The Probability of Winning the Championship is Calculated Based on the Combination of Statistical Features and Symbol Test

Liu Yingjie[*] and Piao Zhe

Department of mathematics, yanbian university, yanji, China.
[*]Corresponding author email id: 15011027040@163.com

Abstract – **This paper analyzes the comprehensive score data of 14 teams, and puts forward a method to calculate the probability of winning the championship. Firstly, the data is pre-analyzed, processed and grouped. Secondly, entropy weight method is used to calculate entropy value and variability coefficient of decision matrix, so as to determine entropy weight. The recent state is then defined using a notation check. Finally, the team's probability of winning the championship is obtained by integrating the team's variability coefficient and recent state.**

Keywords – **Entropy Weight Method, Symbol Test, Statistical Characteristics, Predict.**

I. INTRODUCTION

In the process of rapid social development, entertainment has become more and more important in life, and football and basketball have become more and more popular. In addition to watching the game, there are more and more predictions on the victory or defeat of the game. With the improvement of technical level, the accuracy rate is getting higher and higher. In the study of scholars, it is important to analyze and compare the important factors of winning in the process of confrontation, and then get the main factors that affect the winning and losing. In the bookmakers, various data of teams in recent games, such as goals scored, points scored, steals and passes, are quantified, and further data analysis is carried out to obtain comprehensive scores for comparison. Based on the data of 2019 jilin province college students' mathematical modeling problem A, this paper analyzes the winning situation of each team by using the given comprehensive score data and the method based on statistical features and symbol test in statistics.

II. THE DATA ANALYSIS

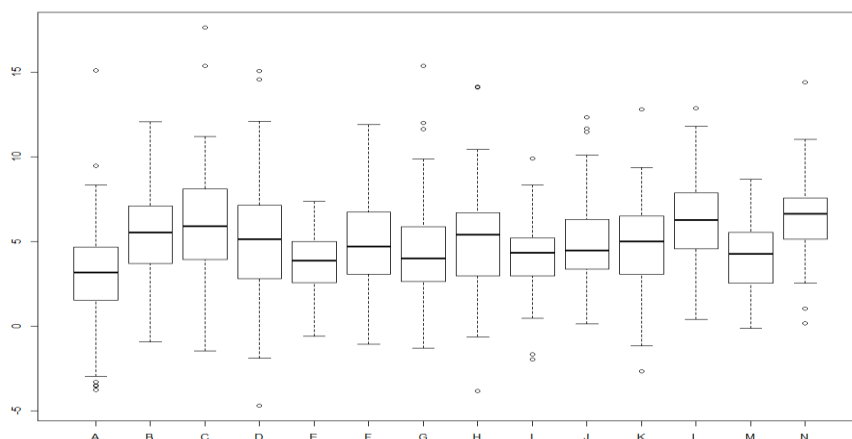2.1. *Remove Outliers by Referring to the Boxplot*

Fig. 1. Boxplot

2.2. *Give the Basic Statistics of Each Season*

Analyze the average, variance, skewness and kurtosis of 14 teams over 5 seasons, the average level of 14 teams is similar, there is no significant difference. Although the stability is slightly different, but the comparison of skewness and kurtosis, the teams have no obvious winning or losing deviation.

III. THE PROBABILITY CALCULATION BASED ON STATISTICAL FEATURES AND SYMBOL TEST

This paper proposes a championship index to measure the probability of each team winning the championship. According to the data and actual situation, whether a team can win the championship or not is affected by two factors: long-term stability and short-term status. This paper USES the coefficient of variation to measure the stability of the team. Recent states are measured in a manner similar to sign checking [1]. The two parts are then given different weights to get the final score, and the probability is calculated from this [2]. The specific calculation formula is as follows: e-champion = $\alpha CV + \beta K$

3.1 *Calculation of Weighted Variation Coefficient*

Since the 100 games are divided into four seasons, each season has different performances and stability. So it's going to have a different impact on the present, so it's going to have a different weight. The weighted variation coefficient was obtained. In order to eliminate subjectivity and maintain objectivity, this paper starts from data and USES entropy weight method to assign values.

1. *Basic Solution Ideas of Entropy Weight Method*

(1) Form a decision matrix

(2) Standardized decision matrix

According to the nature of the data in this paper, the smaller, the better type index is selected. Use the following formula for standardization:

$$V_{ij} = \frac{\max(x_j) - x_{ij}}{\max \ (x_j) \ - \min \ (x_j)}$$

(3) Calculate the characteristic proportion of the ith evaluation object under the j index.

The greater the value difference of a certain indicator j, the greater the effect of the indicator on the evaluated object v_{ij}, that is, the more useful information the indicator provides to the evaluated object. Under the j index, the characteristic proportion of the i evaluation object p_{ij}:

$$p_{ij} = \frac{v_{ij}}{\Sigma_{i=1}^{m} v_{ij}}$$

(4) Calculate the entropy value of the j index e_j:

$$e_j = -\frac{1}{\ln(m)} \sum_{i=1}^{m} p_{ij} \ln \ (p_{ij})$$

(5) Calculate the difference coefficient of the j index d_j:

$$d_j = 1 - e_j$$

(6) Determine the entropy weight of each index:

$$w_j = \frac{d_j}{\sum_{k=1}^{n} d_k} \quad j = 1, 2, \ldots, n$$

2. *Specific Results:*

(1) The proportion of calculated features is:

Table 1. The Eigenvalue

0	0.002344932	0	0
0.088001032	0.08421367	0.073968117	0.085650898
0.081099366	0.056070289	0.073777757	0.064255039
0.051206207	0.044064782	0.071816679	0.031665476
0.075140525	0.109331617	0.096033717	0.081614491
0.087539516	0.062583837	0.062735638	0.051492131
0.049621227	0	0.067945154	0.087808545
0.051946014	0.079260911	0.053483489	0.084670952
0.082997624	0.108165976	0.074186964	0.080276587
0.071975966	0.092929014	0.070430024	0.08581119
0.07245535	0.074146628	0.082908118	0.053130462
0.097961402	0.107055783	0.096569243	0.093552898
0.082921873	0.060424687	0.078558529	0.091336871
0.107133897	0.119407874	0.09758657	0.108734461

(2) The Final Weight is:

0.205398887741022	0.380619855003571	0.18184501887143	0.232136238383973

(3) Final enhancement coefficient of 14 teams was calculated:

Table 2. Coefficient

A	0.9085
B	0.4653
C	0.5433
D	0.6517
E	0.4305
F	0.5623
G	0.6186
H	0.5440

I	0.4515
J	0.4793
K	0.5395
L	0.3840
M	0.4851
N	0.3313

3.2 *Sign Check Defines Recent Status*

1. Purpose of symbol test definition:

Let's say that the median of a population is M_0, if the median of a sample M is equal to M_0, then we accept that the sample is from a population. Their specific inspection method is as follows: first of all, in each sample observation value minus the median overall, it is concluded that the positive and negative difference with positive (+) and negative (-) represented, if the overall median is equal to the sample median, namely, $M = M_0$, then, under the sample observations on the median, should the number of each half, therefore the probability of a plus or minus sign should be each half, a sample size of n, and can use the binomial distribution B (n,1/2) , to calculate a minus sign (or plus) number of probability, which, according to a significant level of α certain assumptions about whether to accept the original decision [3].

2. Therefore, it is assumed that the recent state of the team is better than the median, and sign test is conducted. To get the feasibility of such a determination.

Table 3. The Inspection Results.

Team	Symbol test P-value
A	0.5841
B	0.009605
C	0.002172
D	0.009605
E	0.1917
F	0.1431
G	0.5
H	0.002172
I	0.1189
J	0.02624
K	0.005909
L	0.002172
M	0.01544
N	0.002172

For a given significance level, if the p-value is greater than the significance level, the recent state is good. If it's less than the significance level, it's bad.

3. Therefore, the results of the selected team in the recent 7 games are different from the median, and the symbolic function is used. If it is greater than 0, it is denoted as 1; If it's less than 0, it's negative 1.

Get the following Table:

Table 4. Symbol

Ream\number	7	6	5	4	3	2	1	In shift ratio K
A	-1	1	-1	1	1	1	-1	0.4286
B	-1	-1	-1	1	1	-1	1	0.5714
C	-1	-1	1	1	1	-1	-1	0.5714
D	-1	1	1	-1	1	-1	-1	0.5714
E	-1	-1	-1	-1	-1	1	-1	0.8471
F	1	-1	-1	-1	1	-1	1	0.5714
G	1	-1	-1	-1	1	-1	1	0.5714
H	-1	1	1	-1	-1	1	1	0.4286
I	-1	-1	-1	1	-1	1	1	0.5714
J	-1	1	-1	-1	-1	1	1	0.5714
K	-1	-1	1	1	1	-1	1	0.4286
L	-1	1	-1	1	-1	1	-1	0.5714
M	-1	1	1	1	-1	1	-1	0.4286
N	1	1	1	-1	-1	-1	-1	0.5714

In the basketball game, the team competing for the championship has to make the playoffs, and the team in poor condition cannot enter the playoffs this season. Therefore, four teams of AMKH were deleted according to the recent status. Then delete D and F teams according to last season and this season's combined state. Based on the data, only eight teams are given the probability of winning the championship.

3.3. *Final Calculation:*

State to different coefficients. Finally, by integrating various factors, a and b are selected as (0.3, 0.7) to obtain the final probability of winning the championship [4]:

Table 5. Result

Team	Value	Percentage
B	0.539578329	13.17%
C	0.562975679	12.50%
E	0.722106022	7.95%

Team	Value	Percentage
G	0.585558	11.85%
I	0.535440405	13.28%
J	0.543755724	13.05%
L	0.515189319	13.86%
N	0.499377384	14.33%

IV. CONCLUSION

The final champion of this season is N team, and the eight teams qualified for the final playoff are B team, C team, E team, G team, I team, J team, L team, N team, the final four is N team, L team, I team, B team, the predicted champion is N team. Compared with other methods [5], the starting point of this method is only data, and it does not delve into the internal factors of motion.

REFERENCES

[1] Lu yunqing. Influence of consistency between samples in symbol test and continuity correction statistics and decision-making. 2017 (4) : 90-92.
[2] Tang jianbin. Incident research based on empirical likelihood method southwest university of finance and economics. 2013.
[3] Xue yi, R statistical modeling and R software. Tsinghua university press. 2006: 397-411.
[4] He fengbiao, realization of comprehensive evaluation method MATLAB. China social science press. 2010: 327-333.
[5] Luo hua. Analysis of winning factors of champions league in 2016-2017 season. Shaanxi normal university, 2018.

AUTHORS PROFILE'

First A. author
Liu Yingjie, is a student majoring in statistics, department of mathematics, school of science, yanbian university.

Second B. author
Piao Zhe, is a student majoring in statistics, department of mathematics, school of science, yanbian university.