866		
	Total	603.984	15			

a. Predictors: (Constant), fault, 3point, free throw, steal, 2 point, rebound, block shot, assists, foul; b. Dependent Variable: scores

Regression coefficients test

Regression coefficients significance test t test that is

TABLE 2 : Regression coefficients

Coefficients[a]												
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
(Constant)	-54.856	39.783		-1.379	.217	-152.202	42.490					
2 point	210.290	69.511	.933	3.025	.023	40.204	380.377	.741	.777	.406	.189	5.291
3 point	50.190	31.353	.410	1.601	.161	-26.529	126.909	.475	.547	.215	.274	3.653
Free throw	29.735	19.509	.247	1.524	.178	-18.001	77.472	.327	.528	.204	.683	1.464
Rebound	1.385	.555	.718	2.498	.047	.028	2.743	-.065	.714	.335	.218	4.593
1												
Block shot	-.069	1.186	-.017	-.058	.956	-2.970	2.833	.457	-.024	-.008	.215	4.650
Assists	-1.035	.750	-.428	-1.379	.217	-2.870	.801	.404	-.491	-.185	.187	5.356
Steal	1.077	1.722	.211	.625	.555	-3.136	5.290	.458	.247	.084	.158	6.321
Foul	-.977	1.592	-.253	-.613	.562	-4.874	2.920	.111	-.243	-.082	.106	9.472
Fault	-.638	1.672	-.184	-.382	.716	-4.729	3.453	-.451	-.154	-.051	.077	12.956

a. Dependent Variable: scores

to teat independent variable (explanatory variable) x_i effects on dependent variable y is significant or not. Steps are as following:

(1) Put forward null hypothesis and alternative hypothesis.

$H_0 : b_j = 0(j = 1,2,\cdots,16)$

$H_1 : b_j \neq 0(j = 1,2,\cdots,16)$

(2) It needs to carry out significance test on every regression coefficient.

From TABLE 2 nine variables, 2 point x_1,3 point x_2, free throw x_3, rebound x_4, block shot x_5, assists x_6, steal x_7, foul x_8, fault x_9, their Sig are respectively: 0.023, 0.161, 0.178, 0.047, 0.956, 0.217, 0.555, 0.562, 0.716; their corresponding Sig is t value actual significance level that p value, if given $\alpha = 0.05$

$p_1 = 0.023 < \alpha$ therefore refuse H_0, it is thought independent variable"2 point"regression coefficient is significant.

p_2、p_3、p_5、p_6、p_7、p_8、p_9 all are bigger than α, so accept H_0, it is thought" free throw"," block shot"," assists"," steal"," foul"," fault" regression coefficients are not significant.

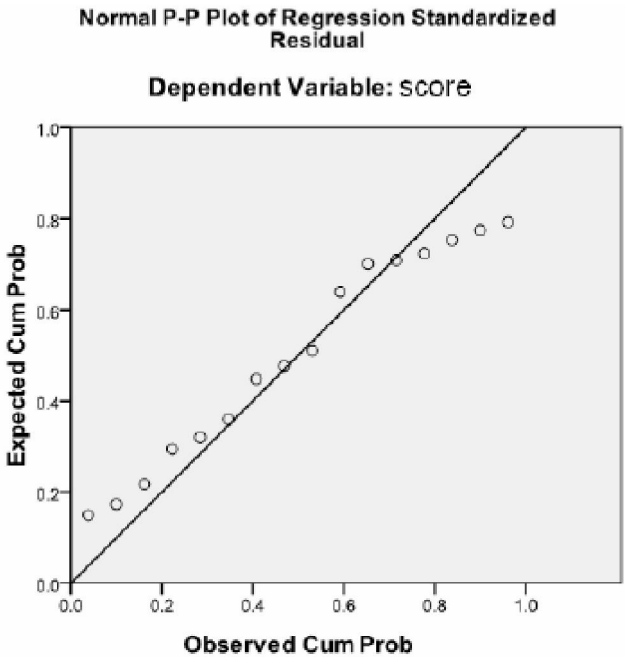

Figure 1 : Scatter

$p_4 = 0.047 > \alpha$ therefore refuse H_0, it is thought that independent variable" rebound" regression coefficient is significant.

FULL PAPER

From above conclusions, it is clear that score y and 2 point x_1, rebound x_4 significance is higher, which have great effects on final scores, other seven items significance is lower, so it effects on total performance is smaller.

Due to most of scatters near to diagonal (Figure 1), it is thought standard residual conforms to normal distribution.

From regression results and analysis results retesting (TABLE 3 and TABLE 4), it gets that score only is significant related to 2 point x_1,rebound x_4, so that consider score and 2 point x_1, rebound x_4 regression linear relationships.

TABLE 3 : Coefficients test results

	Coefficients	Standard error	t Stat	P-value
Intercept	-15.942	28.157	-0.566	0.581
X Variable 1	194.071	42.314	4.586	0.001
X Variable 2	0.561	0.362	1.549	0.145

TABLE 4 : Linear regression model statistical test

R	R square	Adjust R square	Standard error	P value
0.787	0.620	0.561	4.204	0.002

From TABLE 3, it can get regression equation:

$$y = -15.942 + 194.071x_1 + 0.561x_4 \quad (4)$$

Take 16 teams' scores as dependent variable, other technical statistical indicator as 2 point, 3 point shooting percentage, rebound and other 9 items indicators as independent variable, it gets dependent variable y and independent variable x correlation degree. The more correlation coefficient gets closer to 1, it indicates correlation degree is higher, correlation coefficient is 0.787, judgment coefficient is 0.620, by testing, $p < 0.01$, it indicates fitting degree is very good(can refer to TABLE 4).

NBA playoff each team score ability modelling [4]

According to (TABLE 5) score ability regression variance test results indicate p value is 0.002 <0.01, which shows the score ability regression equation is of remarkable significance.

Score ability model test

Take NBA playoffs in 2011 to 2012 16 teams scores and teams rankings as illustration examples (can refer to TABLE 6), it makes rank correlation analysis of it, test results indicate $p < 0.01$, indicates that the score ability model better reflects NBA playoffs in 2011 to 2012 each team score ability that has higher reliability and validity.

TABLE 5 : Variance analysis table

	df	SS	MS	F	Significance F
Regression analysis	2	374.2607557	187.1303779	10.58965952	0.001867109
Residual	13	229.7236193	17.67104764		
Total	15	603.984375			

TABLE 6 : Team combat gains and score ability relationship

	Combat gains	Score ability	Score ability ranking
Spurs	1	100.34	1
Heat	2	96.12	3
Thunder	3	97.89	2
Celtics	4	92.01	6
Grizzlies	5	91.48	8
Lakers	6	90.71	9
Mavericks	7	84.42	15
Clippers	8	93.99	5
Jazz	9	85.03	13
Nuggets	10	94.28	4
Bulls	11	91.84	8
76ers	12	87.09	12
Pacers	13	90.27	10
Hawks	14	85.54	13
Knicks	15	88.25	11
Magic	16	83.24	16

From TABLE 6, it is clear that spurs, heat, thunder the three teams score ability no doubt is the strongest.

MATHEMATICAL STATISTICS

Statistics

Assume that it has a capacity n sample (that is a group of data), record as $x = (x_1, x_2, \cdots, x_n)$; it needs to process it so that can put forward useful information

that used for entity (distribution) parameters estimation and test. Statistics is function by processing and reflects sample quantity feature, it doesn't include any unknown quantity.

(1) It represents location statistics-arithmetic mean value

Arithmetic mean value (is called mean value for short) describing data values average location, record as $\bar{x}$,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (5)$$

(2) It represents variation degree statistics –standard deviation, standard deviation s is defined as:

$$s = \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^{\frac{1}{2}} \quad (6)$$

It is each data and mean value deviation degrees measuring, the deviation can be called as variation.

2013-2014NBA regular seasons each team technical statistics and analysis

From above TABLE 7, it is clear that spurs has advantages in field-goal percentage, 3 point, assists, fault, foul and other technical aspects that attack dominates, while it is dominated in steal, block shot aspects; Heat dominates in field-goal percentage, steal, but is dominated in rebounds, block shot aspects; Thunder dominates in free throw, rebound, block shot, score and other technical aspects, while is dominated in field-goal percentage, 3 point, assists, fault, foul aspects; Pacers lies in middle or lower levels in each technical aspect. In conclusion, it is known that in 2013 to 2014NBA regular seasons spurs and heat have equal performances, thunder is the secondary, pacers is the worst.

2013-2014NBA regular seasons each team players' factors analysis

From above TABLE 8, it is clear that spurs core player dominates in fault, foul aspects; heat core players absolute dominates in field-goal percentage, 3 point, assists, steal aspects, while it is dominated in free throw aspect; Thunder core player dominates in free throw, rebound, assists, block shot and other aspects, while is dominated in fault, foul and other aspects; Pacers almost lies in middle or lower levels in each technical level. In conclusions, it is clear that heat core player court all have best performance, thunder is the secondary, spurs and pacers are worse.

TABLE 7 : 2013-2014NBA regular seasons four teams court each technical indicators[5]

	Spurs	Heat	Thunder	Pacers
Shooting	0.492	0.506	0.466	0.457
3 point	0.399	0.373	0.341	0.363
Free throw	0.774	0.758	0.815	0.791
Front court rebound	9.1	7	11.1	9.7
Back court rebound	33.8	29.6	35.7	34.9
Total rebound	42.9	36.6	46.8	44.7
Assists	25.2	23.4	21.5	20.8
Steal	7.6	9.2	8.1	7.2
Block shot	4.7	4.5	6.2	5.9
Fault	14.6	15	15.8	15.4
Foul	17.5	19.9	22.5	19.9
Score	104.6	104	105.3	98.1

TABLE 8 : 2013-2014NBA regular seasons each team core players' court average technical statistics[5]

	Spurs	Heat	Thunder	Pacers
Shooting	0.498	0.551	0.480	0.460
3 point	0.253	0.388	0.352	0.375
Free throw	0.761	0.757	0.814	0.802
Front court rebound	1.2	1.1	1.5	0.8
Back court rebound	4.9	5.0	6.2	4.7
Total rebound	6.1	6.1	7.7	5.5
Assists	3.7	4.1	4.3	4.0
Steal	0.9	1.3	1.2	1.2
Block shot	0.8	0.7	1.1	0.3
Fault	1.9	2.6	2.9	2.1
Foul	1.7	2.1	2.3	2.3
Score	14.8	20.4	22.1	15.8

From above TABLE 9, it is clear that spurs role player dominates in rebound, assists, steal aspects; heat role players absolute dominates in 3 point, free throw, fault aspects, while it is dominated in rebound and assists aspect; Thunder role player dominates in shooting, rebound and other aspects, while is dominated in 3 point, free throw, foul and other aspects; Pacers dominates in 3 point, while it almost lies in middle or lower

FULL PAPER

TABLE 9 : 2013-2014NBA regular seasons each team role player court average technical statistics[5]

	Spurs	Heat	Thunder	Pacers
Shooting	0.478	0.457	0.472	0.421
3 point	0.251	0.274	0.155	0.277
Free throw	0.678	0.746	0.540	0.696
Front court rebound	0.7	0.4	0.9	0.4
Back court rebound	2.0	1.5	1.9	1.6
Total rebound	2.7	2.0	2.8	2.1
Assists	1.1	0.7	0.8	0.8
Steal	0.5	0.4	0.4	0.4
Block shot	0.3	0.3	0.3	0.2
Fault	0.8	0.6	0.7	0.7
Foul	1.3	1.5	1.8	1.2
Score	5.3	4.5	3.3	4.2

TABLE 10 : 2010-2013 three seasons four teams winning rate[5]

	Spurs	Heat	Thunder	Pacers
2010-2011	74.40%	70.70%	67.10%	45.10%
2011-2012	75.80%	69.70%	71.20%	63.60%
2012-2013	70.70%	80.50%	73.20%	60.50%
Average value	73.63%	73.63%	70.50%	56.40%
Variance	0.000694	0.003561	0.000967	0.009817

levels in other technical aspects. In conclusions, it is clear that spurs, heat, thunder three teams' role player have equal performance, while pacers role player has the worst performance.

Team winning rate analysis

From above TABLE 10, it is clear that spurs and heat two teams winning rate are the same, slightly larger than thunder, far bigger than pacers, while spurs winning rate stability is the highest, the next is thunder, final is heat and pacer, the previous three teams winning rate stability is comparatively higher.

To convenient for analyzing four teams recent three seasons winning rate change status, it draws line graph as Figure 2 shows:

From above figure, it is clear that heat and thunder winning rate are in rising tendency, but heat rises faster, while thunder rises slower. Spurs winning rate totally is in reducing tendency, pacers winning rate is in rising tendency but average level is too lower by comparing with other three teams.

TABLE 11 : Leading player average age, average NBA years pro[5]

Team	Average age	Average NBA years pro
Spurs	31.6	10.4
Heat	31.8	11.4
Thunder	24.6	4.8
Pacers	27.6	6.2

TABLE 12 : Core player average age, average NBA years pro[5]

Team	Average age	Average NBA years pro
Spurs	35.7	14.0
Heat	30.7	11.0
Thunder	25.7	6.3
Pacers	25.3	4.7

Age and experience analysis (TABLE 11 and TABLE 12)

In order to more intuitional compare four teams' leading players' average age and average NBA years

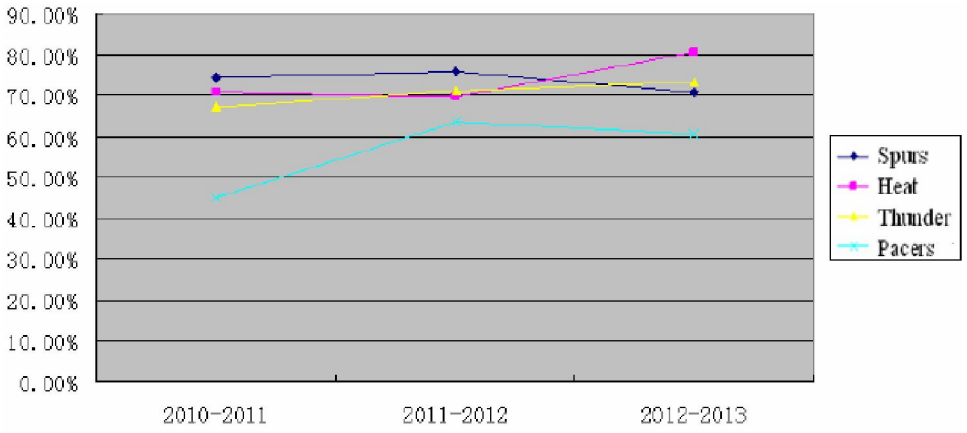

Figure 2 : 2010-2013 three seasons four teams' winning rate changes

FULL PAPER

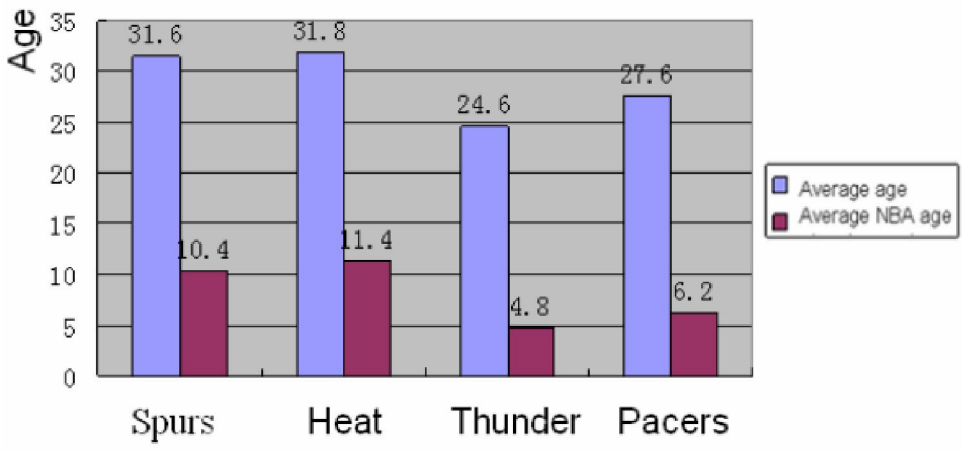

Figure 3 : Leading players' average age, average NBA years pro

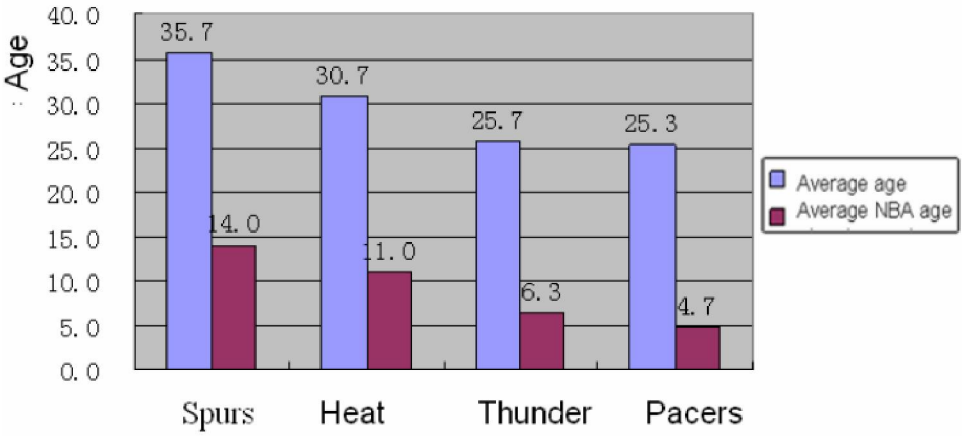

Figure 4 : Core players' average age, average NBA years pro

pro, it draws bar diagram as following Figure 3 shows:

From above Figure 3, it is clear that spurs and heat two teams' leading players' age and NBA years pro difference is not big that physical ability and experience have no big difference. While though thunder and pacers two teams are young, NBA years pro is relative small that experience is not as plentiful as spurs and heat two teams.

In order to more intuitional compare four teams' core players' average age and average NBA years pro, it draws Table 12 bar diagram as following Figure 4 shows:

From above Figure 4, it is clear that no matter average age, or average NBA years pro, heat core player all are in the golden age, young and strong and experienced. Spurs core players though have plentiful experiences, ages are relative older, physical ability is not vigorous, corresponding shorten appearing time. Thunder and pacers two teams core players are both young but experience is not rich.

CONCLUSIONS

From above two models synthesizing, it gets that spurs, heat, thunder three teams score abilities are the strongest, NBA regular seasons in 2013 to 2014 average each technical indicators analyzing, spurs and heat are equal, thunder is the secondary, pacers is the worst. From NBA regular seasons in 2013 to 2014 each team player factor analysis, it is clear that heat player has best performance. From winning rate analysis, it is known that spurs and heat two teams winning rate are the same and the maximum, meanwhile heat is in rising tendency that rises faster. From age and experience analysis, it is clear that no matter average age or average NBA years pro, heat core players are in the golden stage, young and strong as well as experienced, state are the best. From above deduction, NBA grand champion winners in 2013 to 2014 should be heat.

FULL PAPER

REFERENCES

[1] Guo Lei, Liu Jun; A Research on the Relationship Between the Technical Indicators and the Net Scores in CBA Regular Games Held during the 03-04 Season[J]. Journal of Capital College of Physical Education, **17(1)**, 84-86 **(2005)**.

[2] Qing Yong, Lin Chuang; Statistical Analysis between Technical Index and Scores of CBA Champion of the Last Five Years[J]. Journal of Xi'an Institute of Physical Education, **29(3)**, 371-375 **(2012)**.

[3] Zhang Xiang-yang, Xun Sun; Comparative Analysis on the Ability of Offense and Defense between Chinese Man''s Basketball Team and Opponents in the 16th FIBA World Basketball Championships[J]. China Sport Science and Technology, **47(1)**, **(2011)**.

[4] Jin Yi-Ping; Research on the Similarity and Difference Counter-analysis of the Strengths of CBA Men''''s Basketball Team Matches[J]. Journal of Beijing Sport University, **29(7)**, 1001-1003 **(2006)**.

[5] Zhang Feng, Zong Zhen-Ji; A Comparative Study of Offensive Technical Indicators Between Chinese Men''''s Basketball Team and Its Opponents during the 28th Olympics Games[J]. Journal of Capital College of Physical Education, **17(4)**, 62-63, 80 **(2005)**.

[6] Xu Zi-sheng, Ge Yao, Yang E-Ping; Relationship between several targets and league tournament rank during 2005-2006 CBA regular season[J]. Shandong Sports Science & Technology, **29(3)**, 37-39 **(2007)**.

[7] Liu Zhi-bo, Shen Xian, Ji Yin; Influence of Every Technical Indicator of LA Lakers Vs Detroit Pistons of NBA Finals During 03 - 04 Season[J]. Shandong Sports Science & Technology, **26(4)**, 14-16 **(2004)**.