

The Prediction of CBA League Title

Wu Jing¹ Yan Chang² Li Jiangpeng³

¹ Yanbian University School of Geography and Ocean Sciences Jilin Yanbian Korean Autonomous Prefecture 133000;

² Yanbian University School of Geography and Ocean Sciences Jilin Yanbian Korean Autonomous Prefecture 133000;

³ Yanbian University School of Science Jilin Yanbian Korean Autonomous Prefecture 133000)

Abstract: *The CBA professional league is currently the highest level of professional basketball league in China. It's the enjoyment for the audience to predict the winning team. This paper uses the comprehensive scores of fourteen teams in nearly one hundred games to establish a mathematical model through the normal distribution-integration method, and calculated that the 14 basketball teams participating in the CBA professional league won the competition this year. Probability, and then predict the top four teams. At the same time, using the original data given in the annex and its derived multiple indicators, combined with the correlation analysis of SPSS, qualitative analysis of the level of these fourteen teams. Finally, combined with the actual situation of the domestic CBA league, according to the research results, the top four teams have put forward effective championship recommendations.*

Key words: *normal distribution-integral method for probability; ranking prediction; feature matrix weighting; rationalization suggestion*

Date of Submission: 03-08-2019

Date of acceptance: 19-08-2019

I. Introduction

The China Basketball Association, or CBA, was founded in 1995 and is the highest level basketball league in China. From the original 12 teams to more than 20 teams today. Assume that there are 14 teams in the CBA League this year to participate in the regular season and the playoffs, each of which must have a winner, and each team number is fixed. Based on the comprehensive scoring data of each team's presence in each game, try to establish a mathematical model to solve the following problems:

1. Estimate the probability of each team winning the championship and predict the top four teams;
2. Qualitative analysis of the level of fourteen teams in the Chinese CBA League according to the specific situation of the attachment score;
3. Based on the actual situation of the domestic CBA league, according to the results of this study, from the perspective of the probability of winning the championship, the rationale for how to maximize the championship is proposed for the top four predicted.

The fourteen teams were numbered A, B, C, D, E, F, G, H, I, J, K, L, M, and the data was derived from the Mathematical Modeling Association of Jilin University.

I. Estimate the probability of winning, predict the top four teams

1. Problem analysis

In order to estimate the probability of winning each team, the collected data should firstly be analyzed. Due to the difference in the number of data of each team which is because of the fact that some teams did not participate in certain years, and the data fluctuated greatly in the vertical (inter-annual), the hypothesis was originally proposed by researching the new cab system published by the official website^[1]: There is no time link in the data. And because the overall level of a team depends mainly on the strength of the team, and the strength of the team's strength in the short-term fluctuations will not be very large, it is again assumed that the overall level of each team's game is normally distributed. Let the comprehensive score data of a certain team be X, the comprehensive score data of another team be Y, and when $X > Y$, the former team wins. According to this, the probability density function can be constructed to directly calculate the winning probability of any team.

2. Data processing

In the calculation and analysis of data, the abnormal values tend to have adverse effects that cannot be ignored, so the elimination of these values is crucial. Using the function of eliminating the outliers in the SPSS, the original data in the attachment is imported to generate a new variable, and if the absolute value of the new

variable is greater than 2, the data is eliminated. This reduces errors in data calculations and yields valid data that truly reflects database information.

3. Building a model

According to the processed data, using the built-in functions $f = \text{mean}(X)$ and $f = \text{var}(X)$ in MATLAB software, the probability density function of a team's score data can be obtained:

$$f(x) = \frac{1}{\sqrt{2 * \pi * \text{var}(X)}} e^{-\frac{(x - \text{mean}(X))^2}{\text{var}(X)}}$$

Where $\text{Var}(X)$: the variance of the elements in the X vector

4. Calculate the probability of winning and losing between the two teams

If there is a match between the two teams, it is assumed that there is a certain constant between the teams, and the larger of the constants will win. The constant exists in the abscissa of the probability density function and is distributed by probability. Let any two teams z and v, their historical score data is vector x and vector y.

Assume $f(x) = \frac{1}{\sqrt{2 * \pi * \text{var}(X)}} e^{-\frac{(x - \text{mean}(X))^2}{\text{var}(X)}}$ with $f(y) = \frac{1}{\sqrt{2 * \pi * \text{var}(y)}} e^{-\frac{(y - \text{mean}(y))^2}{\text{var}(y)}}$ The

intersection of the abscissa is $d(1,1)$ and $d(2,1)$.

When $\text{mean}(X) < \text{mean}(Y)$:

$$P_y = \int_{-\infty}^D f(x) * \int_D^{\infty} f(y) + \int_D^{t_1} f(x) * \int_{t_1}^{\infty} f(y) + \int_{-\infty}^{t_2} f(x) * \int_{t_2}^D f(y)$$

among them: $t_1 = 3 * \sqrt{\text{var}(X)} + \text{mean}(X)$ $t_2 = \text{mean}(Y) - 3 * \sqrt{\text{var}(Y)}$

Then: $P_x = 1 - P_y$

When $\text{mean}(Y) < \text{mean}(X)$:

$$P_x = \int_{-\infty}^D f(y) * \int_D^{\infty} f(x) + \int_D^{t_2} f(y) * \int_{t_2}^{\infty} f(x) + \int_{-\infty}^{t_1} f(y) * \int_{t_1}^D f(x)$$

among them $t_1 = 3 * \sqrt{\text{var}(X)} + \text{mean}(X)$ $t_2 = \text{mean}(Y) - 3 * \sqrt{\text{var}(Y)}$

Then: $P_y = 1 - P_x$

Where P_y indicates the probability of winning the championship in the basketball team corresponding to the data y

P_x indicates the probability of winning the championship in the basketball team corresponding to the data x

5. Each team wins the probability of winning

First, the team a is regarded as the benchmark and the other teams are calculated to find the probability that the other teams will defeat the team a. The probability that b, c, d, e, f, g, h, i, j, k, l, m, n are victorious with respect to a is recorded as: $p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}, p_{13}, p_{14}$.

Then the probability that the i-th ($i \geq 2$) group wins the championship is:

$$p[i] = \frac{p_i}{\sum \frac{p_i}{1 - p_i} + 1}$$

among them: $p[1] = \frac{1}{\sum \frac{p_i}{1 - p_i} + 1}$

$P_y = 1 - P_x$

According to the calculation, the probability of winning each team is: $p[1] = 5.1593\%$, $p[2] = 8.54669\%$, $p[3] = 7.3299\%$, $p[4] = 5.2375\%$, $p[5] = 7.8633\%$, $p[6] = 6.0370\%$, $p[7] = 6.8024\%$,

P[8]=7.3740%、P[9]=6.2704%、P[10]=6.5540%、P[11]=7.3299%、P[12]=9.195%、P[13]=6.6124%、P[14]=9.6875%

Table 1 The probability of winning each team

Team Name	a team	Team b	Team c	Team d	Team e	Team f	G team
Probability of winning	5.1593%	8.5467%	7.3299%	5.2375%	7.8633%	6.0370%	6.8024%
Team Name	h team	Team i	Team j	K team	l team	m team	n team
Probability of winning	7.3740%	6.2704%	6.5540%	7.3299%	9.195%	6.6124%	9.6875%

The probabilistic results are arranged in order, and the first four names are predicted: n team, l team, b team, e team

II Qualitative analysis of the level of fourteen teams

1. Problem analysis

Qualitative analysis is the analysis of the "quality" of the research object. By finding the mean (m) and variance (S^2) of each team, the score matrix is used, and the weights w and t are obtained by considering the variability of the probability. And determine the level of the fourteen teams by fitting the correlation between w, t and the probability of winning (P[i]).

Sexual analysis. Using the score matrix, the relative strength value obtained is combined with the variance as an indicator of the evaluation. Analyze the level of the team. You can also construct a model between the probability of winning and the weight of the mean and variance to analyze the team's level.

2. Qualitative analysis

On the basis of the existing assumptions, the default data is not directly related to the time, and the data between the groups is randomly given in the historical database of the basketball team. Therefore, the analysis data cannot be established on the basis of time. The attachment data we know is for nearly 100 games in each group.

In the data, the mean and variance of each group are the most direct indicators. By judging the average and variance of a certain team and comparing with other groups, only the strength level of the two groups can be reflected. Therefore, the mean weight and the variance weight between all groups can be established by the operation of the matrix. Because this is the two main reasons that affect strength, and the probability of winning the championship is based on relative strength, the probability of winning the championship for each group is approximately equivalent to the team level.

Table 2 Qualitative analysis of various indicators

index Team	team A	team B	team C	team D	team E	team F	team G
Probability of winning (P[i])	5.1593%	8.5467%	7.3299%	5.2375%	7.8633%	6.0370%	6.8024%
Mean (m)	3.40773	5.486278	5.878735	4.89899	3.90802	4.79936	4.15482
Variance (s^2)	5.74115	4.833452	7.490090	5.19812	2.48921	5.91575	5.520386
Weight of mean (w)	0.0508	0.0806	0.0864	0.0720	0.0547	0.0705	0.0610
Weight of variance (t)	0.0813	0.0703	0.1089	0.1107	0.0326	0.086	0.0803
index Team	team H	team I	team J	team K	team L	team N	team M
Probability of winning (P[i])	7.3740%	6.2704%	6.5540%	7.3299%	9.195%	6.6124%	9.6875%
Mean (m)	5.0114	4.26316	4.45746	4.9455	6.25445	4.1616	6.393854
Variance (s^2)	5.8792	2.75008	3.89462	5.0101	4.98170	3.6357	3.183395
Weight of mean (w)	0.0736	0.0626	0.0655	0.0726	0.0919	0.0611	0.0939
Weight of variance (t)	0.0855	0.04	0.0566	0.0728	0.0724	0.0529	0.0463

Observe the indicators in the table to find:

1). Team N gets the highest probability of winning the championship. Its average value is relatively high and the variance is relatively small, indicating that team N is indeed very excellent. 2). Team E is predicted to be the fourth which has a strong correlation between the winning percentage and the variance. The original data, E

team's comprehensive score fluctuations are small, and there is no negative score in the score. Therefore, team E has a stable and high standard;3).The remaining teams L and B of top fours ,up to their mean values , combined with the original data, the two teams have large fluctuations but no negative points.so that these two teams are also outstanding.

In addition, the SPSS can also be used to import the calculated probability of winning (p) to construct the regression equation, and the correlation coefficient is derived as shown in Figure 1. The p-w relationship is shown in Figure 2. The p-t relationship is shown in Figure 3.

Figure 1. Correlation coefficient

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.120	1.056		2.955	.013
	w	87.434	14.005	.842	6.243	.000
	t	-31.027	7.857	-.533	-3.949	.002

a. Dependent Variable: p

It can be seen that the reliability level is good, according to which the model $p=3.12+87.434w-31.027t$ can be constructed.

Figure 2. P-w diagram

Partial Regression Plot

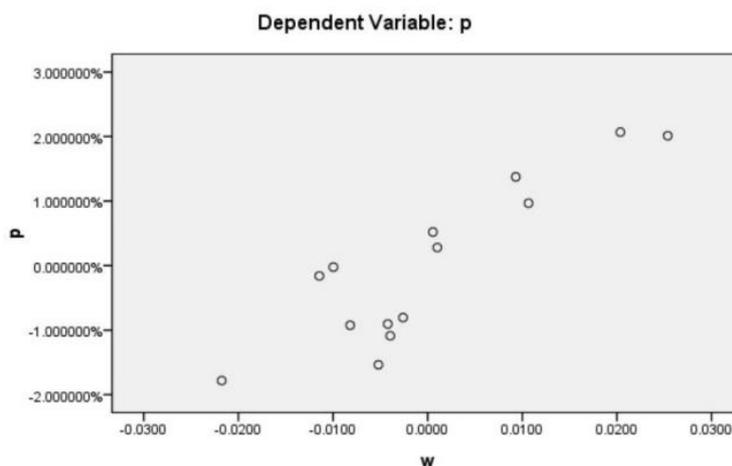
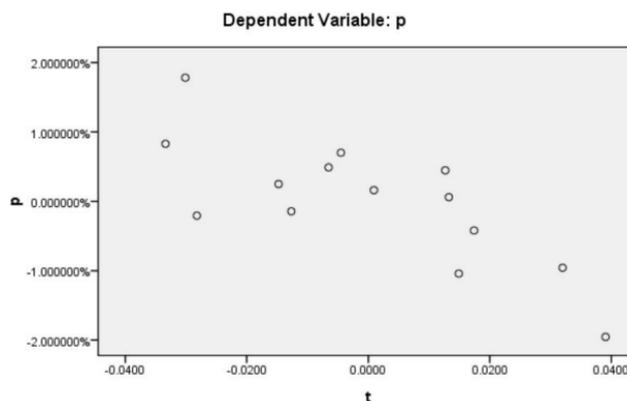


Figure 3. p-t diagram

Partial Regression Plot



According to the figure, the degree of crowning and the weight of the mean and the weight of the variance are well fitted. That is, the probability of winning is inversely proportional to the weight of the variance and proportional to the weight of the mean. Because the latter two are effective indicators to reflect the team's level, the probability of winning indirectly reflects the level of the team, and its ranking can be used as a reference for the team's horizontal ranking. Therefore, the level of the team from high to low is: N, L, B, E, K, H, C, G, M, J, I, F, D, A.

III The top four winners of the rationalization proposal

It is necessary to combine the actual situation of the domestic CBA league to propose the rationalization proposal for the top four teams to win the championship. Find relevant documents and find the main factors affecting the team's assessment: offensive, defensive, combining, coaching, endurance, etc.^[2] The main factor is the team's average score. In other words, there are three main factors that determine the overall score of the team's presence: the team's average score, the player's skill level, coaching level or foreign aid. Starting from these three aspects, combined with the CBA basketball league system, the four top teams have put forward the rationalization proposal of how to win the most.

The advice for a team is based on the team's existing data and is based on an analysis of the team. Based on the conclusions of the previous papers and the research results of the scholars, the paper proposes suggestions on how to win the top four questions from the following four perspectives:

1. Average scoring. In the case where each player's playing time is the same, the coach should try to score players with higher rates play, play the maximum performance of each player, improve the team's offensive efficiency and effective life. The medium rate reduces the team's intrusion rate, free throw rate, foul rate, and turnover rate. According to the analysis of question two, team N which has the highest probability of winning, the training of scoring items should be strengthened. For the second and third rankings for the team L and B, the performance of these two teams in each game is not very stable. The coach should focus on strengthening the actual training to ensure the stability of the game. For the last team——team E, played more stable, but the overall level is not as high as the first three teams, and technical training should be strengthened.

2. Home and away win rate. According to the survey data, CBA has a home win rate of more than 60% in each season, and the away win rate is below 40%. The home win rate of each season is significantly higher than the home court CBA home advantage, the home field effect is outstanding^[4]. According to this rule, team coaches should make reasonable arrangements for the players at home and choose reasonable tactics.

3. Game system. The organizer should appropriately increase the number of rounds of the game. As is known, the winning percentage of the top four basketball teams is higher than other teams. If the teams with different strengths compete, the team with better strength also has the probability of failure inevitably. If you increase the number of matches between teams, the rankings after the game will accord to the strength of the team increasingly. In the regular season of the CBA, the rounds of the game are appropriately increased.

Suppose a winning team has a probability of winning $p=2/3$

If a match is negative, the winning probability of the strong team is $p_1=2/3$

If the three games win two systems, the probability of winning the strong team is:

$$P_2 = 1 - (1 - P)^3 - 3 * P(1 - P)^2 = 7/20$$

Obviously: $p_1 < p_2$

By analogy, it is found that the number of wins in the first four basketball teams will increase.

4. At the end of the regular season, the fifth to tenth place will be played, the top four will be selected, and the playoffs will be held. After the regular season, eight top teams will be selected for the playoffs. The significance of the game is to eliminate the weak team and leave a strong team. For these fourteen teams, after all rounds of the regular season, the odds must be eliminated. If you want to increase the probability of winning the first four teams, you need these four strong basketball teams and weaker teams to play (according to recommendation 3). We can eliminate the top two teams after all rounds of the regular season. The 8 teams between the 5th and 6th teams will play the points game and choose the four teams to win the playoffs with the first four teams.

IV Model evaluation

1. The shortcomings of the model

1). When the Matlab software processes the data, due to the accuracy of the software, the data will have rounding errors and truncation errors;

2). In the construction of the first model, when the mean and variance of the two teams' comprehensive score data are similar, there will be certain errors;

3). When using SPSS for correlation analysis, due to insufficient data, the accuracy of the fitted model is not enough;

2. Advantages of the model

- 1). The first model relates the integral operation to the probability density function of the normal distribution, which has certain originality;
- 2). When solving the problem 2, the matrix calculation weight is used which simplify complicated calculation;
- 3). Using SPSS to reflect the mean weight, variance weight and team strength on the image indirectly.

References

- [1]. <http://www.cba.net.cn/>
- [2]. Xia Hanmin. Team evaluation system based on fuzzy neural network[J]. Computer Simulation, 2007, 24(12): 163-165. DOI: 10.3969/j.issn.1006-9348.2007.12.043. [3] <https://m.taodocs.com/p-86529320.html>
- [3]. Li Liyang. Research on cba main field effect [d]. Capital Institute of Physical Education, 2014.
- [4]. Fan Minyun, Wang Wei, Zhou Wu. Research on the Influence of the Introduction of Foreign Players on CBA League[J]. Journal of Beijing Sport University, 2007(11): 1561-1563

IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG) is UGC approved Journal with Sl. No. 5021, Journal no. 49115.

Wu Jing. " The Predicton of CBA League Title. "IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG) 7.4 (2019): 45-50.