Zhang H and Wang A. (2024) LEVERAGING STATISTICAL THEORY IN SPORTS COMPETITIONS: AN ANALYSIS OF PROBABILISTIC MODELS AND MULTIPLE REGRESSION WITHIN THE FRAMEWORK OF BIG DATA. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 24 (95) pp. 19-41.
DOI: https://doi.org/10.15366/rimcafd2024.95.002

ORIGINAL

LEVERAGING STATISTICAL THEORY IN SPORTS COMPETITIONS: AN ANALYSIS OF PROBABILISTIC MODELS AND MULTIPLE REGRESSION WITHIN THE FRAMEWORK OF BIG DATA

Hanzhe Zhang [1], Ailing Wang [2,*]

[1] Faculty of Sports Science and Technology, Bangkok Thonburi University, Bangkok 10170, Thailand
[2] Physical Education Department, Qingdao University of Technology, Qingdao 266000, Shandong, China
E-mail: zhz_15053239396@163.com

Recibido 02 de Junio de 2023 **Received** June 02, 2023
Aceptado 02 de Febrero de 2024 **Accepted** February 02, 2024

ABSTRACT

CBA is a sports event that allows fans to enjoy themselves and players to give full play, and traditional Chinese cultural values have a profound influence on it. This paper takes the 100 sets of historical rating data of the fourteen teams in CBA league as the basic basis, firstly, we simply deal with the 100 sets of historical rating data and use Excel function formula to find out the mean, extreme deviation and variance of each team, then we carry out SAS normal test, and we find that except for the very few data with large deviation, the historical rating data satisfy the normal distribution. Through the outlier algorithm to screen the values, compare the confidence intervals as well as carry out hypothesis testing, to objectively and scientifically explore the probability of each team winning the championship in the CBA league. Compare the probability of winning the championship of these fourteen teams and predict the top four teams in the CBA league to ensure that the prediction results are as reasonable as possible. With the help of hierarchical analysis to qualitatively analyze the level of each team, and then through cluster analysis to compare these data, and combined with the trend of the development of the world's basketball movement, the use of multiple regression and SPSS to analyze the level of the team's factors, in-depth thinking about the league, a more reasonable to give a more scientific to improve the probability of the team's winning the championship, and to promote better development of the basketball movement.

KEYWORDS: Interval estimation; hierarchical analysis; probabilistic modeling; multiple regression; statistical software

1. INTRODUCTION

Nowadays, big data has an irreplaceable role and status in the new era of sports competition, and the use of big data for sports analysis has become the key to winning for basketball teams in the world's sports leagues. The American NBA tournament has already established a set of mature data analysis system (JING & FANG, 2019). The official website has sufficient data resources, rapid updating, analyzable data visualization charts, while the stadium tracking equipment can record the trajectory of any ball running in the game and can be presented (Ju, 2008), greatly sublimating the fans' viewing experience in the broadcast screen, the video data to make up for the cognitive differences that may be caused by a single piece of data, and the data analysis team will be able to deal with these complete data resources to ultimately form the team's needs of the data report, which can be applied to the team's game, operation, training, and training. The data analysis team will process these data resources and finally form the data report needed by the team, which can be applied to the team's game, operation, training and other aspects to help the team improve its strength and promote the development of the NBA league (SUN, CHEN, & FANG, 2001). The CBA league has made a lot of achievements in data analysis in recent years, but there is still a big gap compared with the NBA league, which is mainly reflected in the lack of awareness of paying attention to the data, the scarcity of data analysis talents (Tong, 2002), the weak application of high-tech products, the limited access to data resources, insufficient data dimensions (L. Gao, Dai, Xu, Zhou, & Li, 2021), and the lack of data application level. Insufficient data dimension, narrower data application level and other issues.

As a result, a data analysis system with the characteristics of CBA league has not been formed. The above problems can be solved by the following aspects, firstly, CBA league can invite domestic and foreign data analysis experts to hold regular training courses to strengthen the team's awareness of data analysis, and at the same time cultivate data analysis talents (J. Li & Ma, 2017). Secondly, CBA League can actively cooperate with sports science and technology companies to improve the high-tech application level of the league, and learn from sports science and technology companies so as to do "inviting in and going out", and develop a data analysis system suitable for its own application according to the actual situation of CBA League (Zhang & Mao, 1995). Finally, increase the investment of funds, the introduction of video tracking equipment, expanding the application of data, the establishment of youth training game video library (Duan, 2020; C. Gao, Dynasty, & Li, 2017), data statistics website, help youth basketball training, digging basketball talent, scientific selection for the CBA league to play a role in the decision-making

process, to enhance the overall competitiveness of the league (CAO, WANG, & TU, 2015; Huo, 2021).

2. Research Significance

High-tech era of "big data" under the background of the inevitable requirements of the development of competitive sports since the new century, the development of high-tech impact on all aspects of our social life, cloud computing as a representative of scientific and technological innovation means for the long and heavy data to give a new value. Science and technology is the first productive force, also in the field of sports, in the international sports powerhouse, the development of science and technology and scientific and technological innovation has always been the source of sports development.

The ultimate goal of competitive sports is to win the game, and the prerequisite for winning the game is who is more adept at utilizing resources to help their own, "the wise man seeks to follow the times, the fools move against the times" Today the world's new round of scientific and technological revolution is in full swing, competitive sports only to grasp the pulse of the times in line with the times, embrace the development of "Big Data" will be the inevitable choice for the sustainable development of competitive sports. CBA professional league to the world basketball high level development of the reality of the need for big data era, the use of science and technology fusion data analysis has been widely used in all areas of life. NBA as the highest level of the world's basketball professional league, in all aspects of the world's basketball development direction, the NBA is now the world's highest level of professional league, the NBA is now the world's highest level of basketball development direction.

3. Study of the problem

3.1 Research hypothesis

1. assuming that there are no surprises in the predicted next game, and that there are no sudden cold streaks or comebacks by the teams. 2. assume that the umpire for that game is the same person as the historical scoring umpire. 3. Assume that the data given in the question is true and reliable. 4. Assume that all the data given are due to the real strength of the teams, and that there is no fake soccer.

3.2 Probabilistic model

According to the CBA team data obtained through the National Sports Bureau and various clubs, pre-excluding their own relevant factors affecting the playing time play, each team data presented by the trend of the known amount of time elements, so choose to use SPSS for time series analysis, the specific

results are shown in Figure 1-1: It seems like you're describing a statistical analysis conducted on Chinese Basketball Association (CBA) team data using SPSS for time series analysis. However, you mentioned Figure 1-1 without providing any additional information or context.

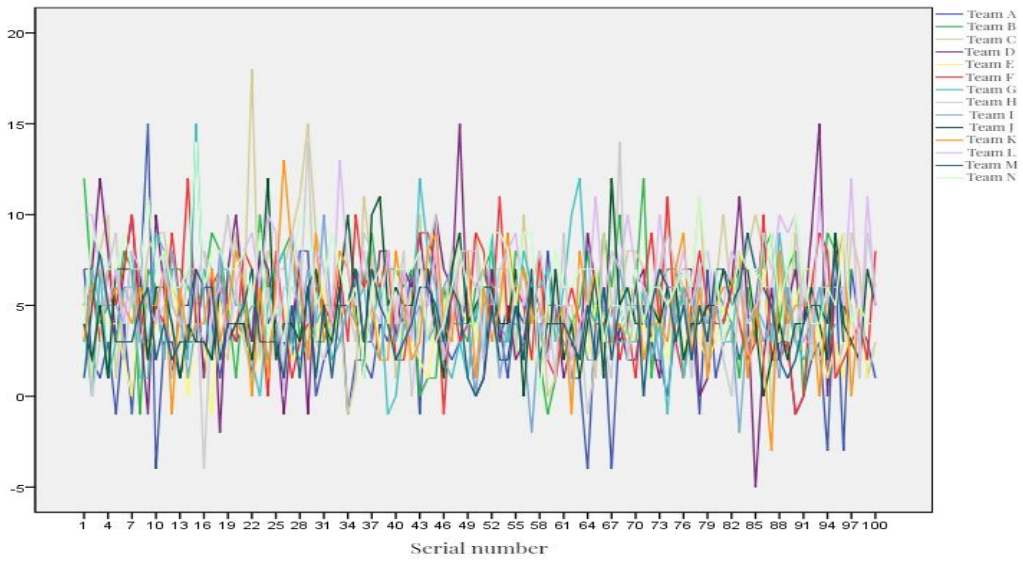

Figure 1

The article by comparing the data with time in the analysis of the results, found that the data and the time element it's not cyclical and there is no obvious pattern, but we accidentally found the data by analyzing the image that there are extremely fluctuating numerical points, meaning that there may be sample points that do not reasonably represent the overall data. The data was selected to obtain the mathematical expectation and variance for each team, and the results are as follows:

Table 1

N	1	2	3	4	5	6	7
HAVE EXPECTATIONS	3.18	5.38	5.94	5.01	3.79	4.89	4.43
N	8	9	10	11	12	13	14
HAVE EXPECTATIONS	5.05	4.16	4.79	4.83	6.32	4.12	6.45

Table 2

N	1	2	3	4	5	6	7
VARIANCE (STATISTICS)	3.01	2.58	3.25	3.35	1.70	2.76	2.80
N	8	9	10	11	12	13	14
VARIANCE (STATISTICS)	2.90	1.97	2.41	2.59	2.43	2.02	2.19

The presence of sample points with extremely fluctuating values was clearly observed by comparing the data with the desired expectation and variance, and a discrete factor test was chosen to examine the degree of

dispersion in the data for each team:

Definition 1: kth distance of object p: Choose an integer k. The kth distance of object p can be denoted as $k-distance(p)$. There exists an object o in the dataset and denote the distance between objects p and o as $\mathrm{d}(\mathrm{p},\mathrm{o})$. If the following condition is satisfied:

$$\text{Then } \mathrm{d}(\mathrm{p},\mathrm{o}) = k-distance(p).$$

1. in the set of numbers, in addition to object p, there exists at least k objects $ó$, satisfying $\mathrm{d}(\mathrm{p},ó,) \le \mathrm{d}(\mathrm{p},\mathrm{o})$; and

2. in the set of numbers, in addition to object p, there exists at most $k-1$ objects $ó$, satisfying $\mathrm{d}(\mathrm{p},ó,) < \mathrm{d}(\mathrm{p},\mathrm{o})$;

Definition 2: The reachable distance of object p with respect to object o, expressed as follows:

$$reach-dis_k(\mathrm{p},\mathrm{o}) = \max\left\{k-dis(\mathrm{p}),\mathrm{d}(\mathrm{p},\mathrm{o})\right\}$$

Definition 3: The local reachable density of an object p, expressed as follows:

$$lrd_k(\mathrm{p}) = \frac{k}{\sum_{o\in N_k(\mathrm{p})} reach-dis_k(\mathrm{p},\mathrm{o})}$$

After calculating its density value, from the above definition, it can be seen that when a fixed positive integer k is selected, the region in which the object p is located is relatively sparse, then the value of $k-distance(p)$. is relatively large, and thus the value of $reach-dis_k(p,o)$ is large; thus the value of $Ird_k(p)$ is small.

As the data conforming to the normal distribution is locally reachable to have a relatively small density, it is possible to round off the values outside of the 3σ, i.e., the random variables are as follows:

$$\begin{aligned} \mathrm{P}(|\mathrm{X}-\mu|<\mathrm{k}\sigma) &= \Phi(\mathrm{k})-\Phi(-\mathrm{k}) \\ &= 2\Phi(\mathrm{k})-1 \\ &= \begin{cases} 0.6826, k=1 \\ 0.9545, k=2 \\ 0.9973, k=3 \end{cases} \end{aligned}$$

Through the data fluctuation value of the data of the larger data rounding, to get a more scientific analysis of the data, at the same time, due to rounding off the value of the region is small, the number of samples lost is small, in order

to make the analysis of its data more accurate,

In the statistical theory that does not affect the trend of the situation, the choice of rounding off the range outside the range of the sample value. In order to further analyze the data statistically, the data were processed through SAS, and the normality test was performed.

Through the SAS test of the data, combined with the P-value of 0.0174 and 0.1012, against the test criteria of statistics, it was basically concluded that the group conformed to normality in the data analysis.

Through the collation and analysis test of the data of each group, it is generally considered that the test of normality is met, which is further verified by SPSS, proving that statistical analysis can be carried out with the help of the normality of the data.

Table 3

NORMALITY TEST							
	TEAM A	KOLMOGOROV-SMIRNOV[A]			SHAPIRO-WILK		
		STATISTIC	DF	SIG.	STATISTIC	DF	SIG.
V1	-4	.260	2	.			
	-1	.178	6	.200*	.908	6	.426
	0	.279	3	.	.939	3	.522
	1	.215	9	.200*	.926	9	.446
	2	.223	13	.076	.928	13	.323
	3	.175	14	.200*	.873	14	.046
	4	.138	15	.200*	.950	15	.524
	5	.160	7	.200*	.972	7	.914
	6	.146	4	.	1.000	4	.999
	7	.213	6	.200*	.955	6	.781
	8	.338	3	.	.852	3	.246

*. This is the lower limit of the true significant level.

A. lilliefors significant level correction

B.v1 is constant when team a = -3. It has been ignored.

It has been ignored. D. V1 is constant when team a = 9. It has been ignored.

At the same time with the help of MATLAB software to plot each group of data can be plotted normal distribution graph, to the first group of data plotted graphical results as an example:

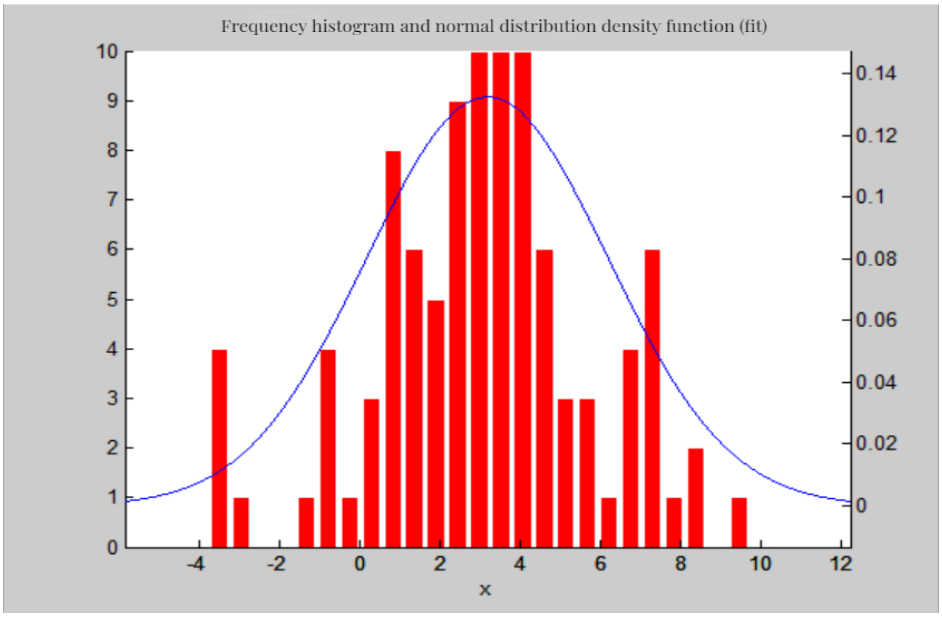

Figure 2

The data is estimated as it conforms to a normal distribution and thus interval estimation, since the overall data obeys a normal distribution:

$$P(\mathrm{c} \le \mathrm{G} \le \mathrm{d}) = \Phi(\mathrm{d}) - \Phi(\mathrm{c}) = 1 - \alpha$$

After a deformation of the inequality:

$$P_{\mu} = (\bar{x} - \mathrm{d}\sigma / \sqrt{n} \le \mu \le \bar{x} - c\sigma / \sqrt{n}) = 1 - \alpha$$

The length of this interval is $(c-d)\sigma/\sqrt{n}$,Since the standard normal distribution is single-peaked and symmetric, then under condition $\phi(d) - \phi(c) = 1 - \alpha; d - c$: is minimized when $d = -c = u_{1-\alpha/2}$, which gives:

$$\left[\bar{x} - u_{1-\alpha/2}\sigma / \sqrt{n}, \bar{x} + u_{1-\alpha/2}\sigma / \sqrt{n}\right]$$

This is a symmetric interval of radius $u_{1-\alpha/2}\,\sigma/\sqrt{n}$, often denoted as $\bar{x} \pm u_{1-\alpha/2}\,\sigma/\sqrt{n}$. For this problem the value of the confidence interval is specified to be 0.95, then $1 - \alpha = 0.95,\ \alpha = 0.05$.

Upon checking Table $\mu_{0.975} = 1.96$, the confidence intervals for each group can also be calculated as:

$$\bar{x} \pm u_{1-\alpha/2}\sigma / \sqrt{n} = \bar{x} \pm 1.96 \times \sigma / \sqrt{n}$$

Confidence intervals were calculated for each team:

Table 4

TEAMS	TEAM A	TEAM B	TEAM C	TEAM D	TEAM E	TEAM F	TEAM G
UPPER	3.77427	5.89409	6.58338	5.67641	4.12115	5.43169	4.99787
CONFIDENCE	738	2662	8642	3448	5046	1973	9309
LOWER	2.59299	4.86883	5.31095	4.34207	3.45022	4.34864	3.87064
CONFIDENCE	4777	4111	9064	5232	225	2045	628
TEAMS	Team H	Team I	Team J	Team K	Team L	Team M	Team N
UPPER	5.90641	5.04883	6.36214	6.57088	3.33785	5.41524	5.49345
CONFIDENCE	3017	4469	7885	183	7114	964	7441
LOWER	5.90641	5.04883	6.36214	6.57088	3.33785	5.41524	5.49345
CONFIDENCE	3017	4469	7885	183	7114	964	7441

We validate the accuracy of the confidence intervals for each team in the resulting CBA league through hypothesis testing:

For the one-sided test (1):

$$u = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Therefore, this test function is a function of μ: for $\mu \ \epsilon(-\infty, +\infty)$:

$$\begin{aligned} g(\mu) &= P_\mu\left(\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq u_{1-\alpha}\right) \\ &= P_\mu\left(\frac{\bar{x} - \mu + \mu - \mu_0}{\sigma / \sqrt{n}} \geq u_{1-\alpha}\right) \\ &= P_\mu\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \geq \frac{\mu - \mu_0}{\sigma / \sqrt{n}} \geq u_{1-\alpha}\right) \\ &= 1 - \Phi\left(\sqrt{n}(\mu_0 - \mu) / \sigma + u_{1-\alpha}\right) \end{aligned}$$

The potential function is an increasing function of μ. By the increasing function property, the test is a test with level α. For the one-sided test (2) $\mu \ \epsilon(-\infty, +\infty)$::

$$\begin{aligned} g(\mu) &= P_\mu\left(\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq u_\alpha\right) \\ &= P_\mu\left(\frac{\bar{x} - \mu + \mu - \mu_0}{\sigma / \sqrt{n}} \leq u_\alpha\right) \\ &= P_\mu\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \geq \frac{\mu - \mu_0}{\sigma / \sqrt{n}} \leq u_\alpha\right) \\ &= 1 - \Phi\left(\sqrt{n}(\mu_0 - \mu) / \sigma + u_\alpha\right) \end{aligned}$$

The potential function is a decreasing function of μ. By the increasing

function property, then the test is a test with level α. For the two-sided test (3) then its rejection domain should be

For $\mu \in (-\infty, +\infty)$:

$$g(\mu) = P_{\mu}\left(\frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \geq u_{1-\alpha/2}\right)$$
$$= 1 - P_{\mu}\left(\frac{\mu - \mu_0}{\sigma / \sqrt{n}} - u_{1-\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \leq \frac{\mu - \mu_0}{\sigma / \sqrt{n}} + u_{1-\alpha/2}\right)$$
$$= 1 - \Phi\left(\sqrt{n}(\mu_0 - \mu) / \sigma + u_{1-\alpha/2}\right) + \Phi\left(\sqrt{n}(\mu_0 - \mu) / \sigma - u_{1-\alpha/2}\right)$$

Further test the normal population hypothesis by setting $x_1, \ldots, x_m$ to be a sample from a normal population $N(\mu_1, \sigma_1^2)$ and $y_1, \ldots, x_n$ to be a sample from another normal population $N(\mu_2, \sigma_2^2)$. The two samples are independent of each other. Consider the following three types of testing problems:

$$I \quad H_0 : \mu_1 - \mu_2 \leq 0 \ \ VS \ \ \mu_1 - \mu_2 > 0 \quad (5)$$
$$II \quad H_0 : \mu_1 - \mu_2 \geq 0 \ \ VS \ \ \mu_1 - \mu_2 < 0 \quad (6)$$
$$III \quad H_0 : \mu_1 - \mu_2 = 0 \ \ VS \ \ \mu_1 - \mu_2 \neq 0 \quad (7)$$

For a two-sample u-test when σ_1, σ_2 is known, the distribution of point estimates of $\bar{x} - \bar{y}$ at this point in μ_1, μ_2 is perfectly known:

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} - \frac{\sigma_2^2}{n}\right)$$

From this, the u-test can be used and the test statistic is:

$$u = (\bar{x} - \bar{y}) / \sqrt{\frac{\sigma_1^2}{m} - \frac{\sigma_2^2}{n}}$$

At $\mu_1 - \mu_2$, $u \sim N(0,1)$. The rejection domain of the test depends on the specifics of the alternative hypothesis, and for test problem I shown in (5), the rejection domain and p-value of the test are respectively:

$$W_I = \{u \geq u_{1-\alpha}\}, p_I = 1 - \Phi(u_o)$$

where $u_0 = (\bar{x} - \bar{y}) / \sqrt{\frac{\sigma_1^2}{m} - \frac{\sigma_2^2}{n}}$ is the value of the test statistic computed from the sample. For the test problem II shown in (6), the rejection domain and p-value of the test are respectively

$$W_{II} = \{u \le u_{1-\alpha}\}, p_{II} = 1 - \Phi(u_o)$$

For test problem III of (7), the test and rejection domains and p-values are respectively

$$W_{III} = \{u \ge u_{1-\alpha}\}, p_{III} = 2(1 - \Phi(|u_o|))$$

In turn, we get a two-by-two comparison of teams against each other, which reconfirms the rankings as well as getting each team's probability of winning the title:

Table 5

	A	B	C	D	E	F	G	H	I	G	K	L	M	N
A	0.5	0.3483	0.33	0.3897	0.4522	0.3859	0.4168	0.3783	0.4247	0.3859	0.3859	0.2843	0.4286	0.2676
B	0.6517	0.5	0.4641	0.5239	0.6443	0.5359	0.5359	0.5675	0.5239	0.6026	0.5438	0.5398	0.4286	0.6064
C	0.67	0.5359	0.5	0.4129	0.5557	0.6664	0.5675	0.5987	0.5557	0.5331	0.5793	0.5753	0.4761	0.6331
D	0.6103	0.4761	0.5871	0.5	0.4641	0.5948	0.504	0.5359	0.5	0.5596	0.512	0.512	0.4129	0.5636
E	0.5478	0.3557	0.4443	0.5359	0.5	0.4013	0.4052	0.4443	0.3936	0.4602	0.4052	0.4052	0.2709	0.4681
F	0.6141	0.4641	0.3336	0.4052	0.5987	0.5	0.2483	0.5319	0.2483	0.5319	0.4919	0.4404	0.504	0.504
G	0.5832	0.4641	0.4325	0.496	0.5948	0.7517	0.5	0.3936	0.5636	0.3783	0.4602	0.5199	0.4761	0.4721
H	0.6217	0.4325	0.4013	0.4641	0.5557	0.4681	0.6064	0.5	0.3594	0.5239	0.3446	0.5714	0.516	0.516
I	0.5753	0.4761	0.4443	0.5	0.6064	0.7517	0.4364	0.6406	0.5	0.409	0.5714	0.3936	0.4443	0.4443
G	0.6141	0.3974	0.3669	0.4404	0.5398	0.4681	0.6217	0.4761	0.3936	0.5	0.3121	0.504	0.2912	0.5
K	0.6141	0.4562	0.4207	0.488	0.5948	0.5081	0.5398	0.6554	0.4286	0.6879	0.5	0.3783	0.5596	0.3707
L	0.7157	0.4602	0.4247	0.488	0.5948	0.5596	0.4801	0.4286	0.6064	0.496	0.6217	0.5	0.6879	0.492
M	0.5714	0.5714	0.5239	0.5871	0.7291	0.496	0.5239	0.484	0.5557	0.7088	0.4404	0.3121	0.5	0.2912
N	0.7324	0.3936	0.3669	0.4364	0.5319	0.496	0.5279	0.484	0.5557	0.5	0.6293	0.508	0.7088	0.5

Table 6

SPORTS TEAM	PROBABILITY OF WINNING THE CHAMPIONSHIP	RANKING
N	0.08155	1
L	0.078489	2
C	0.07728	3
B	0.075428	4
D	0.075293	5
M	0.074515	6
I	0.073494	7
K	0.073414	8
G	0.072554	9
H	0.070433	10
J	0.065519	11
F	0.065278	12
E	0.06184	13
A	0.054913	14

Through the model establishment and operation, we successfully obtained the probability of each team winning the championship as well as finally determined the top four teams are: N>L>C>B. Step1 chooses to take the four rating criteria of expectation, variance, extreme deviation and median for the level test of each team; the specific hierarchical structure chart is as follows:

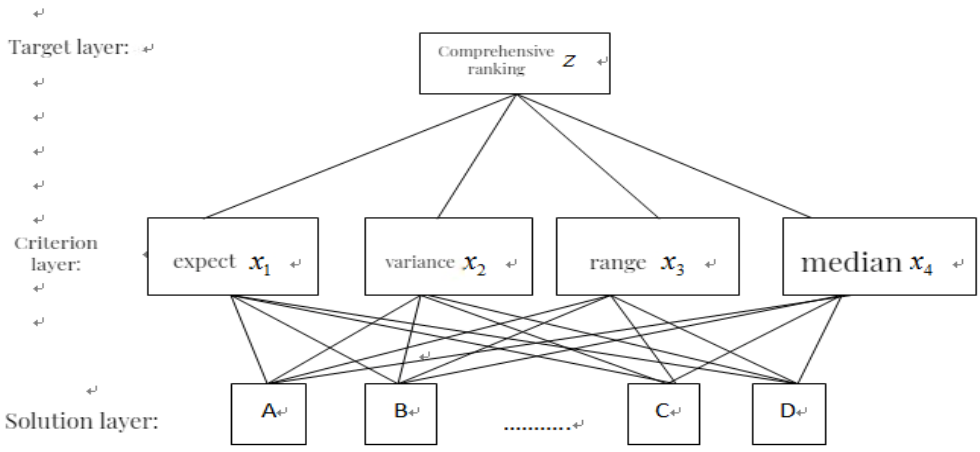

Figure 3

Construct a pairwise comparison matrix formed by comparing expectation, variance, extreme variance, and median two by two:

$$C = \begin{pmatrix} 1 & 3 & 6 & 4 \\ \frac{1}{3} & 1 & 3 & 2 \\ \frac{1}{6} & \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 2 & 1 \end{pmatrix}$$

Step 2 consistency test: firstly, calculate the maximum eigenvalue λ of pairwise comparison matrix C is 4.0310. secondly, through the formula:

$$CI = \frac{\lambda - n}{n-1}$$

The value of the consistency index CI was calculated to be 0.0103. by using Eq:

$$CR = \frac{CI}{RI}$$

The value of the consistency ratio CR was calculated as 0.0115 < 0.1 passing the consistency test. Where the random consistency index RI satisfies the following table:

Table 7

N↩	**3**↩	**4**↩	**5**↩	**6**↩	**7**↩	**8**↩	↩
RI↩	0.58↩	0.90↩	1.12↩	1.24↩	1.32↩	1.41↩	↩
N↩	9↩	10↩	11↩	12↩	13↩	14↩	↩
RI↩	1.45↩	1.49↩	1.52↩	1.54↩	1.56↩	1.58	↩

Step3 Determine the weights of each criterion: The eigenvector corresponding to the largest eigenvalue of the pairwise comparison matrix C is calculated:

$$m = (-0.8964, -0.3658, -0.1253, -0.2169)$$

Normalizing it yields:

$$m^* = (0.5587, 0.2280, 0.0781, 0.1352)$$

The resulting weights of each criterion layer on the target layer, i.e., the weights of expectation, variance, extreme variance, median on the target layer are 0.5587, 0.2280, 0.0781, 0.1352, respectively.

Step 4 construct the pairwise comparison matrix p_1 of each target layer A to N team to criterion layer expectation x_1 see appendix, with the above can be calculated its normalized eigenvector as

$$m_1^T = (0.0456, 0.0778, \cdots, 0.0629, 0.0927)$$

After calculating CI=0.0045, CR=0.0028<0.1, its passes the consistency test.

Step5 constructed the pairwise comparison matrix p_2 of the variance x_2 of each target layer A to the N team to the criterion layer is shown in the Appendix, and its normalized eigenvectors can be calculated in the above way as

$$m_2^T = (0.0838, 0.0699, \cdots, 0.0590, 0.0549)$$

After calculating CI=0.0119, CR=0.0075<0.1, it passes the consistency test.

Step 6 As the extreme difference is a criterion that plays a negative role in the sum score, so the larger the extreme difference indicates that its composite score should be smaller, the normalization of the extreme difference that is its inverse, constructed each target layer A to the N team on the criterion layer extreme difference x_3 of the pairwise comparison matrix p_3 is shown in the Appendix, with the above can be calculated as its normalized eigenvectors are

$$m_3^T = (0.0523, 0.0737, \cdots, 0.1040, 0.0665)$$

After calculating CI=0.0222, CR=0.0141<0.1, its passes the consistency test. Step 7 constructed the pairwise comparison matrix p_4 of the median x_4 of each target layer A to N team to criterion layer is shown in the Appendix, and its normalized eigenvectors can be calculated in the above way as

$$m_4^T = (0.0433, 0.0819, \cdots, 0.0639, 0.0933)$$

After calculating CI=-0.0025 and CR=-0.0016<0.1, it passes the consistency test.

Step 8 The normalized composite scores and composite rankings of the teams were calculated by Equation $z = m^* * (m_1^T, m_2^T, m_3^T, m_4^T)^T$ in the following table:

Table 8(a): Comprehensive Ranking

TEAMS	SCORES	RANKINGS
N	0.260969	1
L	0.241157	2
J	0.206898	3
H	0.197637	4
F	0.191606	5
B	0.171505	6

Table 8(b): Comprehensive Ranking

C	0.166992	7
D	0.166421	8
A	0.141879	9
E	0.133012	10
G	0.128519	11
K	0.115472	12
I	0.110349	13
M	0.10948	14

As can be seen from the table: we get the team level score and the comprehensive strength ranking through the hierarchical analysis of the four criteria, successfully comparing the various factors to determine the level of the team and the differences between the teams, showing a complete picture of the characteristics of each team pair and their own strength.

3.3 regression model

CBA is a place for fans to enjoy themselves and for players to give full play, and traditional Chinese cultural values have a profound influence on it, with the league emphasizing more emphasis on the collective interests of the team, on the mutual cooperation among players, and on setting a better example for the youth and competing better for the patriotism. With the help of the given historical scoring data and the above research results, improve the team to improve the main factors of winning the championship. Take the long and make up for the shortcomings, strengthen the short board, give the team in the CBA game the maximum probability of winning the championship of the rationalization of the proposal. First of all, the clustering analysis of the team, based on the data obtained for data processing, the specific results are shown in Fig:

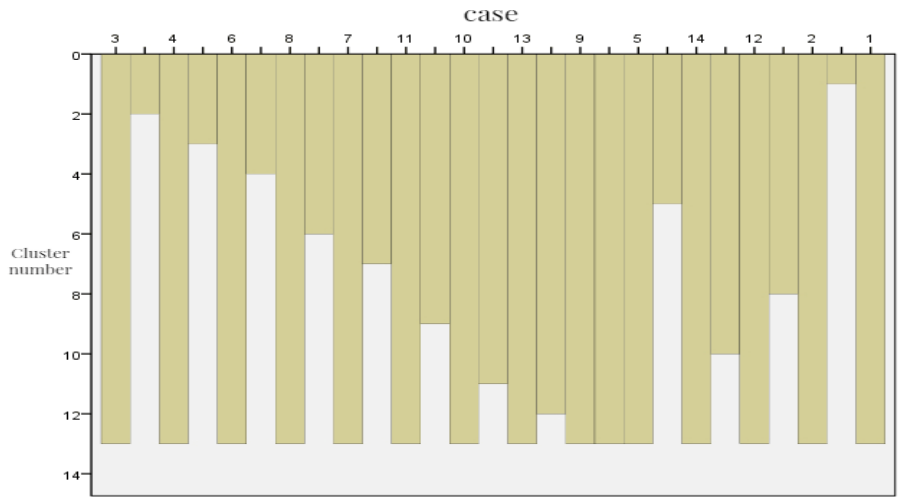

Figure 4

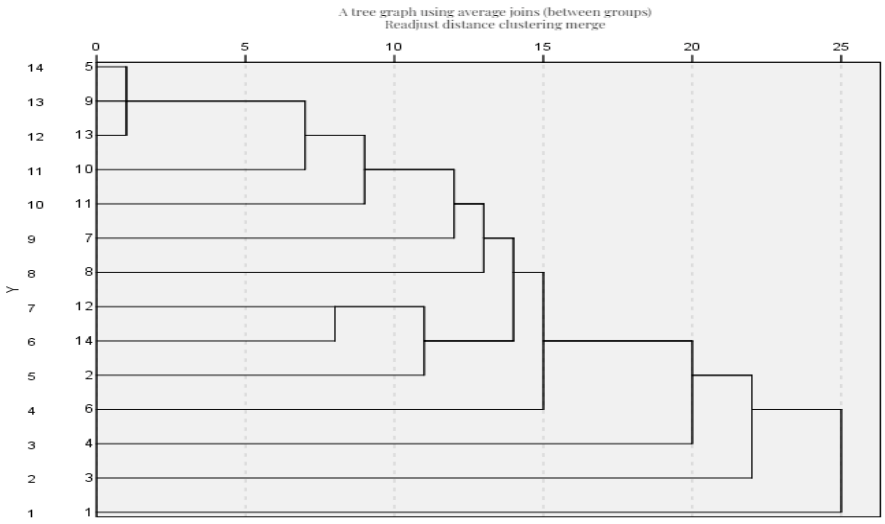

Figure 5

Through the cluster analysis image can be seen, if you select 20-25 as a measure, there are two kinds of clustering standards, through the observation of the data and the senior scholars to ask for advice, we believe that it is more reasonable to select the more back-end of the two types of stratification, selected team A and team L for the two-level analysis of the impact of the elements of the championship, the two teams are now selected for the team's data of the same time of the 20 wins with its tactical strategy, team mentality, key players and other three elements are analyzed, taking Team A as an example: Through the data found that the data basically meet the normal distribution, and then the data through the SPSS non-parametric test - normality test, the specific results are shown in the table:

Table 9: normality test

	KOLMOGOROV SMIRNOV (V)A			**SHAPIRO WILKE.**		
	statisticians	degrees of freedom	significance	statisticians	degrees of freedom	significance
TEAM A	.144	20	.000	.938	20	.000

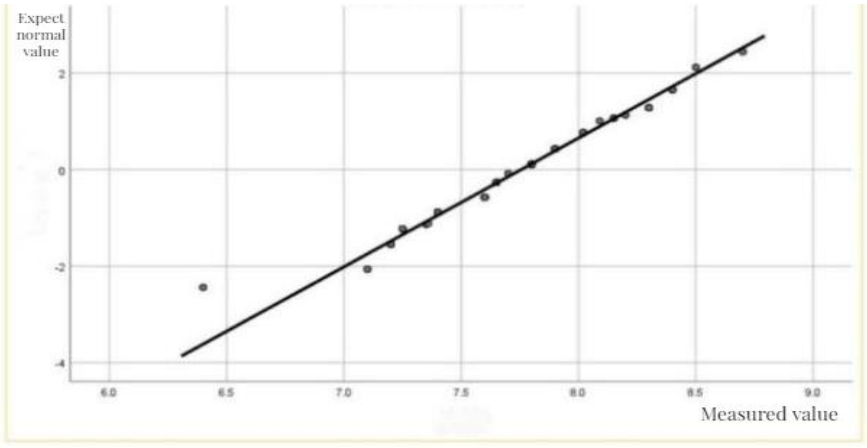

Figure 6

Through the normality test of SPSS (significance less than 0.05) and normal Q-Q plot, we find that the level of data of team A basically meets the normal distribution, and we can see that the degree of goodness of fit, but we can find that there is a break in the level of team A, and the higher level and lower level of the field deviate from the center of normality, which suggests that we should pay attention to the stability of the players in the training and the mentality of the players, so that the level of the players to be able to play stably. On the basis of the normal overall distribution, we want to explore the relationship between the team level, tactical strategy, team mentality, key players, especially the impact of the remaining elements on the team's game level, as well as whether there is an impact between the other elements of the situation, choose to carry out the independence test:

Table 10: chi-square (math.) test

	VALUE	DEGREES OF FREEDOM	ASYMPTOTIC SIGNIFICANCE (BILATERAL)
PEARSON'S CHI-SQUARE (MATH.)	876.000[a]	162	.000
LIKELIHOOD RATIO	225.860	162	.000
LINEAR CORRELATION	18.000	1	.000
NUMBER OF VALID CASES	20		

Table 11: Symmetry measurement

		VALUE	ASYMPTOTIC SIGNIFICANCE
NOMINAL TO NOMINAL	Phi	4.359	.000
	Clem V	1.000	.000
NUMBER OF ACTIVE CASES		20	

Thereupon, we conducted the consistency test between the dimensions, taking the team level dimension and tactical strategy as an example, and in the chi-square test table on the data, the value of the Pearson chi-square test was 876.000, and the significance value was 0.000 < 0.05, which rejected the pre-hostage hypothesis that there is no independence between the elements of the team level and several other dimensions.

According to the Clem V coefficient effect size rating table, dfmin is 1 in this question, and the Clem V coefficient in the symmetric measurement table is 1.000, which exceeds the standard of 0.50, then the effect size is a large effect size, and the credibility is high, and it can be assumed that the team level dimensions and tactical strategies are not independent from each other, and we continue to verify the correlation between the dimensions.

Table 12: Tests for between-subject's effects

DEPENDENT VARIABLE: LEVEL OF CIVICS								
SOURCE	**TYPE III SUM OF SQUARES**	**DEGREES OF FREEDOM**	**EQUAL SQUARE**	**F**	**SIGNIFICANCE**	**PARTIAL ETA SQUARE**	**NONPARAMETRIC CENTER**	**REAL POWER[B]**
MODIFIED MODEL	31.696[a]	2	10.565	163.413	.000	.710	490.240	1.000
INTERCEPT	17.407	1	17.407	269.223	.000	.574	269.223	1.000
TACTICAL STRATEGY	4.333	2	2.166	33.508	.102	.251	67.015	1.000
ERROR	12.931	17	.065					
TOTAL	1100.000	20						
CORRECTED TOTAL	44.627	20						
a. R-squared = .710 (adjusted R-squared = .706)								
b. Calculated using Alpha=.05								

As can be seen from the table of the test of inter-group effects, the F value of the indicator of the test of variance chi-square is 85.258, while the significance is 0.320, which exceeds the standard of 0.05, and the pre-hostage hypothesis is accepted. Conclusion: there is a variance chi-square relationship between the groups, which can be analyzed by ANOVA.

The F-value of the experimental variable "tactical strategy" is 4.333, with a significance of 0.102, which is greater than 0.05, and the hypothesis is rejected. Conclusion: There is a significant difference between the groups of tactical strategy, and tactical strategy has a significant effect on the dependent variable "team level". Calculation of Principle $\eta^2 = \frac{SS}{SS_a + SS_{error}}$ and the standardized table of effect sizes according to J. Cohen's Eta:.

Table 13: Standardized Table of Effects

ETA VALUE	EFFECT SIZE
0.01—0.06	Small effect size
0.06—0.14	Medium effect size
0.14 AND ABOVE	large effect size

The partial Eta square in the between-subjects effect test table has the statistical efficacy of an analysis of variance (ANOVA), and the value of the partial Eta square for tactical strategy is 0.251, which exceeds the criterion of 0.14, and should be considered a large effect size.

The effect size indicates that the statistical conclusion (i.e., the rejection of the original hypothesis) has a high degree of confidence, proving that the conclusion has a very high degree of confidence (the rejection of the original hypothesis), and that "tactical strategy" has a significant effect on "team level".

In order to judge the size of the correlation and the specific correlation, we carry out the correlation test, first of all, the correlation analysis of the team level and tactical strategy, team mentality, key players, aiming to get the influence relationship and the degree of influence between the elements, the specific influence relationship and correlation coefficients are shown in the table:

Table 14: correlation analysis

		TEAM A	TACTICAL STRATEGY	TEAM MENTALITY	KEY PLAYERS
TEAM A	Pearson correlation	1	.837	.382	.678
	Sig. (two-tailed)		.003	.275	.031
TACTICAL STRATEGY	Pearson correlation	.837	1	.470	.522
	Sig. (two-tailed)	.003		.170	.122
TEAM MENTALITY	Pearson correlation	.382	.470	1	.740
	Sig. (two-tailed)	.275	.170		.015
KEY PLAYERS	Pearson correlation	.678	.522	.740	1
	Sig. (two-tailed)	.031	.122	.015	

According to Cohen's point of view, to calculate the effect size and statistical test power of the cumulative difference correlation coefficient, the reference basis of the standard is: the effect size of the cumulative difference correlation is the coefficient r.

When the cumulative difference correlation coefficient r is significant, r=0.10 is a small effect size, r=0.30 is a medium effect size, and r=0.50 is a large effect size. In order to better get the influence coefficient and intuitive expression, regression analysis is carried out, and the specific data and analysis results are shown in the table:

From the statistical results, it can be seen that the compound correlation coefficient R is 0.922, the corresponding sample coefficient of determination R^2 is 0.850, and its adjusted coefficient of determination $\overline{R^2}$ is 0.774, and from the data analysis, it can be seen that the model regression explains a relatively high percentage.

Table 15: ANOVA[a]

MODEL		SQUARE SUM	DEGREES OF FREEDOM	EQUAL SQUARE	F	SIGNIFICANCE
1	Regression (math.)	6.953	3	2.318	11.293	.007[b]
	Residuals	1.231	6	.205		
	Total	8.184	9			
a. Dependent variable: team level						
b. Predictor variables: (constant), tactical strategy, team mentality, key players						

The F-value corresponding to the regression equation is 11.293, while the significance is 0.007, which is lower than the pairwise standardized value of 0.05, and the pre hypothesis is rejected. By analyzing this, the regression equation we established is valid.

Table 16: Ratio[a]

		UNSTANDARDIZED COEFFICIENT		STANDARDIZED COEFFICIENT		
MODEL		B	STANDARD ERROR	BEAT	T	SIGNIFICANCE
1	(Constant)	.858	.437		1.963	.000
	Tactical Strategy	2.231	.855	.709	2.461	.013
	Team mentality	1.459	1.480	.395	1.625	.006
	Key Players	2.173	1.301	.600	1.856	.029
a. Dependent variable: team level						

Through the regression analysis table, we can see that its F-test and R-test significance is good and can indicate the change of the function. And through the coefficient matrix we can get the correlation coefficient of each standardization, and the significance except for the family social education coefficient significance is slightly lower, the significance of other factors is better, but the significance of all the related variables is lower than 0.05, the expression is more reliable. If we set the unknown quantities team level Y, tactical strategy X_1, team mentality X_2, key players X_3, the regression expression can be obtained from the correlation coefficient:

$$Y = 0.709X_1 + 0.395X_2 + 0.6X_3 + 0.858$$

Throughout these historical scoring data, from the current situation and development trend of basketball, the factors affecting the team's probability of winning the championship in the CBA league can be the psychological state of the athletes, the optimal regulation of the game state, the environment of the

game venue, the audience factors, refereeing factors and so on. (JIA et al., 2022; W. LIU, 2021) Take the psychological state of athletes as an example, generally in the home game, athletes are more vigorous, in the home crowd's cheering and cheering can be fully or even super level play, can be fully committed to the game. This psychological state of the athlete controls the whole game situation, and the advantage will be revealed in the game. On the contrary, in away games, athletes are prone to unstable psychological state, tactical cooperation is not enough, which ultimately leads to the loss of the game (Liu, 2020). At the same time, the layout of tactical strategy is the key element to win the whole game. Technology is the basis of tactics, and tactics are the expression of technology. To realize the purpose of tactics not only to have a certain tactical awareness of the domination (Y. Li, 2021), but also must require players to master the appropriate technical ability to use, there is no tactics without technology. The improvement of tactics in turn can stimulate the development of technology. On the other hand, the improvement of technical level is also bound to promote the development of tactics, a certain level of technology can only be adapted to the corresponding tactical requirements. The two are complementary, otherwise it may be unrealistic (C. LI et al., 2019; Ning, 2013).

4. Model Evaluation and Extension

4.1 Model Benefits

1. Avoiding the influence of subjective factors on the actual estimation results, it is more objective. 2. The model has a wide range, not only applicable to the CBA league, but also applicable to many sports competitions, with a Great reference significance. 3. The mathematical and statistical model does not have strict requirements for the competition (basketball) program. 4. It is flexible and highly adjustable. 5. We can estimate the next trend objectively by referring to its historical data. 6. The model algorithm is not complicated, convenient and practical.

4.2 Model Drawbacks

Estimates of volatile and unstable teams are subject to large errors, and there is no guarantee that the final result will be convincing to all.

5. Other current data analysis issues and responses

5.1 Existing issues

Idea is the guide of behavior; a certain development behavior is led by a certain development concept. The use of data analysis in major international sports events to help the game to get the victory has been commonplace (P. LIU, 2021), four years and three times won the NBA championship Warriors are

known as: data analysis to create a great team; 2019 Chinese women's volleyball World Cup volleyball team won the title behind the use of a large number of data collection and analysis techniques. Reached the point of pure fire, compared to the CBA due to the lack of data collection and data analysis application of the awareness of our current data analysis has not been elevated to the due strategic height, history proves: do not follow the trend of the times will inevitably be eliminated by the times (Hanping, Chun, & Yong, 2018). Our CBA league executives, club owners and professionals should firstly recognize the importance of data analysis from the level of conscious thinking, and elevate the awareness of data analysis to the strategic level, so as to help the development of CBA league with data integration technology (Chang et al., 2018).

5.2 Countermeasures

1. Expand the training of data analysis applications: "One year's plan is like a tree valley; ten years' plan is like a tree; life-long plan is like a tree", the cultivation of talents is a long-term process, but the benefits from the long-term interests are incalculable. The quality of data analysis talents will directly determine the efficiency of CBA data analysis application, therefore (Manzano-Moreno et al., 2021), CBA league should firstly organize regular training courses on basketball data analysis from the inside, and invite experts in basketball data at home and abroad to explain the advanced concepts and application methods of data analysis. Externally, the CBA league should encourage universities and colleges to open sports big data analysis application courses, and cultivate talents who understand both basketball knowledge and statistics to give back to the CBA league, so that colleges and universities can gradually become a fertile ground for cultivating CBA data application talents (Bonne & Wong, 2012).

2. Increase investment in data analysis hardware construction: Science and technology is the first productive force, the NBA has been skilled in the use of high-tech products to help the league data analysis, CBA as the world's third largest basketball league should take the initiative to accept high-tech products, the use of high-tech products to help the league data analysis applications. For example, the introduction of player tracking equipment he can record the player's offensive area, running distance (Gai et al., 2018; Langevin et al., 2017). Intelligent video data system it can recognize the court player's finishing style, tactical play. In training, wearable devices can be used to assist training, monitor heart rate to improve the training effect to prevent injuries. In short, the use of high-tech products is the CBA league data analysis application to the high-end development is not a choice.

3. Enriching the application field of data analysis: CBA team data analysis application mainly focuses on the grasp of the opponent's foreign aid

offensive characteristics on the field of play, tactical play formulation, etc., and should gradually expand the application of data analysis, used to enhance the team's daily training efficiency, evaluation of the efficiency of the players' lineup combinations, optimization of the fan's data experience, the development of the draft player scouting report and the team's scientific attraction and other aspects. For example, the development of the draft player scouting report, with the current strength of college players to enhance the annual CBA draft has become a key channel for the team to strengthen their own strength, through the scouting report team can understand in detail the comprehensive strength of the draft players, which gives the players the opportunity to show themselves and help the team to select the appropriate construction and development of the players.

Reference

Bonne, N. J., & Wong, D. T. (2012). Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome Medicine, 4*, 1-12.

CAO, D., WANG, G., & TU, C. (2015). Q&A on production technology of flat mushroom. *Beijing:Chemical Industry Press*, 107-109.

Chang, Y.-A., Weng, S.-L., Yang, S.-F., Chou, C.-H., Huang, W.-C., Tu, S.-J., . . . Huang, H.-D. (2018). A three–microRNA signature as a potential biomarker for the early detection of oral cancer. *International Journal of Molecular Sciences, 19*(3), 758.

Duan, S. (2020). Status and trend of precision poverty alleviation in edible mushroom industry in the new era. *China Edible Mushroom, 39*(1), 192-195.

Gai, C., Camussi, F., Broccoletti, R., Gambino, A., Cabras, M., Molinaro, L., . . . Arduino, P. G. (2018). Salivary extracellular vesicle-associated miRNAs as potential biomarkers in oral squamous cell carcinoma. *BMC cancer, 18*, 1-11.

Gao, C., Dynasty, J., & Li, H. (2017). High quality production technology of flat mushroom. *Beijing:China Science and Technology Press*, 101-103.

Gao, L., Dai, S., Xu, X., Zhou, J., & Li, M. (2021). Research on coupled control method of greenhouse temperature and humidity. *Agricultural Mechanization Research, 43*(12), 24-30.

Hanping, M., Chun, J., & Yong, C. (2018). Analysis and Prospect of Research Progress on Greenhouse Environment Control Methods [J]. *Journal of Agricultural Machinery*.

Huo, G. (2021). Implications of NBA data analysis for CBA application in the era of big data. *Chengdu Institute of Physical Education*.

JIA, Y., SU, Y., ZHANG, R., LI, P., WANG, F., & LU, M. (2022). BP neural network optimization model simulation of reference crop evapotranspiration under restricted meteorological data conditions:A case study in Beijing-Tianjin-Hebei region. *China Agricultural*

Meteorology, 43(01), 1-16.

JING, L., & FANG, Q. (2019). Neural network-based CFD temperature uniformity prediction model for mushroom room. *Chinese Journal of Agricultural Mechanical Chemistry, 40*(06), 71-75.

Ju, R. (2008). High-yield cultivation technology of flat mushroom. *Beijing:Jindun Publishing House*, 115-116.

Langevin, S., Kuhnell, D., Parry, T., Biesiada, J., Huang, S., Wise-Draper, T., . . . Kasper, S. (2017). Comprehensive microRNA-sequencing of exosomes derived from head and neck carcinoma cells in vitro reveals common secretion profiles and potential utility as salivary biomarkers. *Oncotarget, 8*(47), 82459.

LI, C., TAN, Q., BIAN, Y., XIE, V., LIU, Z., & LI, Y. (2019). Current status and outlook of edible mushroom factoryization in China. *Mycological Research, 17*(01), 1-10+12.

Li, J., & Ma, T. (2017). Detecting the development trend of core technology using local outlier factor algorithm--Taking China's wind energy patent data as an example. *Journal of Intelligence, 36*(3), 119-124.

Li, Y. (2021). The sustainable development of edible fungi industry in China in the post epidemic era. *Mycological Research, 19*(01), 1-5.

LIU, P. (2021). Research on environmental monitoring and greenhouse wind control system for cucumber growth in facilities. *Shandong Agricultural University*.

LIU, W. (2021). Points of spring mushroom production management of flat mushrooms planted in autumn with fermentation material. *Edible Mushrooms, 43*(05), 44-45.

Liu, X. (2020). Research on international competitiveness of Chinese edible mushroom industry under the background of "Belt and Road". *Chinese edible fungus, 39*(09), 169-171+175.

Manzano-Moreno, F. J., Costela-Ruiz, V. J., García-Recio, E., Olmedo-Gaya, M. V., Ruiz, C., & Reyes-Botella, C. (2021). Role of salivary MicroRNA and cytokines in the diagnosis and prognosis of oral squamous cell carcinoma. *International Journal of Molecular Sciences, 22*(22), 12215.

Ning, L. (2013). Forecast model of minimum temperature inside greenhouse based on principal component regression. *Chinese Journal of Agrometeorology, 34*(03), 306.

SUN, M., CHEN, J., & FANG, M. (2001). The Development Trend of World Basketball Athletics in the 21st Century--An Introduction to the Present Situation of Basketball and Countermeasures in China. *Sports Science*(1), 44-46.

Tong, Z. (2002). Humanistic Sports: Culture of Sports Interpretation. *China Customs Press*.

Zhang, L., & Mao, Z. (1995). The relationship between physical exercise and mental health (a review). *Journal of Guangzhou Sports Institute*(4), 42-47.