

Original Article

## STATISTICAL MODELING OF FOOTBALL MATCH PERFORMANCE DATA

Chinedu Obinna Ekezie

Science and Laboratory Technology  
Department, Imo State Polytechnic,  
Umuagwo, Imo State, Nigeria

DOI: <https://doi.org/10.5281/zenodo.17151519>

**Abstract:** This study is on the modelling of the recorded football match statistics of the group stage of the 2014 FIFA world cup. The data for the analysis was obtained from BBC sports. The researchers applied multiple regressions in the analysis, at 5% significant level. Two independent variables; ball possession and shot on target against the dependent variable goal. ANOVA and t-test statistic were used to test the hypothesis using STATA package version 13. The model shows positive linear relation between goal and shot on target, ball possession showed a weak negative relation on goal. ANOVA showed significant effect on goal. T-test showed that ball possession has no significant effect on goal while shot on target has a very significant effect on goal. The researchers recommend that football managers should go for more attacking and midfield players. And also, train players on how to hit target.

**Keywords:** Multiple Regressions, Ball possession, shot on target, Goal and STATA package.

### Introduction

Football as popular as it is in the world today, is no more just for entertainment and physical fitness, but a lucrative business both for the team owner(s), manager(s), players and the fans. Because of its lucrative nature today, team managers tend to adopt the best tactics that will give their team upper hand over their opponent.

Among these tactics includes possession football and attacking football.

Menon (2015) in his article "Possession Vs Counter-attack: who wins?" defines possession football as essentially revolving around a basic tenet-pass and move. Its aim is to ensure that the ball is kept in the team's possession for the longest time. He cited Barcelona Fc and Spain as the teams that have taken possession football to almost fanatic extreme. He noted that their strength are on their strong midfield anchored by Xavi, Iniesta and Busquets. He concludes by saying that there is constant movement of every player, given the man in possession of the ball at least two shot passes open at the same time to ensure that the ball is constantly moving away from the opposition.

On the other hand, some football analysts are of the opinion that attacking football is more result oriented than other tactics.

They argue that having much possession without attempting opponent's goal post will definitely not yield the expected result which is "goal scoring". Pep Guardiola (First season at Bayern Munich) in one of his statements said "I loathe all that passing for the sake of it, at that tiki taka. It's so much rubbish and has no purpose. You have to pass the ball with a clear intention with the aim of making it into the opposition's goal. It's not about passing for the sake of it." This agrees with Menon (2015) position that Pep Guardiola's heydays at Barcelona

## **Original Article**

were a combination of strong midfield along with the equally competent passing and selfless running of their forwards.

**Hart (2015)** in his article “Champion’s league analysis: Counter attacking football has usurped the possession game for Europe’s elite teams, UEFA evidence finds”. Noted that evidence in the Europa league shows that the team in the last 16 was the two with the lowest average possession-Sevilla and Dnipropetrovsk who reached the final. Also, European under21 championship, England had the highest possession ratio of the eight teams but went home after the group stage.

Many researcher works have been carried on football, being one of the most popular sporting events in the world. A lot of empirical questions have been raised by researchers. As such, the following findings were observed.

**Green Hough, Birch, Chapman and Rowlands (2002)** noted that Extremal distributions have been observed in a variety of complex systems and the results may then inform the modelling of football games.

**Lago and Martin (2007)** in their work on “Determinants of possession of the ball in soccer” observed that teams have more possession when they are losing matches than when winning or drawing, and the identity of the opponent matters – the worse the opponent, the greater the possession of the ball.

**Paulis, Perea and Mendo (2007)** in their “Diachronic analysis of interaction contexts in „06 World Championship” observed that „The succession of interaction contexts did not occur by chance but according to an internal logic, which could be explained by certain offensive contexts that appeared to be closer to scoring opportunities, and by other more propitious ball recovering situations”.

**Rowlinson and O'Donoghue (2007)** in there research on “Performance profiles of soccer players in 2006 UEFA Champions League and 2006 FIFA World Cup Tournaments” tried to know if UEFA Champions League tournament is a more quality competition than the FIFA World Cup. They observed that “There was much greater variability between player performances at the FIFA World Cup than during the UEFA Champions League. The current study does not provide sufficient evidence to support the view that there is difference in quality of the players” performances between the two tournaments”.

**Sajadi and Rahnama (2007)** in their work on “Analysis of goals in 2006 FIFA World Cup” concludes that “since highest number of goals was scored from the penalty zone (62%) and by the direct shots coaches should pay more attention to this area of the pitch and also on shots as well”,

**Hughes and Maloney (2007)** in “A technical analysis of elite male soccer players by position and success” observed that “The technically best team did not win the tournament; the successful teams during the early rounds had higher technique scores in all positions but in the semifinals and finals the losing teams had the higher technique scores”.

**Randers, Mujika, Hewitt, Santisteban, Bischoff1, Solano, Zubillaga, Peltola, Krusturp and Merce, Garcia, Pardo, Gallach, Mundina and González(2007)** in “Analysis of technical-tactical parameters in young soccer players” studied if soccer player’s achievement depends on variables such as psychology, physical condition, coordination and cognition. And they observed that “The present study showed no significant difference among the means of independent variables analyzed. Nonetheless, significant relations were found among the technical and tactical variables studied. It would be interesting to prolong observation times in future investigations to increase the likelihood of finding significant differences among the independent variables analyzed”.

**Frencken and Lemmink (2007)** tested the hypothesis that changes in the centre position of teams and changes in surface area precede goal-scoring opportunities. And conclude that Changes in the center position of teams, accompanied by changes in surface area, indicate goal-scoring opportunities.

## Original Article

**Lago (2009)** in “The influence of match location, quality of opposition, and match status on possession strategies in professional association football” finds that strategies in soccer are influenced by match variables, and teams alter their playing style during the game accordingly.

**Mohr1 (2010)** in “Application of four different football match analysis systems: A comparative study” advise that any comparisons of results between different match analysis systems should be done with caution.

**Bell, Brooks and Markham (2011)** in “The Performance of Football Club Managers: Skill or Luck?” Their findings suggested that there are a considerable number of highly skilled managers but also several who perform below expectations.

**Llu, Gomez, Lago-penas & Sampalo (2015)** in “Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup” observed that for all the games, nine match statistics had clearly positive effects on the probability of winning (Shot, Shot on Target, Shot from Counter Attack, Shot from Inside Area, Ball Possession, Short Pass, Average Pass Streak, Aerial Advantage and Tackle), four had clearly negative effects (Shot Blocked, Cross, Dribble and Red Card), other 12 statistics had either trivial or unclear effects. While for the close games, the effects of Aerial Advantage and Yellow Card turned to trivial and clearly negative, respectively.

**Papahristodoulou** (website material) observed that “Because the “shots on goal” is the strongest variable to goals scored, the statistics show that the following variables (for both teams) are strongly correlated with that variable. First, both teams should try to keep the ball within the team and shot on goal, only when the opportunity appears. They should avoid taking chances by shooting from non-favourable positions”.

In this study, the researchers will model the recorded football match statistics of the first round of FIFA 2014 world cup, with the view to know the effects of the two independent variables ball possession and shot on target have in determining the goal(s) of football match played in 2014 FIFA world cup.

The finding will help football managers to know the department to invest more when considering players to buy.

### Model

The researchers will combine two independent factors; ball possession and shot on target to know their effects on football match goal(s).

### Assumptions

Football team has three departments namely; defense, mid field and attack. Effective performances of these three departments will guarantee win for the team. Poor performance or average performance of one or two department(s) will either result to lose or draw for the team depending on the department(s) that is or are involved. Below average performances of the three departments is a sure loss to the team.

Hence we assume that;

- Ball possession shows the strength of the mid field
- Shot on target shows the strength of attack and defense
- Goal shows the strength of the attack

Therefore the possible outcomes of a football game are winning, lose and draw.

Win is a function of 65% and above ball possession plus 65% and above, plus 65% and above on goal attempts.

Loose is a function of 35% or less ball possession plus 35% or less goal attempts.

Draw lies between 45% to 55% ball possession and between 45% to 55% shot on target.

65% and above indicates strong mid field.

35% or less ball possession indicates weak mid field.

65% or above shot on target indicates strong mid field plus active attack on a weak defense.

35% or less shot on target indicates weak attack on strong defense.

Between 45% to 55% ball possessions indicate fairly equal mid field strength.

Between 45% to 55% shot on target indicates fairly equal mid field and attacking strength on weak defense.

## Original Article

### Definition of terms

Let; attack = A, mid field = M, defense = D, ball possession =  $\beta$ , shot on target =  $\rho$ , goal = y.

Let;

$\gamma' \geq 65\%$  ball possession

$\gamma' \leq 35\%$  ball possession

$45\% \leq \gamma' \leq 55\%$  ball possession

$\gamma' \geq 65\%$  shot on target

$\gamma' \leq 35\%$  shot on target

$45\% \leq \gamma' \leq 55\%$  shot on target

Since win, lose or draw is determined by number of goals, therefore let;

$\gamma' = \text{win}, \quad \gamma = \text{lose} \quad \gamma^- = \text{draw}.$

Hence we have the following equations;

$$\gamma' = \beta + \rho \quad 1$$

$$\gamma = \beta + \rho \quad 2$$

$$\gamma^- = \beta + \rho \quad 3$$

### Error term

We also take into consideration that sometimes, human factors may affect or alter the outcome of the match when the above conditions in the three equations have been met. Such human factors may come as a result of bias officiating from the referee and his assistant causing a team to win, lose or draw a match the outcome would have been otherwise.

Let  $R$  = fair officiating

$R^+$  = bias officiating

Therefore, equations 1 to 3 above becomes

$$\gamma' = \beta + \rho + \bar{R} \quad 4$$

$$\gamma = \beta + \rho + \bar{R} \quad 5$$

$$\gamma^- = \beta + \rho + \bar{R} \quad 6$$

Equation 4 to 6 is a generally expected if all things being equal.

But the officiating error may alter the equations and hence, the following may be possible.

$$\gamma' = \beta + \rho + R^+ \quad 7$$

$$\gamma = \beta + \rho + R^+ \quad 8$$

$$\gamma^- = \beta + \rho + R^+ \quad 9$$

$$\gamma' = \beta + \rho + R^+ \quad 10$$

$$\gamma^- = \beta + \rho + R^+ \quad 11$$

The possible outcomes different from what we have in equations 1 to 3 are still possible even when the officiating is normal which are very rare. Such are;

$$\gamma' = \beta + \rho + \bar{R} \quad 12$$

$$\gamma = \beta + \rho + \bar{R} \quad 13$$

$$\gamma^- = \beta + \rho + \bar{R} \quad 14$$

$$\gamma' = \beta + \rho + \bar{R} \quad 15$$

### Research Questions

1. Are goals, ball possession and shot on target linearly related?
2. Does ball possession have significant effect on goal?

## Original Article

3. Does goal attempt have significant effect on goal?
4. Does ball possession and goal attempt have significant effect on goal?
5. What is the level of the proportion of the variation in goal by the combination of ball possession and goal attempt?

### Hypothesis

1. **H<sub>0</sub>**: ball possession and shot on target have no significant effect on goal scored in 2014 FIFA World cup.  
**H<sub>1</sub>**: ball possession and shot on target have significant effect on goal scored in 2014 FIFA World cup.
2. **H<sub>0</sub>**: ball possession has no significant effect on goal scored in 2014 FIFA World cup.  
**H<sub>1</sub>**: ball possession has significant effect on goal scored in 2014 FIFA World cup.
3. **H<sub>0</sub>**: shot on target has no significant effect on goal scored in 2014 FIFA World cup. **H<sub>1</sub>**:  
shot on target has significant effect on goal scored in 2014 FIFA World cup.

We are considering the level of relationship between three variables goal the dependent, ball possession and shot on target the independent variables. In this case, we have two independent variables which require multiple regressions.

Multiple regression according to **Egbulonu** (2005) is a measure of relationship between one dependent variable (y) and K independent variable x, where  $K \geq 2$ . The model is given as;

$$y = \beta_0 + \beta_{1x_1} + \beta_{2x_2} + \beta_{3x_3} + \dots + \beta_{Kx_K} + e \quad 16$$

Therefore, a multiple regression with  $K=2$  become:

$$y = \beta_0 + \beta_{1x_1} + \beta_{2x_2} + e \quad 17$$

Since in the game of football, there are three possible outcomes win, lose or draw. A team win is other team loses. That is equation 4, 5 and 6 above.

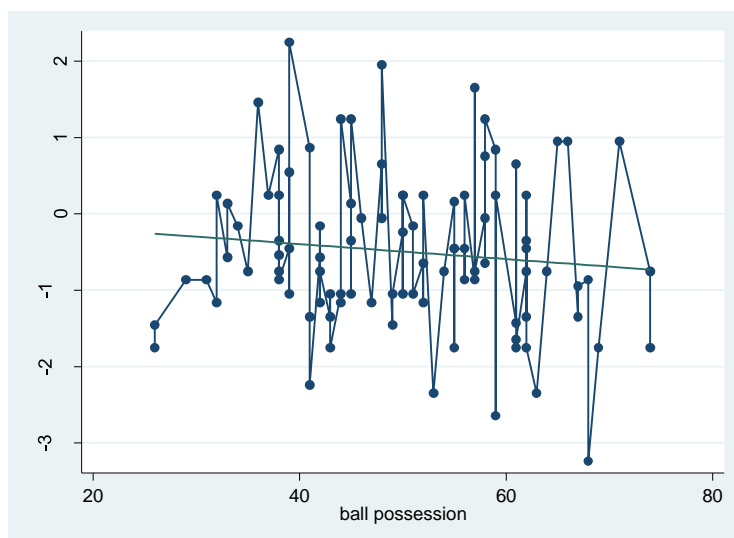
### Method of Data collection

The data as obtained from <http://www.bbc.com/sport/football/25285036>. The researcher extracted only the group stage matches of the 2014 FIFA world cup hosted in Brazil.

### Data Presentation/Analysis

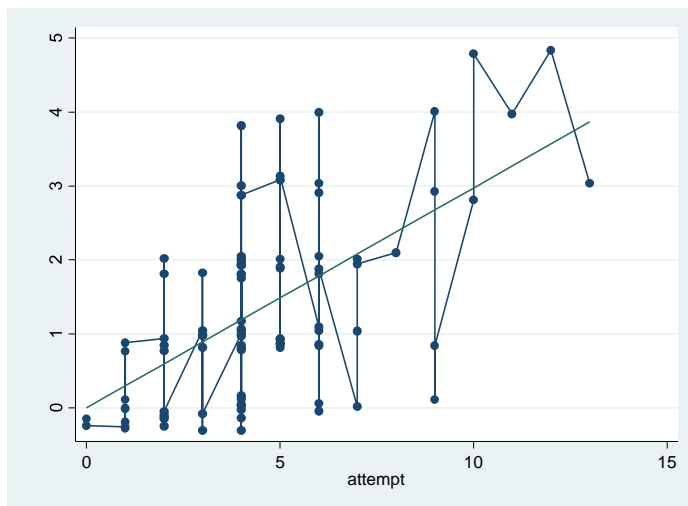
The data is presented in the two line plots below.

**Fig.1.** fitted line plot showing goals on ball possession



## Original Article

**Fig 2:** Fitted line plot showing goals on attempt



The data was analysis using STATA version 13 and the result as shown below:

**Table 1**

Source	SS	df	MS	p
Model	51.222	2	25.611	<0.0001
Residual	100.111	93	1.076	
Total	151.333	95	1.593	

n=96, R-squared = 0.3385, Adj R-squared = 0.3242 Root MSE= 1.038

Factor	coef	Std. Er	t	p	95% CI
Constant	0.566	0.484	1.17	0.245	-0.394 1.526
Possession	-0.010	0.010	-1.04	0.310	-0.029 0.009
Target	0.298	0.0432	6.88	<0.001	0.212 0.383

The model becomes;

$$\text{Goal(s)} = 0.566 - 0.010\beta + 0.298\beta_p$$

### Discussion and Conclusion

1. **H<sub>0</sub>:** ball possession and shot on target has no significant effect on goal scored in 2014 FIFA World cup.

**H<sub>1</sub>:** ball possession and goal attempt has significant effect on goal scored in 2014 FIFA World cup.

From table 1, Prob > F = 0.0000 shows that the variables combined have significant effect on goal. Therefore, we reject H<sub>0</sub> and conclude that ball possession and shot on target have significant effect on goals scored in 2014 FIFA world cup.



## Original Article

Hypothesis 2 and 3 requires individual test to ascertain the level of relationship each of the variables have on goals. A T-test was carried as shown in table2.

### Hence the result:

2. **H<sub>0</sub>**: ball possession has no significant effect on goals scored in 2014 FIFA World cup. **H<sub>1</sub>**: ball possession has significant effect on goals scored in 2014 FIFA World cup. From table 2, the ball possession has p-value 0.301, which shows no significant effect. We therefore reject H<sub>1</sub> and conclude that ball possession has no significant effect on goals scored in 2014 FIFA world cup.
3. **H<sub>0</sub>**: shot on target has no significant effect on goal scored in 2014 FIFA World cup. **H<sub>1</sub>**: shot on target has significant effect on goals scored in 2014 FIFA World cup.

From table 2, shot on target has p-value 0.000, which shows significant effect. We therefore reject H<sub>0</sub>, and conclude that shot on target has significant effect on goals scored in 2014 FIFA World cup.

### Observations/Recommendations

Base on the findings above, we have ascertained from the ANOVA that there is linear relationship between the two independent variables and goals, but the t-test specifically showed that it is shot on target that has significant effect on goals. We observe the followings;

1. That having higher ball possession does not guarantee a team winning a football match.
2. That teams that have strong attack have greater chances of winning a football match.
3. That a team that has the greater ball possession but lacks the ability to attack may end up losing or drawing football matches
4. That a team that has the greater ball possession and stronger attacks has the greatest chance of winning matches.

These observations agree with Hart's article. We therefore, recommend that football mangers should go for more attacking and mid-field players. And also, train players on how to hit targets.

### References

- Anirudh, M.(2015). *Possession Vs Counter-attack: who wins?* BBC sports. Retrieve from [www.sportkeeda.com/keeda/anirudhmenon](http://www.sportkeeda.com/keeda/anirudhmenon).  
<http://www.bbc.com/sport/football/25285036>.
- Bell. (2011). *The Performance of Football Club Managers: Skill or Luck? ICMA Centre Discussion Papers in Finance DP*
- Christos, P.(n.d). *An analysis of UEFA Champions League match statistics retrieved from www.diva-portal.org/smash/get/diva2:37338/fulltext02*.
- Egbulonu. (2005). *Basic Econometric Methods*. Peace Publishers, Owerri. Nigeria.
- Frencken .(2007) . *Successful performance in soccer: Team kinematics of goal-scoring opportunities in small-sided soccer games. Journal of Sports Science and Medicine*. Greenhough. (2002). *Football goal distributions and extremal statistics. Physica A* 316.
- Hughes. (2007). *A technical analysis of elite male soccer players by position and success. Journal of Sports Science and Medicine*.
- Lago. ( 2007). *Determinanants of Possession of the ball in Soccer.Journal of Sports Sciences* .

**Original Article**

- Lago. (2009). *The influence of match location, quality of opposition, and match status on Possession strategies in professional association football* . *Journal of Sports Sciences*.
- Llu. ( 2015) .*Match statistics related to winning in the group stage of 2014. Brazil FIFA World Cup*.
- Merce (2007) . *Analysis of technical-tactical parameters in young soccer players*. *Journal of Sports Science and Medicine*.
- Paulis (2007) . *Diachronic analysis of interaction contexts in '06 World Championship* *Journal of Sports Science and Medicine*.
- Randers . ( 2010). *Application of four different football match analysis systems: A Comparative Study* *Journal of Sports Sciences*.
- Rowlinson. (2007). *Performance profiles of soccer players in 2006 UEFA Champions League and 2006 FIFA World Cup Tournaments*. *Journal of Sports Science and Medicine*.
- Sajadi .(2007) . *Analysis of goals in 2006 FIFA World Cup*. *Journal of Sports Science and Medicine*.
- Simon, H.(2015). *Champion's league analysis: Counter attacking football has usurped the possession game for Europe's elite teams, UEFA evidence finds* retrieve from [www.independent.co.uk/author/simon-hart](http://www.independent.co.uk/author/simon-hart).