ISSN : 0974 - 7435 Volume 10 Issue 10

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(10), 2014 [4768-4777]

Based on factor analysis of SAS the world cup soccer team comprehensive strength evaluation research

Yong Li, Shanshan Li
Shandong Sport University, Jinan 250102, Shandong, (CHINA)

ABSTRACT

2010 South Africa World Cup has ended, the field of the performance of the 32 teams is not the same. In order to more objectively evaluate the comprehensive competition of the comprehensive competitiveness of 32 teams and technology level, teams of comprehensive competitiveness evaluation model is established in this paper. Through the relations of the related events and technical indicators data statistical analysis, factor analysis model is established, and by using SAS software programming indicates that the four main factors affect the team the outcome, the number of foot ball, ball rate, passing the success rate and steals the success rate. To quantify the engagement team in the history of the processing, combined with the feature of the data, and the 2010 World Cup game, establish a fuzzy comprehensive evaluation model, and find out seven good luck team, respectively: Mexico, Uruguay, Argentina, Paraguay, Switzerland, the United States, Serbia. Finally to 32 teams on the field of integrated performance, give the different weights of the four major factors, calculated 32 teams of different values, and thus it is concluded that the comprehensive strength of the 32 teams, one of the top three, respectively is Italy, Brazil and Argentina.

KEYWORDS

Comprehensive strength; Evaluation model; Rank; Factor analysis model; Fuzzy comprehensive evaluation.

© Trade Science Inc.

INTRODUCTION

With matador lifted the trophy, the World Cup in South Africa, for the 32 teams have different harvest, so to speak. Spain is newly crowned champions, the Netherlands continued the uncrowned king, let people disappointed in Argentina and Uruguay at last moment. All this success or failure, in addition to the game for some events there is a certain amount of luck, more is determined by the performance of the players on the pitch. Hou Weidong (2005) the fourth FIFA women's World Cup football match on the teams tactics has carried on the comprehensive evaluation using ability, to explore a more reasonable and effective method of quantitative evaluation team tactics ability. Chang-quan wang, lang, Li Sun South Asia (2007) of the Chinese women's team and high level women's offensive capability has carried on the comparative analysis, some evaluation methods are obtained. Hou Huisheng (2008) on the 18th World Cup finals last 32, 64 games, the tactics for the 13 indicators of statistics, reveal the developing trend of the world football tactics at the same time, explore a more reasonable and effective quantitative method to evaluate the quality of the team tactics.

By convention, FIFA and some official institutions through its website after the game and some portal outward announced the 2010 FIFA World Cup in South Africa of all relationships and the technical statistics, such as the goals, the number of shooting, passing, steals, etc. This article, based on the related data of each team and statistical analysis of data, through the establishment of the comprehensive competitiveness of participating team evaluation model to evaluate the comprehensive strength of the teams.

FACTOR ANALYSIS MODEL

Model establishment and solution

By searching and analyzing relative data, corresponding factors are as following aspects: goal, shoot, shoot on target, corner kick, pass, steal, stealing success rate, ball handling rate, lose, offside, foul, yellow card, and red card.

In order to find out most important influence factors, the paper will establish factor analysis model that is selecting least common factors from intricate influence factors to simplify variables' number and structure, so that using relative coefficient matrix to find out fewer factors that can best decide results.

To establish factor analysis model, firstly it should find out relative factors' relative data. We search thirteen indicators relative data in 2010 world cup official website, considering that teams' participated competition sessions are different, and causes pass, number of shooting, corner kick and others nonidentity, so we make following processing:

For goal, number of shooting, number of corner kicking, offside, foul, yellow card and red card, all calculate according to average number per competition

For ball handling, number of passing, handle them as corresponding ball handling rate, passing rate

Steal is divided into stealing average number and stealing success rate two parts

Pass is divided into number of passing and passing success rate two parts

According to above rules, we get correlation data after processing as following TABLE 1.

TABLE 1 : Each correlation factor data

Team	Goal	Shooting	Shoot on target	Corner Kick	Pass	Passing success rate	Steal	Stealing success rate	Ball handling rate	Lose	Foul	Offside	Yellow card	Red card
Spain	1.1429	18	5	8	615.8571	82.40%	18	82.50%	66.30%	0.2857	11.7143	1.7143	1.1429	0
Argentina	2	18.6	7.8	6.6	522.8	82.90%	22.8	76.30%	62.40%	1.2	13	2.4	1.4	0
Brazil	1.8	17.8	6.2	6.8	479	81.10%	21.2	70.80%	59.70%	0.8	15.4	1.4	1.4	0.4
Germany	2.2857	15.1429	5.2857	6.2857	471.7143	78.80%	22.4286	75.20%	51.70%	0.7143	10.7143	3.2857	1.5714	0.1429

Mexico	1	13.5	3.5	4.25	462	78.20%	25.75	78.60%	57.80%	1.25	20.75	3.5	2.25	0

In the following, start to establish model: Set $x_i(i=1,2,...,14)14$ individual variables that are expressed as:

$$Xi = a_{i1}F1 + a_{i2}F2 + ... + a_{im}F_m + \varepsilon_i (m \le 14)$$

That $X = AF + \varepsilon, F_1, F_2, ...F_m$ is called common factor, is unobservable variables, $A=(a_{ij})_{14\times m}$ is called factor loading matrix, a_{ij} represents the i variable in the j factor loading, ε_i is special factor, which is the part that cannot be contained by former m pieces of common factors, and meet a $Cov(F,\varepsilon)=0, F, \varepsilon$ are uncorrelated. (Among them: X_i: relative factors that affect team winning or losing, $i=1,2,3,4.....14$, F_m: common factor).

① $x=(x_1,x_2,...,x_{14})'$ is observable random vector, and mean vector $E(x)=0$, covariance matrix $\text{cov}(F)=\sum,$ and covariance matrix $\sum$ and relative matrix R are equal;

② $F=(F_1,F_2,...,F_m)'$, $m<p$, is unobservable vector, its mean vector $E(F)=0$, covariance matrix $\text{cov}(F)=1$, that vector F each component is mutual independent from each other;

③ $\varepsilon=(\varepsilon_1,\varepsilon_2,...,\varepsilon_{14})'$ and F are mutual independent from each other, and $E(\varepsilon)=0$, ε covariance matrix $\sum_\varepsilon$ is diagonal matrix, that is:

$$\text{cov}(\varepsilon)=\sum_\varepsilon=\begin{bmatrix}\sigma^2{}_{11}.....0\\ .....\sigma^2{}_{22}\\ 0.....\sigma^2{}_{14\,14}\end{bmatrix}$$

It shows ε each component is also independent from each other, then model:

$$\begin{cases}x_1=a_{11}F_1+a_{22}F_2+...+a_{1m}F_m+\varepsilon_1\\ x_2=a_{21}F_1+a_{22}F_2+...+a_{2m}F_m+\varepsilon_2\\ .....\\ x_{13}=a_{13\,1}F_1+a_{13\,2}F_2+...+a_{13\,m}F_m+\varepsilon_p\end{cases}$$

It is called factor model, its matrix form is:

$$x=AF+\varepsilon$$

Among them: $x_i=(x_1,x_2,...,x_p)'$, $F=(F_1,F_2,...,F_m)'$, $\varepsilon=(\varepsilon_1,\varepsilon_2,...,\varepsilon_p)'$

$$A=\begin{bmatrix} a_{11}, a_{12}.....a_{1m} \\ a_{21}, a_{22}.....a_{2m} \\ ... \\ a_{13\,1}, a_{13\,2}, ...a_{13\,m} \end{bmatrix}$$

Relative factors symbol description is as TABLE 2 show.

TABLE 2 : Relative factor symbol description

Symbol	Symbol description
JQ	Number of goals
SM	Number of shooting
SZ	Number of shooting on target
JIAOQ	Number of corner kicking
CQ	Number of passing
CQC	Passing success rate
QD	Number of stealing
CQC	Stealing success rate
KQL	Ball handling rate
SQ	Goals against
FG	Number of foul
YW	Number of offside
HUANGP	Number of yellow card
HONGP	Number of red card

Utilize SAS software to program; it solves results as TABLE 3 show.

TABLE 3 : SAS solution result table

JQ	SM	SZ	JIAOQ	CQ	CQC	QD
0.9176695	0.9518433	0.8063619	0.9662925	0.9774484	0.8949144	0.9819358
QDC	KQL	SQ	FG	YW	HUANGP	HONGP
0.9409011	0.9826882	0.8782212	0.9029488	0.9639159	0.8423043	0.8833767

Result analysis

Software program result requires four larger factors to be key factors, therefore rank each kind of factors from high to low, take top four factors as key factors that are respectively KQL、QD、CQ、JIAOQ;but according to factor parameters, it is clear that YW is also higher, and in practical competition, offside impact are also larger, therefore it also regard YW as an important factor to measure team strength, but it cannot be regarded as key factor.

FUZZY COMPREHENSIVE EVALUATION MODELS

Model analysis

In order to define in competition winning team whether has a matter of luck, we according to historical fighting records, fighting two teams' strength differences to judge, and to select the competition abnormal teams as we think that might win by luck. However a team current strength is not fully up to historical records. It also related to whether recently the team introduces star player, famous coach and short term assembled training and other ways to improve team strength that let the team to win rather by luck. It further is eliminated from luck questioning. In this way, we according to fighting teams' previous fighting records and the world cup fighting teams' standings, preliminary judge abnormal teams, and then according to problem one solved key factors, establish team strength fuzzy comprehensive evaluation model to make final judgment whether the team has a matter of luck or not.

Model establishment and solution

By above problem analysis, we firstly should make quantitative handling with teams' historical fighting records. We handling the data according to its provided data features like following, now we only select three groups of data as TABLE 4.

TABLE 4 : Two teams' fighting historical record

Country	Fighting sessions	Win	Draw	Lose	Gain the ball	Lose the ball	Goal difference	Winning rate
Netherlands VS Japan	1	1	0	0	3	0	3	100.00%
Netherlands VS Denmark	27	11	10	6	55	36	19	64.71%
Netherlands VS Cameroon	2	1	1	0	1	0	1	100.00%

By above TABLE 4 handling result, we need to select seemingly abnormal team combination, how to select abnormal teams, we according to previous customs that have following some principles:

Two fighting teams historical fighting arrives at more than 10 games and strength difference is bigger, one party wins that belongs to abnormality.

If two teams have no fighting records or in fighting make draw that the two belong to approximate strength teams, winning or losing result in this time don't be references.

Two parties fighting times are lower than two, and both the two have no big scores when win, the winning result at this time don't be considered.

Two parties fighting is above five games and less than 10 games, while standings fluctuate around 50% that strengths are approximate, it would not be taken into consideration.

Meanwhile, combine with above principles; we put forward winning rate formula:

$$\text{Winning} = \frac{wins}{\text{total number}}$$

Combine with World Cup in 2010 the searched each team competition record, compare with historical records, according to above principle, preliminary gets seemingly abnormal teams as following TABLE 5.

TABLE 5 : Primary selection abnormal teams

Country	Netherlands VS Slovakia
Mexico VS France	Serbia VS Germany
South Africa VS France	Ghana VS Australia

Uruguay VS Mexico	Paraguay VS Italy
Argentina VS Nigeria	Portugal VS Brazil
America VS England	Switzerland VS Spain

Remark: Fighting teams, former one is abnormal team, totally has eleven teams

Now according to above obtained data, we solve team winning or losing influence factors: corner kick, ball handling rate, steal, and pass, establish team fuzzy comprehensive evaluation model:

According to multiple factors, make comprehensive evaluation on fighting teams two parties performed comprehensive strength at this time and compare. Teams' fuzzy evaluation model generally includes following three aspects: record teams' winning or losing main influence factors are four that as:

$U = \{u_1, u_2, u_3, u_4\}$: u_1: corner kick; u_2 : ball handling rate; u_3: pass; u_4 : Steal

And due to each factor position is different; its function is also different. Generally consider endowing weights to measure, we take problem one solved corner kick, ball handling rate, pass and steal weights to be recorded as:

$$\mathrm{A}=\{\mathrm{a}_1, a_2, a_3, a_4\} = \{0.96629254, 0.98193582, 0.98268825, 0.97744849\}$$

While set existing single factor evaluation set $r_{ij} = \{u_{i1}, u_{i2}, u_{i3}, u_{i4}\}$ so that construct comprehensive evaluation matrix:

$$R = \begin{bmatrix} r_{11,} r_{12}, r_{13,} r_{14} \\ r_{21,} r_{22,} r_{23,} r_{24} \\ \cdots\cdots\cdots\cdots \\ r_{221}, r_{222}, r_{233}, r_{234} \end{bmatrix}$$

Among them r_{ij} represents the i team the j each factor weight percentage. Such as following TABLE 6's Mexico vs. France.

TABLE 6 : Mexico vs France

Country	Corner kick	Ball handling rate	Stealing success rate	Passing success rate
Mexico VS France	1/7	48/52	76/72.2	76.4/80.4
South Africa VS France	5/3	56/44	68.8/71.4	84.9/83.2
Uruguay VS Mexico	7/6	33/67	71.4/83.3	72.7/84.8
Argentina VS Nigeria	10/4	65/35	71/86.4	87.1/75.1
America VS England	4/8	44/57	65.4/69.2	66/76.9
Serbia VS Germany	1/7	42/58	63.2/93.8	83.1/84.4
Ghana VS Australia	6/1	54/46	92.9/76.9	84.1/75.3
Paraguay VS Italy	4/8	43/57	70.8/83.9	63.9/70.3

Portugal VS Brazil	4/7	31.5/68.5	62.5/50.0	77.2/88.9
Switzerland VS Spain	3/12	26/74	68/87.5	70.1/89.5
Netherlands VS Slovakia	5/2	52/48	72.2/92.3	80.6/79.3

Among them, $r_{11=\frac{1}{1+7}=\frac{1}{8}}$ other also successive calculate. Its algorithm is: $b_j = \sum_{i=1}^{4} a_i * r_{ij} \left(j = 1,2,3...,22 \right)$

The model gives balance consideration to all factors according to weights sizes, so that solves the team with maximum total, which is also means bigger results ones don't exist luck, on the contrary is the result of luck. According to algorithm, solved each team numerical values can refer to following TABLE 7.

TABLE 7 : Results whether is matter of luck or not

Country	Corner kick	Ball handling rate	Stealing success rate	Passing success rate	Former team score	Later team score	Luck or not
Mexico VS France	1/7	48/52	76/72.2	76.4/80.4			Yes
South Africa VS France	5/3	56/44	68.8/71.4	84.9/83.2	2.129564	1.78831	No
Uruguay VS Mexico	7/6	33/67	71.4/83.3	72.7/84.8			Yes
Argentina VS Nigeria	10/4	65/35	71/86.4	87.1/75.1			Yes
America VS England	4/8	44/57	65.4/69.2	66/76.9			Yes
Serbia VS Germany	1/7	42/58	63.2/93.8	83.1/84.4	1.4136	2.494379	Yes
Ghana VS Australia	6/1	54/46	92.9/76.9	84.1/75.3			No
Paraguay VS Italy	4/8	43/57	70.8/83.9	63.9/70.3			Yes
Portugal VS Brazil	4/7	31.5/68.5	62.5/50.0	77.2/88.9	1.66056	2.29147	Yes
Switzerland VS Spain	3/12	26/74	68/87.5	70.1/89.5			Yes
Netherlands VS Slovakia	5/2	52/48	72.2/92.3	80.6/79.3	2.12463	1.79346	No

From calculated data TABLE 7, we see that Mexico, Uruguay, Argentina, America, Ghana, Paraguay have no figures that's because their four items results are smaller than later ones, it can see whether they are lucky or not without calculating; by calculating, it gets teams has luck factors in competition process those are: Mexico, Uruguay, Argentina, America, Serbia, Paraguay and Switzerland so on seven teams.

COMPREHENSIVE STRENGTH EVALUATION MODELS

To rank every team according to thirty-two teams performance in the field, it should have an objective function that reflects team comprehensive strength, utilize key factors and important factors that impact on team winning or losing to define objective function value, when objective function value gets bigger, it is supposed that its comprehensive strength is the strongest, and the rank is more forward.

In order to correctly define objective function that reflects team comprehensive strength, we respectively put forward two kinds of model to consider it:

Model 1

According to every key factor concrete parameter, team integral, and single item key factor ranking to establish model, and get objective function that reflects team comprehensive strength as:

$$W_k = \sum_{i=1}^{n} A_{ki}(33 - j)$$

33-j is single rank obtained score; relative symbol description is as TABLE 8 show.

TABLE 8 : Relative symbols description

Symbol	Symbol description
W_k	According to Model one team comprehensive strength
i	Team strength influence factors (i from one to six respectively represent corner kick, pass, steal, ball handling rate, integral, offside)
j	Each factor rank in thirty-two teams
n	Key factors or important factors number
A_k	Represent thirty-two countries (k from 1 to 32 respectively represents Spain, Argentina, Brazil)

According to searched data, utilize $W_k = \sum_{i=1}^{n} A_{ki}(33 - j)$ to calculate and get results as TABLE 9 show.

TABLE 9 : Model one obtained comprehensive strength rank

Country	Rank	Country	Rank
Spain	1	Portugal	17
Argentina	2	Australia	18
Brazil	3	Slovakia	19
Germany	4	South Korea	20
Mexico	5	France	21
Italy	6	Algeria	22
England	7	Denmark	23
Cameroon	8	Slovenia	24
Netherlands	9	Uruguay	25
Ghana	10	Switzerland	26
South Africa	11	Greece	27
Ivory Coast	12	Japan	28
Paraguay	13	North Korea	29
Chile	14	Nigeria	30
Serbia	15	Honduras	31
America	16	New Zealand	32

Model 2

According to every key factor weights and combine with team every key factor rank, it establishes model, gets objective function that reflects team strength:

$$M_k = \sum_{c=1}^{n} A_{kc}(33 - j)$$

33-j is single rank obtained scores

Relative symbols description is as following TABLE 10 shows.

TABLE 10 : Relative symbols description

Symbol	Symbol description
M_k	Team strength according to Model two
c	Each key factor weight (c from 1to 6 respectively represent corner kick, pass, steal, ball handling rate, integral, offside)
j	Each factor rank in these thirty-two teams
A_k	Represents 32 countries (k from 1 to32 respectively represents Spain, Argentina, Brazil)
n	Key factors or important factors number

According to TABLE 3 data and utilize formula:

$$M_k = \sum_{c=1}^{n} A_{kc}(33 - j)$$

It gets rank results as following TABLE 11 shows.

TABLE 11 : Team strength rank according to Model two

Country	Rank	Country	Rank
Italy	1	Greece	17
Brazil	2	Uruguay	18
Argentina	3	Slovakia	19
Spain	4	Denmark	20
England	5	France	21
Germany	6	Algeria	22
Mexico	7	South Africa	23
Ghana	8	Australia	24
Cameroon	9	Switzerland	25
Chile	10	Serbia	26
Ivory Coast	11	Nigeria	27
South Korea	12	Slovenia	28
Paraguay	13	Japan	29
Netherlands	14	North Korea	30
Portugal	15	New Zealand	31
America	16	Honduras	32

CONCLUSION

In this paper the mathematical model of mature, such as factor analysis, fuzzy comprehensive evaluation method, etc. To the complexity of the variables are integrated to a small number of several factors, and makes the team's overall strength has a reasonable analysis and ranking. In the process of solving the problem, through two kinds of methods to solve the same problem, can do a comparison, so that the solution of the problem is more reasonable. On the basis of the original model can be considered appropriate to increase the number of each soccer team has a star player, and good coaches level two factors, so that we can make the model more accurate reflection of the actual situation of the game.

REFERENCES

[1] Chen Jian, Yao Song-Ping; Journal of Shanghai Physical Education Institute, **33(5), (2009)**.
[2] Wang Luning, et al; China Sport Science, **19(3)**, 37-40 **(1999)**.
[3] Mao Jie, Shan Shuguang; Journal of Wuhan Institute of Physical Education, **46(2)**, 70-73, 87, **(2012)**.
[4] Zhang Lei; China Sport Science and Technology, **42(1)**, 50-52 **(2006)**.
[5] Fu Fan-Fei, Yu Zhen-Feng, Zhang Quan-Ning; Shandong Sports Science & Technology, **28(2)**, 24-25 **(2006)**.
[6] Wang Luning, et al; China Sport Science, **19(3)**, 37-40 **(1999)**.
[7] Li Ji-Hui. Journal of Shenyang Sport University, **25(2)**, 63-66 **(2006)**.
[8] Yang Yue-Qing, Rao Han-Fu, Len Ji-Lan; Journal of Hubei Sports Science, **22(2)**, 204-205 **(2003)**.
[9] Xiaomin Zhang; Journal of Chemical and Pharmaceutical Research, **5(12)**, 8-14 **(2013)**.
[10] Wang Bo, Zhao Yulin; Journal of Chemical and Pharmaceutical Research, **5(12)**, 21-26 **(2013)**.
[11] Mingming Guo; Journal of Chemical and Pharmaceutical Research, **5(12)**, 64-69 **(2013)**.
[12] Bing Zhang, S.Zhang, G.Lu; Journal of Chemical and Pharmaceutical Research, **5(9)**, 256-262 **(2013)**.
[13] Bing Zhang; Journal of Chemical and Pharmaceutical Research, **5(2)**, 649-659 **(2014)**.