World Cup in Brazil the Technical Indicators and the Results of Correlation Analysis

Chunfeng Xia*

Department of Humanities and Social Sciences, Zhejiang Industry Polytechnic College, Shaoxing 312000, Zhejiang, China

Abstract: Using the methods of documentary, mathematical statistics method for the World Cup football match statistics 20 technical indicators and the relationship between the team the outcome and correlation analysis, and then more football offensive and defensive technique index respectively, and then carries on the analysis, the result shows that shot, on target, possession, passing, corner of the five is one of the largest offensive technical index correlation with the result of the game, and the correlation between the result of the match and the biggest defensive foul technical indicators have interception, yellow card and 3 items. The entire statistic index is combined with the team winning games, principal component analysis; found that steals, possession, passing the closest relationship with the team winning games.

Keywords: World Cup, Technical indicators, Correlation, Principal component analysis

1. INTRODUCTION

World cup is a football match of highest level in the world, it goes together with Olympic Games and F1 to be known as global three top matches. It is held once per four years, any FIFA member state (region) can send delegation to register for the match. The statue was cast of pure gold, it weighted 1800 grams with the height of 30 centimeters, and erected in the pedestal of marble [1]. Schedule of world cup divides into preliminaries stage and finals stage, two stages world cup's preliminaries stage divide into six divisions to carry out that are respectively Europe, South America, Asia, Africa, North America and Oceania divisions, every division is required to regulate preliminaries rules according to the division own practical situation, and each FIFA member state delegation that has already registered for world cup should engage in preliminaries in located division and fight for qualification places in world cup finals stage [2]. The qualification places in world cup finals stage are by far 32 pieces, in finals stage, host country can be directly qualified for such stage, except for host country, other places are allocated by FIFA according to each preliminaries division football level, different preliminaries divisions will have different amounts of qualification places in finals stage. The finals stage 32 teams are divided into 8 groups by drawing lots, every group has four teams and they carry out grouping integral matches, every group top two teams totally sixteen teams are qualified to quarter-finals; after entering into quarter-finals, sixteen teams define schedule according to fixed rules, no single game and draw the knockout, decide until the title [3].

It is well-known that the 20th world cup host country is Brazil, they are totally 32 teams in the world cup and carry out 64 football games, finally Germany team wins the world cup with powerful strength. And meanwhile, in the world cup, FIFA professional analysis team makes scientific and reasonable analysis according to technical indicators that every game reflects, from which important technical indicators affect match that get involve in are possession, steal, pass, shooting, assist, foul and so on sixteen items. The paper bases on this, carries out correlation analysis of the sixteen items important technical indicators that affect match and match result, it gets connections and rules between the sixteen indicators and match result.

2. RESEARCH OBJECTS AND METHODS

Take the 20th football world cup that is Brazil football world cup in 2014, its participation in finals 32 teams totally 64 games as research objects. Acquire correlation data of Brazil world cup in 2014 from Netease world cup data center website, carry out correlation analysis of each technical indicator and teams match result, analyze each technical indicator influence on match result.

2.1. Research Methods

Documents literature: By searching and consulting lots of relative 2014 Brazil world cup, football attacking techniques, football defense techniques, tactics time factor, correlation analysis and other aspects mass literatures, carry out carefully read and analyze relative information to provide references for studying.

Video observation: By watching match video in CNTV, and through video observing applied sports techniques, tactics statistical methods, make statistics of match technical

indicators (goal, shoot, shoot on target, save, pass and others 16 pieces).

2.2. Data Statistics Method

1) By observing and organizing Brazil world cup teams data statistics, use EXCEL software to do statistical handling with each technical indicator, and form the data into table.

2) Apply IBM SPSS software to do correlation handling and significance test on each statistical index indicator and teams matches results relationships, by correlation analysis, it studies the session world cup each technical indicator and teams matches results correlations, and finds out positive and negative correlation indicators as well as significant correlation indicators important indicators.

3) Correlation coefficient, it refers to statistical indicator that reflect two variables linear correlation close degree on the condition of linear correlation is using r to express.

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

4) Correlation coefficient r value range is $-1 \leq r \leq 1$, $r > 0$, is positive correlation, $r > 0$ is negative correlation, $|r| = 0$ indicates no linear relations here, $|r| = 1$ indicates fully linear correlated.

$$\begin{cases} |r| \leq 0.3 \\ 0.3 < |r| \leq 0.5 \\ 0.5 < |r| \leq 0.8 \\ |r| > 0.8 \end{cases} \begin{array}{l} \text{No correlation} \\ \ \text{Low correlation} \\ \ \ \text{Significant correlation} \\ \ \ \text{High correlation} \end{array}$$

5)

6) Note:"**"indicates significant correlated in 0.01 level (bilateral) ;"*" indicates significant correlated in 0.05 level (bilateral).

3. RESULT AND ANALYSIS

Form table and organize with data network's data, due to data is enormous, here just list out two teams of great disparities in goal, from Table **1**, it is clear that the world cup Brazil and Germany team match performance statues, Germany wins in the world cup as chariot rolls host Brazil team.

3.1. Data Analysis

On the premise of ensuring comparative analysis speed and accuracy, the paper utilizes average integral statistical method to handle with each ranking region each continent teams match performance. Data processing result shows that in each ranking region, each continent teams average integrals are successively 16.140, 7.130, 2.190, 0.502. Therefore European team scores are the first, American teams are the second, African teams rank the third, and then Asian teams scores are at the bottom.

Table 1. 2014 Brazil football world cup Brazil and Germany two teams each technical indicator statistics.

Item	Brazil	Germany
Goal	11	18
Assist	9	13
Shoot on target	71	69
Shoot	110	96
Save	27	43
Off side	14	17
Yellow card	14	6
Red card	0	0
Pass	2287	3595
Steal	115	57
Foul	123	90
Intercept	513	498
Corner	43	37
Fumble	14	4
Possession percentage	54.1	56.9
Clearance kick	56	59

3.2. Each Ranking Section Teams Technical Indicators Comparison

1) Defense indicators comparison

Foul, red card, intercept, yellow card and steal form into defense technical indicators. By vertical comparing these five indicators, obtained results show that yellow card, intercept and foul these three indicators have significant differences. Therefore, it means that yellow card, intercept and foul these three indicators have great impacts on match results. Teams that rank in the top ten, their control in yellow card and foul are obviously higher than other ranking regions teams, result refers to Table **2**. Thereupon, it is clear that defense is very important to get good results in the match, defense levels decide ranking in the match.

2) Attack indicators comparison

Pass, shoot on target, steal, shoot, corner, off side and possession percentage form into attack technical indicators. By vertical comparing these seven indicators, obtained result shows corner, possession percentage, shoot, pass and shoot on target these five indicators have significant differences. Therefore, it shows that current each country football team more advocates attacking, and important factors to win the match is to ensure enough high shooting success rate and pass efficiency, result refers to Table **3**.

Table 2.   Each ranking section defense technical indicators comparison.

Ranking	1-10	10-20	20-32	F	P
Foul	15.36	19.21	18.53	2.16	0.05
Intercept	3.65	3.21	2.21	2.55	0.05
Yellow card	2.16	2.34	3.29	2.98	0.03

Table 3.   Each ranking section attack technical indicators comparison.

Ranking	1-10	10-20	20-32	F	P
Shoot	13.7	12.64	11.31	6.57	0
Shoot on target	8.16	7.04	6.53	1.49	0.22
Pass	78.54	78.94	72.46	9.41	0
Corner	5.61	5.49	4.71	5.59	0
Possession percentage	51.87	53.43	41.82	10.27	0
Steal	19.72	16.82	16.07	1.82	0.15
Off side	3.44	2.5	3.42	1.67	0.18

Table 4.   Each continent teams indicators comparison result.

	America	Asia	Africa	Europe	F	P
Shoot	10.57	10.81	11.88	12.47	1.37	0.26
Shoot on target	4.36	5.06	4.81	6.05	2.00	0.12
Pass	77.99	75.41	75.80	77.42	1.07	0.37
Corner	4.50	3.81	4.31	5.91	4.20	0.01
Possession percentage	49.28	49.51	47.16	51.13	0.89	0.45
Steal	16.07	18.19	15.25	18.27	1.25	0.30
Off side	2.80	2.25	4.69	2.89	3.46	0.01
Foul	18.20	18.69	21.34	17.26	2.11	0.10
Intercept	3.07	2.56	2.75	3.55	1.59	0.19
Yellow card	2.50	2.50	3.47	2.55	3.67	0.01
Red card	0.17	0.06	0.31	0.26	1.01	0.39

3.3. Make Statistics and Comparison of Each Continent Teams Technical Indicators

1) Defense technical indicators comparison

By horizontal comparing each continent teams technical statistical indicators, obtained results show that it has significant differences in yellow card such indicator, it specific reflects in African each team has most yellow cards in every match, almost 3.5 times per match, which is obviously more than other continents teams and can refer to Table **4**.

Above result shows that in foul, African teams and other continents teams obviously have significant differences, the rest three continents teams have insignificant differences, specific results refer to Table **5**. It shows that African players emotional fluctuation is bigger, action is also bigger therefore they are worse in controlling match.

2) Attack technical statistical indicators comparison

By horizontal comparing each team, obtained results show that continents have significant differences in off side and corner, especially for African teams, and other continents teams have insignificant differences. Similarly, it gets that each team in Africa advocates attacking, however, success rates are lower. In the amount of corner, European teams are stronger, almost get 5.9 pieces per game, specific

Table 5.   Each continents technical statistical indicators multiple comparison table.

Affiliated Continent	Yellow Card	Corner	Off Side
Asia	1.38**	0.5	2.44**
Europe	1.33**	-1.06*	1.79**
Africa and America	1.38**	-0.19	1.89**
America and Europe	-0.05	-1.41*	-0.09
Asia	0.05	0.69	0.55
Europe and Asia	0.05	2.10**	0.64

*It represents to arrive at significant correlation in the level 0.01;**It represents to arrive at significant correlation in the level 0.05.

result refers to Table **4**. Above result shows that European teams and rest three continents teams have significant differences in corner indicator, which indicates their attack efficiency is very high.

3.4.   Match Each Technical Statistical Indicator Correlation Analysis

Regards data statistical website group stage and knockout match as a whole to make correlation analysis, apply IBM SPSS software to make analysis and calculation, it gets following table data.

According to Table **6** statistical analysis, in this world cup group game stage, goal, shoot, shoot on target, assist, off side, possession percentage, pass these seven technical indicators and teams winning or losing are in positive correlations, which shows these technical indicators play positive roles in match result. Goal, assist and fumble these three technical indicators are in highly correlation with teams winning or losing. It is clear that in group game stage, goal and assist are upmost offensive indicators, teams with more goals will have more possibility to win, assist represents teams cooperative ability. And fumble is upmost defensive indicator, excessive fumble means bigger possibility of failure, except more goals to make up. Shoot on target, intercept, possession percentage, pass, steal and yellow card as well as others these six technical indicators are in moderate correlations with teams winning or losing, shoot, corner, foul, off side, block, red card and others six technical indicators are in low correlation with teams winning or losing . In group game stage, shoot on target effects on teams winning or losing are larger than shoot that because shoot on target represents shooting success rate, shooting success rate gets bigger, wishes to teams winning will be larger, and in world cup attention to shooting success rate is superior to shooting, save, possession percentage, pass , steal and other indicators are direct reflection of teams match controlling ability, are important foundation that teams make the goal. Corner, foul, block, red card and others lower correlation technical indicators correlations are lower that is because in world cup, these indicators quantities are relative fewer, differences among teams are not big, while pass is closely related to each team technical and tactics playing, its reflected correlation is also lower.

3.5. Principal Component Estimate

Principal component estimate was proposed by Massy in 1965, it was one kind of linear biased estimate in regression coefficient parameter. Principal component estimate adopted method is transforming original regression independent variables to another group of variables that are principal components, select a part of important principal components of them as new independent variables (now abandon a part of less influential independent variables, which actually arrive at the purpose of dimension reduction), and then use least square method to estimate model parameters after selecting principal components, finally transform back to original model to solve parameters estimation.

Set it has p pieces of regression variables $x_1, x_2, \ldots x_p$, , its value in the i test is:

$$x_{i1}, x_{i2}, \cdots x_{ip} \quad (i = 1, 2, \cdots, n)$$

Write them into matrix :

$$X = (x_1, x_2, \cdots, x_p) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

Considering linear model $Y = \hat{a}_0 1 + X\hat{a} + \mathring{a}$, $\mathring{a} \sim N(0, \acute{o}^2 I)$

Among them, Y is $n \times 1$ vector, $\hat{a}_0$ is unknown parameter, now:

$$\hat{a}_0 = \overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

For independent variables any one linear combination:

$$z = c_1 x_1 + c_2 x_2 + \cdots + c_p x_p, \quad \sum_{j=1}^{p} c_j^2 = 1$$

Regard z as new variable. Then z value in the i test is:

$$z_i = c_1 x_{i1} + c_2 x_{i2} + \cdots + c_p x_{ip} \quad (i = 1, 2, \cdots, n)$$

Record $w = (c_1, c_2, \cdots, c_p)^T$, then:

Table 6.   2014 Brazil football world cup each technical indicator and teams winning or losing correlation.

Technical Indicator	Pearson Correlation	Significance (Bilateral)	*n*
Goal	0.647**	0	32
Fumble	-0.186**	0	32
Assist	0.767**	0	32
Shoot on target	0.447*	0.028	32
Shoot	0.139	0.162	32
Save	-0.823	0.326	32
Steal	-0.205	0.079	32
Intercept	0.127	0.463	32
Foul	-0.075	0.74	32
Block	-0.042	0.341	32
Off side	0.267	0.159	32
Pass	0.397*	0.051	32
Possession percentage	0.499**	0.026	32
Yellow card	-0.168	0.156	32
Red card	-0.76	0.614	32
Corner	0.440*	0.961	32

*It represents to arrive at significant correlation in the level 0.01;**It represents to arrive at significant correlation in the level 0.05.

$$M_2^* = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n}z_i^2 = \frac{1}{n}(Xw)^T(Xw)$$

For new variable z, if its values changes under n times of test are not big, that is to say M_2^* is smaller, the new variable can be eliminated. On the contrary, M_2^* is bigger, then the new variable has larger changes, its effects are relative remarkable. Notes that the selection of c_i affects z_i value. Therefore in practical situation, selected $c_i = (i = 1,2,\cdots,p)$ should ensure that M_2^* gets the maximum value. Then it shows new variables effects on the newly-built regression model is larger.

Group game performance analysis, make statistics of this world cup group game performances to form a table, it gets Table **7**. Combine winning games of them with above offensive technical statistical indicators to make principal component estimate analysis by utilizing Matlab software.

By software calculating, it can get each technical indicator and match winning games correlation coefficient matrix. Then utilize Matlab software to calculate matrix feature value, matrix seven feature values are respectively:

0.0207,0.0770,0.2091,0.2843,0.6128,1.0716,4.7246

By above, it is clear that in feature values, the front two are smaller and tend to zero, the rest five sum accumulative contribution rate gets closer to 0.96, therefore the paper gets rid of the front two principal components, rest five feature values corresponding feature vectors are as following

$$n_1^T = (0.0189,0.1850,0.1369,0.1785,-0.8782,0.2453,0.2890)$$

$$n_2^T = (-0.1952,-0.1834,0.7463,-0.1610,0.0450,0.4529,-0.3717)$$

$$n_3^T = (0.1492,0.1970,0.2285,-0.8454,-0.0852,0-0.3000,0.2735)$$

$$n_4^T = (0.0294,0.0298,-0.4959,-0.3767,0.0970,0.7703,0.0876)$$

$$n_5^T = (0.4477,0.4440,0.3385,0.2895,0.4170,0.2245,0.4240)$$

Make linear regression of match data, it gets:

$$\hat{y} = -4.249 + 0.038x_1 - 0.034x_2 + 0.020x_3 + 0.077x_4 + 0.051x_6 + 0.052x_7$$

By above Table **8**, it is clear that in football match each statistical indicators, the ones that are most closely related to winning or losing are steal, possession percentage and pass these three items in attack technical indicators.

CONCLUSION

By making statistics and researching on this world cup data, combining with high correlation indicators, moderate correlation indicators and low correlation indicators, it finds that teams are consistent in goal, shoot and other indicators, and different in steal and other indicators.

According to correlation analysis between match result influential main technical indicators and result, obtained result shows assist and goal have largest correlations with match result, and then is pass success rate, times of shooting on target, shoot, possession percentage, corner, yellow card, intercept and else seven items, other technical statistical

Table 7.   Brazil football world cup each technical indicator and teams winning or losing correlation coefficients.

Name of Team	Winning Game	Name of Team	Winning Game	Name of Team	Winning Game
Brazil	2	Costa Rica	2	Germany	2
Mexico	2	Uruguay	2	America	1
Croatia	1	Italy	1	Portugal	1
Cameroon	-3	England	-2	Ghana	-2
Netherlands Netherlands	3	France	2	Belgium	3
Chile	2	Switzerland	2	Algeria	1
Spain	1	Ecuador	1	Russia	-1
Australia	-3	Honduras	-3	Republic of Korea	-2
Columbia	3	Argentina	3		
Greece	1	Nigeria	1		
Ivory coast	1	Bosnia and Herzegovina	1		
Japan	-2	Iran	-2		

Table 8.   Correlation coefficient matrix.

	Shoot on Target	Shoot	Steal	Off Side	Pass	Possession Percentage	Corner
Shoot on target	1	0.976**	0.677**	0.531**	0.864**	0.447*	0.921**
Shoot	0.976**	1	0.684**	0.507**	0.827**	0.444*	0.929**
Steal	0.677**	0.684**	1	0.516**	0.589**	0.011	0.605**
Off side	0.531**	0.507**	0.516**	1	0.541**	0.140	0.433*
Pass	0.864**	0.827**	0.589**	0.541**	1	0.499**	0.781**
Possession percentage	0.447*	0.444*	0.011	0.140	0.499**	1	0.440*
Corner	0.921**	0.929**	0.605**	0.433*	0.781**	0.440*	1

*It represents to arrive at significant correlation in the level 0.01;**It represents to arrive at significant correlation in the level 0.05.

indicators have no big effects on match result. In yellow card, African teams have significant differences with each continent teams, they obtained amount of yellow cards per game is larger than each continent teams, which shows defense is very important in the match, defense quality and efficiency affect trend of match. Each ranking region's each continent teams have significant differences in possession percentage, shoot on target, pass, shoot and corner, which shows master match attacking pace, enhance pass efficiency and shoot success rate are key to match.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

None declared.

REFERENCES

[1]   K. Liu, "Analysis of soccer globalization on world soccer sports development," *Bulletin of Sport Science & Technology*, vol. 16, no. 12, pp. 18-20, 2008.
[2]   G. Zhao, X. Han, J. Chen, L. Xie, Y. Suo, and X. Hu, "Reason analysis and countermeasures on misbehavior of football fans," *Bulletin of Sport Science & Technology*, vol. 19, no. 2, pp. 10-12, 2011.
[3]   X. Zhou, and Z. Mao, "The deviant behavior classification and the trend analysis of chinese soccer fans on the soccer field," *Sports & Science*, vol. 32, no. 6, pp. 103-106, 2011.