A study of the Performance of Premier League teams measured by the goals scored in the first 45 minutes and the second 45 minutes of the games in the period 2000-2019

Nicholas Stylianides

February 2020

Abstract

This paper will provide an insight into the performance of Premier League teams since 2000 in terms of the number of goals scored in each half of the game to assist football team managers and players developing their game strategies. It will also convince spectators to remain in the football stadium until the referee whistles the end of the game. Firstly, for the entire period of 2000 – 2019; secondly, a comparison of the first nine years, 2001 – 2010 with the next nine years, 2011 – 2019; and thirdly, the study is extended to compare the performance of the teams in the extra time of the first half with the added time of the second half, in terms of the goals scored.

1 Introduction

Football is one of the most watched sports in the world and millions of people go to the stadium every year to support their team.

The Premier League was founded in 20 of February in 1992. It is often referred to as the English Premier League or the EPL and is the highest division of the English football league system. Premier League is formed by 20 football clubs. The season begins in August and ends in May and each football club plays 38 matches, 19 of which are played in the home stadium and the other 19 in the opponent's stadium. [1]

My aim in this paper is to present information on the number of goals scored by all teams during each half of the game and testing the following hypothesis:

1. Test the hypothesis that more goals are scored in the second half of the games than in the first half for the period 2000 - 2019,
2. Test the hypothesis that more goals are scored in the period 2011 – 2019 than in the period 2001 – 2010 and
3. Test the hypothesis that more goals are scored in the overtime of the second half than in the overtime of the first half for the period 2000 – 2019.

2 Choosing a league

Football is a very competitive sport and attracts attention worldwide. Each country has its own league consisting of a number of professional football clubs registered in the said country such as: Premier League (UK) [2], Bundesliga (Germany), Serie A (Italy), La Liga (Spain), and Ligue 1 (France). These leagues are some of the most competitive leagues in the world. The Premier League was chosen at random by a random generator for this paper and all information provided in this paper refers to the performance of the said league.

3 Methodology

The methodology I used to test the three hypotheses indicated above is based on the use of the student t – distribution (independent sample):

$$t = \frac{|x_1 - x_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Once the Premier League was chosen as the testing football league, Table 1 was prepared presenting the number of goals scored by Premier League teams in each half of the game for the period 2000 – 2019 and is the data I used for my study. [3] At the end of the table, I have calculated the Mean, the Standard Deviation and the Variance of the 19 periods. The formulae for the mean and the standard deviation are shown below:

Mean: [4]

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Standard Deviation: [5]

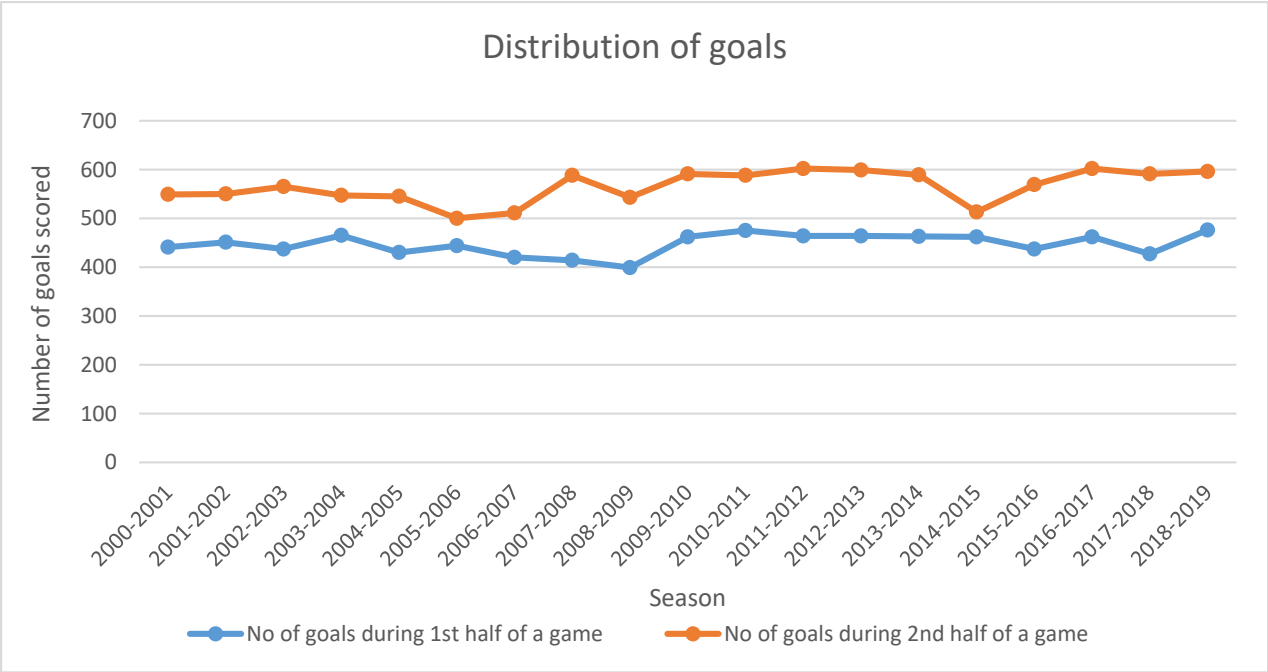
$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Table 1

Season	No of goals during 1st half of a game	No of goals during 2nd half of a game	Total no of goals in a season
2000-2001	441	549	990
2001-2002	451	550	1001
2002-2003	437	565	1002
2003-2004	465	547	1012
2004-2005	430	545	975
2005-2006	444	500	944
2006-2007	420	511	931
2007-2008	414	588	1002
2008-2009	399	543	942
2009-2010	462	591	1053
2010-2011	475	588	1063
2011-2012	464	602	1066
2012-2013	464	599	1063
2013-2014	463	589	1052
2014-2015	462	513	975
2015-2016	437	569	1006
2016-2017	462	602	1064
2017-2018	427	591	1018
2018-2019	476	596	1072
Total	8493	10738	19231
Mean	447	565.1579	N/A
Standard Deviation	21.76899	32.82151	N/A
Variance	473.88889	1077.251	N/A
n	19	19	N/A

A schematic representation of the distribution of goals in the first half and in the second half since 2000 is provided in Graph 1.

Graph 1



The graph shows that the number of goals scored in 2nd half are greater than the number of goals scored in the 1st half. It also shows that over the nineteen years under investigation the number of goals scored are steady around 560 goals in the 2nd half and around 450 in the 1st half.

1. <u>Test the hypothesis that more goals are scored in the second half of the games than the first half over the period 2000 - 2019,</u>

The data in Table 1 show that the number of scoring goals in the second half is higher than the number of goals in the first half for all seasons since 2000. I used the t – distribution with the following hypothesis:

$$H_0: \mu_{first\ half} = \mu_{second\ half}$$

$$H_1: \mu_{first\ half} \neq \mu_{second\ half}$$

with a significant figure α = 0,05 and a degree of freedom equal to 36. In doing my calculation of the mean, standard derivation and variance between the two halves I used the t – distribution formula which gave me the answer

$$t = \frac{|447 - 565.1579|}{\sqrt{\frac{473.8889}{19} + \frac{1077.251}{19}}} = 13.0597.$$

Because I was doing a two-tail test, I had to divide my significant figure by two. By comparing it with my t_{stat} I can see that $t_{stat} > t_{critical}$. I conclude that I am going to reject my H_0 and accept H_1. This shows that more goals were scored in the second half compared to the first half over the period 2000-2019.

2. Test the hypothesis that more goals are scored in the period 2011 – 2019 than in the period 2001 – 2010

To proceed with the 2nd hypothesis I used the data divided in two time intervals i.e. from 2001 to 2010 and from 2011 to 2019. I have calculated the average number of goals scored in the first half and in the second half for:

(a) the nine-year period from 2001-2010
(b) the nine-year period from 2011 to 2019.

Additionally, I calculated the Standard Deviation and Variance of these periods. This information is presented in Table 2.

Table 2

Period	1st half	2nd half	Total	Average 1st half	Average 2nd half	Standard deviation 1st half	Standard Deviation 2nd half	Variance 1st half	Variance 2nd half
2001-2010	3922	4940	8052	435.7778	548.8889	22.22486	30.50182	493.944	930.3611
2010-2019	4130	5249	10189	458.8889	583.2222	16.3453	28.21249	267.1111	795.9444

To investigate if the teams scored the same number of goals in the first half of each period I used the t – distribution:

$$t = \frac{|x_1 - x_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

My H_0 being the average number of goals scored in the first half of period 2001 – 2010 is equal to the average number of goals scored in the first half of period 2011-2019. My H_1 hypothesis is that the average number of goals scored in the first half of the first period (2001-2010) is different from the average number of goals scored in the period 2011-2019.

$$H_0: \mu_{first\ half\ 2001-2010} = \mu_{first\ half\ 2011-2019}$$

$$H_1: \mu_{first\ half\ 2001-2010} \neq \mu_{first\ half\ 2011-2019}$$

$$t = \frac{|435.7778 - 458.8889|}{\sqrt{\frac{493.9444}{9} + \frac{267.1111}{9}}} = 2.51324.$$

By applying the t – distribution and comparing it with the significant figure of α = 0,05 and degree of freedom 16, I conclude that we reject the null hypothesis and accept the hypothesis that the two first halves of each 9 years are not the same.

Moreover, I have made another t –distribution comparing the goals scored in the second half of each group of the two periods. I have used the same hypothesis as before but this time I have taken the goals scored in the second half of each period of 9 seasons. I calculated the t – distribution t = 2.7901 which is bigger than t critical. Based on this calculation I reject the null hypothesis and accept the hypothesis H_1 i.e. the number of goals scored in the second half of the first period is different from the number of goals scored in the second period for the last 9 seasons.

4 Further Findings

5 Test the hypothesis that more goals are scored in the extra time of the second half than the extra time of the first half for the period 2000 – 2019.

In addition to the above, I was also very interested to compare performance data of Premier League teams in the extra time of each half of a game for all seasons since 2000.

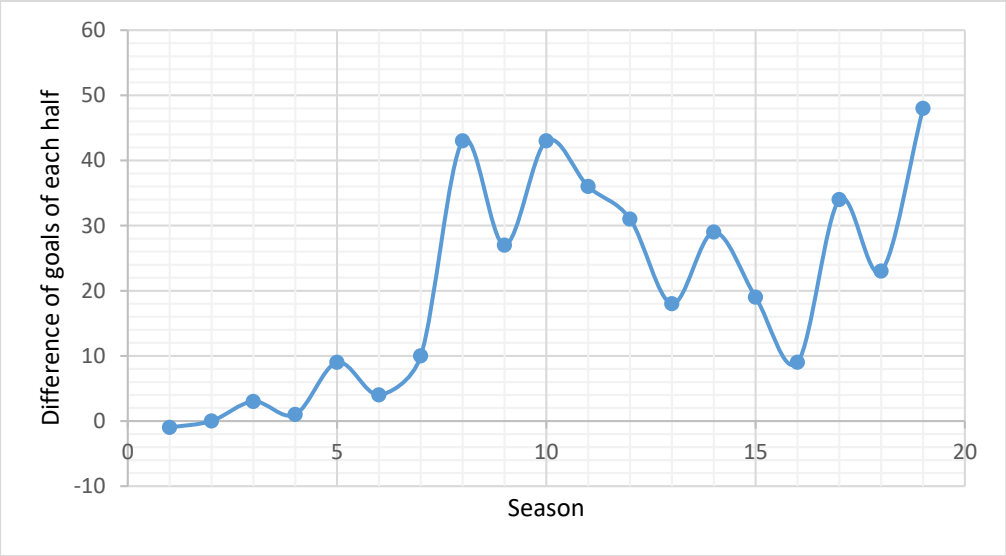
Table 3 shows that, with the exception of season 2000 -2001, the number of goals scored in the extra time of the second half is higher than the goals scored in the extra time of the first half of the game. The highest number of goals scored at the final minutes of the first half was in season 2015-2016 i.e. 37 goals (8% of the total goals scored in the first half of the game). On the other hand, the highest number of goals scored at the final stage of the game was recorded in the last season 2018-2019 i.e. 71 goals (12% of the total goals scored in the second half of

the game). Graph 2 shows the difference in the number of goals scored between the extra time of the second half with the number of goals scored in the extra time of the first half.

Table 3

Season	45+	90+
2000-2001	1	0
2001-2002	0	0
2002-2003	0	3
2003-2004	1	2
2004-2005	0	9
2005-2006	0	4
2006-2007	16	26
2007-2008	13	56
2008-2009	24	51
2009-2010	24	67
2010-2011	17	53
2011-2012	25	56
2012-2013	27	45
2013-2014	23	52
2014-2015	21	40
2015-2016	37	46
2016-2017	33	67
2017-2018	26	49
2018-2019	23	71

Graph 2



5 Conclusion

Based on the above data and the calculations carried out we conclude that all teams of the Premier League try to score more goals at the final stages of a game to secure the victory or try to tie the game. One may argue that the above conclusion is expected, considering that teams play harder in the second half to win the game. Moreover, teams usually change their tactics during the second half to be more effective, such as pulling a defender and putting in an attacker to increase the chance of the team to score higher.

The only season that the above conclusion does not apply is 2000 – 2001. In this season the teams scored 1 goal in the extra time of the first half and no goals in the extra time of the second half. Possible explanations may be that:

(a) there was not enough extra time after the second half to score goals

(b) teams were more eager to secure the victory early on in the game

(c) the teams followed different tactics than in the following years.

References

[1] En Wikipedia.org 2020 *Premier League history*. Available at:

https://en.wikipedia.org/wiki/Premier_League

[2] Premier League.com 2020. *Premier League* Available at:

https://www.premierleague.com/

[3] Transfermarkt.co.uk 2020 *Distribution of goals in a 90 minutes' match*. Available at:

https://www.transfermarkt.co.uk/premier-league/torverteilungminuten/wettbewerb/GB1/plus/1

https://www.transfermarkt.co.uk/premier-league/torverteilungminuten/wettbewerb/GB1

[4] Statistics how to 2020 *Sample mean formula*. Available at:

https://www.statisticshowto.datasciencecentral.com/sample-mean/

[5] Math is fun.com 2020 *Standard deviation formula for sample*. Available at:

https://www.mathsisfun.com/data/standard-deviation-formulas.html