ORIGINAL ARTICLE

The effect of some statistical parameters on soccer success using the example of the Turkey Super League*

MEHMET ALPASLAN[1], FIRAT AKCAN[2]
[1]Gaziantep University, Faculty of Sport Sciences. Gaziantep –Turkey. Orcid: 0000-0003-3276-2520 (asln.mehmetalp@gmail.com)
[2]Gaziantep University, Faculty of Sport Sciences. Gaziantep –Turkey. Orcid: 0000-0002-3610-0973 (firatakcan@yahoo.com)
Corresponding author, Dr. Fırat AKCAN, Email: firatakcan@yahoo.com, Cell: +905304684243

ABSTRACT

Background: Soccer is one of the most popular sports in the world. Studies on soccer match statistics have a long history. In recent years, clubs have established match analysis units related to performance in soccer and have started to formulate training plans and match strategies in light of the resultant information.

Aim: The aim of this study is to investigated the effects of some statistical data, i.e. technical and conditional features, on soccer success.

Method: The effects of various statistical data on teams' performance in the last three seasons, measured by their end-of-season scores, were examined. The average effects of teams per game, total running distances (TRD), sprint distances (SD), high intensity running distances (HIRD), accurate shot per game (SPG), number of accurate passes per game (APG), pass percentage per game (PPG), number of key passes per game (KPG), and possession of the ball (PB) on total points (TP) at the end of the season were examined. SPSS22 was used for data analysis, which revealed no relationship between total seasonal scores and the conditional parameters, whereas a strong positive correlation was found with the technical parameters ($p < 0.05$)

Results: There was no statistically significant difference in the effect of conditional properties on total end-of-season score, while the technical characteristics yielded a statistically significant prediction ($p < 0.05$).

Conclusion: Consequently, considering the Turkey Super League's last three seasons, it was observed that the technical features positively affected success, whereas the conditional features were not determinant in success. Teams that want to be successful should therefore develop their technical characteristics.

Keywords: soccer, statistics, technical features, conditional features, Turkey Super League

INTRODUCTION

Soccer is one of the most popular sports in the world[1-2]. In order to be successful in soccer, scientific programs should be developed, and training plans and match strategies should be devised accordingly[3]. Studies on soccer match statistics have a long history[4-5]. In recent years, clubs have established match analysis units related to performance in soccer and have started to formulate training plans and match strategies in light of the resultant information[6].

Statistics has inevitably found a place in soccer, as in almost every field, and has become an indispensable element in soccer clubs[7-8]. Statistics has led to a better understanding of soccer and contributed to the creation of policies for the sport's development. It is especially important to improve performance on the field[9-10]. In sports clubs where soccer is played at a high level, soccer managers definitely include statistical experts on their teams. As members of the technical team, such experts report to the manager with statistical data reflecting the team's overall performance as well as individual players' performance. Managers can then shape their decisions in line with the content of the statistical reports. Statisticians also collect information aboutrival teams[11]. It is thought that soccer managers who understand the importance hierarchy pertaining to the factors determinant in soccer success are in an advantageous position. Although soccer is a universal game, it may differ according to the characteristics of each region where it is played. It is therefore crucial to identify these differences and understand their effects on success[12]. On this premise, this study examined the relationship between teams' end-of-season total scores and their technical and conditional properties. The effects of statistical data reflecting teams' technical and conditional characteristics on the total score at the end of the season were also examined. This study is the first to use a detailed data set covering three seasons in Turkey. A review of similar studies conducted worldwide indicated that no similar study is available in the literature. It is thought that this study's results can contribute to a better understanding of soccer in general. Moreover, identifying the successful teams in the Turkey Super League will contribute to the identification of areas where performance can be enhanced based on statistics.

MATERIALS AND METHODS

Research Design: The scanning method was used in this study. Scanning model Scanning models are a research approach that aims to describe a past or present situation as it exists.

Sample: The sample comprises the match statistics of all matches played in the Turkey Super League during the 2017–2018 and 2019 seasons. The Turkey Super League consists of 18 teams that compete in the national soccer championship. Teams play each other twice per season (i.e. on a balanced schedule), once during Rounds 1 through 17 and once during Rounds 18 through 34 when the status of home and visiting team is reversed. The outcome of a match (i.e. win, draw, or loss) determines the number of points that are allocated to the teams; no points are allocated to the loser of a match, but each team receives one or three points in the case of a draw or a win, respectively. A season lasts for about ten months (September through July), and teams typically play one match per week. The number of points accumulated during a season determines the team's league ranking.

Data Collecting Design: Match statistics were obtained from the Sentio Sports Analytics Company. The written consent was obtained from the Company, due to using their analysis data. According to the Company data collection rules, professional technicians (at least 5 years of experience in performance analysis in soccer) gather the variables and save to the data bank of the Company. The company uses the inter-operator reliability of the company's observational system (OPTA Client System) to collect match statistics data. It is stated that the reliability of the data collected through this system is at the appropriate Kappa level[13]. Total points (TP): total points achieved by each team after 34 games played in a full season consisting of 18 private teams. Total running distance (TRD): the average of each team's total running distance in the game at the end of 34 games played in a full season consisting of 18 private teams in km. Sprint distance (SD): the average of the total sprint distance of each team in the game at the end of 34 games played in a full season consisting of 18 private teams in km. High-intensity running distance (HIRD): the total speed of each team in the game at the end of 34 matches in a full season consisting of 18 teams is the average of 20 km/h or more running distance per game in km. Accurate shots per game (SPG): the average number of shots for each team that reached the goal after 34 matches in a full season consisting of 18 private teams. Number of accurate passes per game

(APG): expression of the average number of passes per match made by each team at the end of 34 games played in a full season consisting of 18 private teams. Pass percentage per game (PPG): percentage expression of the average of the total passes per game for each team at the end of 34 games played in a full season consisting of 18 private teams. Number of key passes per game (KPG): the average number of key passes per match made by each team at the end of 34 games played in a full season consisting of 18 teams. Possession of the ball (PB): percentage expression of the average of the total time for which each team played with the ball over 34 games in a full season consisting of 18 teams.

Statistical Analysis: The SPSS22 package program was used to analyse the data sets. The Shapiro Wilk test was used to test the data sets' normality assumptions. After making the assumption of normality, the average and standard deviation values of the data are given. Pearson's correlation test was used to test the relationship between datasets, and multiple regression analysis was applied to examine the parameters' effects on the total end-of-season score. In addition, the one-way ANOVA test was applied in order to understand whether there was a significant difference in the parameters between seasons, and the LSD test, one of the post hoc tests, was used to identify the seasons in which there was a significant difference.

RESULTS

The frequency and percentage values of the answers given to the questionnaire questions by the participating football players are presented in tables.

Table 1. Mean and standard deviation values of the parameters.

			TP	TRD km	SD km	HIRD km	SPG	APG	PPG %	KPG	PB
			N = 18	*N* = 18	*N* = 18	*N* = 18	*N* = 18	*N* = 18	*N* = 18	*N* = 18	*N* = 18
2016 2017		Mean.	47.27	110.08	2.60	4.69	4.30	372.88	81.27	7.77	49.94
		Std. dev.	14.70	2.46	0.21	0.26	0.69	57.02	2.24	1.21	4.34
2017 2018		Mean.	47.16	109.00	2.51	4.58	4.48	384.38	81.88	7.27	49.88
		Std. dev.	16.56	1.53	0.14	0.18	0.82	56.76	2.32	1.49	4.56
2018 2019		Mean.	46.05	109.58	2.48	4.47	4.53	384.50	82.44	6.73	49.83
		Std. dev.	11.92	2.14	0.15	0.25	0.64	44.31	1.46	1.32	4.01

Table 2. Analysis of the relationship between TP and other parameters.

	Variable			TRD km	SD km	HIRD km	SPG	APG	PPG %	KPG	PB
	2016 2017	TP	*r*	-0.18	-0.18	0.22	**0.75**	**0.68**	**0.58**	**0.74**	**0.76**
			p	0.47	0.45	0.37	**0.01**	**0.02**	**0.01**	**0.01**	**0.01**
			N	18	18	18	18	18	18	18	18
2017 2018		TP	*r*	0.201	0.139	0.28	**0.76**	**0.81**	**0.72**	**0.83**	**0.78**
			p	0.423	0.58	0.25	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
			N	18	18	18	18	18	18	18	18
2018 2019		TP	*r*	-0.17	0.05	0.12	**0.61**	**0.70**	**0.65**	**0.72**	**0.71**
			p	0.49	0.81	0.62	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
			N	18	18	18	18	18	18	18	18

Table 2 shows that there is no significant relationship between TRD, SD, HIRD, and TP in any season ($p < 0.05$). In the 2016–2017 season, there was a moderately positive and significant relationship between PPG and TP, while a strong positive and significant relationship was observed between other parameters and TP, except for TRD, SD, and HIRD ($p < 0.05$). For the 2017–2018 season, there was a very strong positive and significant relationship between PPG and KPG and TP, and between other parameters and TP, except TRD, SD, and HIRD ($p < 0.05$). A strong positive and significant relationship was found between all parameters and TP in the 2018–2019 season, except TRD, SD, and HIRD ($p < 0.05$).

Table 3. Results of multiple regression analysis to estimate the effect of TRD, SD, SPG, and KPG on TP.

Variable		Adjs. R^2	B	Std. dev.	t	p	VIF
2016–2017	Constant*	**0.519**	-68.450	126.304	-0.542	0.596	
	TRD		0.645	1.073	0.601	0.557	1.146
	SD		-10.210	11.506	-0.887	0.390	1.010
	SPG		16.575	3.788	4.375	**0.001**	1.141
2017–2018	Constant*	**0.427**	-386.221	222.718	-1.734	0.105	
	TRD		-0.047	2.112	-0.22	0.983	1.131
	SD		7.751	20.958	0.370	0.717	1.062
	KPG		5.117	1.358	3.769	**0.002**	1.076
2018–2019	Constant*	**0.412**	-250.850	163.020	-1.539	0.146	
	TK		-1.797	1.181	-1.522	0.150	1.309
	SD		14.960	16.495	0.907	0.380	1.301
	KPG		5.539	1.520	3.643	**0.003**	1.007

Table 3 shows that the conditional parameters TRD and SD have no effect on TP in all seasons (p < 0.05). On the other hand, it is observed that the technical features of SPG and KPG have a significant effect on TP in all seasons (p < 0.05). In addition, based on the strong correlation between the technical features of APG, PPG, and PB with KPG and SPG, the effect of the related parameters on TP at the end of the season is thought to be highly predictable.

Table 4 shows that a statistically significant difference was observed only in high-intensity running distance between the 2016–2017 season and the 2018–2019 season, while no statistically significant difference was found in other parameters in any season (p < 0.05).

Table 4. Results of ANOVA test to understand the between-season differences.

Variable	2016–2017 (1)		2017–2018 (2)		2018–2019 (3)		p	Difference
N = 18	Av.	S. dev. (±)	av.	S. dev. (±)	Av.	S. dev. (±)		
TP	47.28	14.708	47.17	16.564	46.06	11.924	0.962	-
TRD (km)	110.08	2.468	109.03	1.531	109.58	2.148	0.308	-
SD (km)	2.60	0.216	2.51	0.149	2.48	0.153	0.094	
HIRD (km)	4.69	0.269	4.58	0.181	4.47	0.259	0.024	1-3
SPG	4.30	0.697	4.48	0.823	4.53	0.641	0.600	-
APG	372.88	57.02	384.38	56.760	384.50	44.312	0.753	-
PPG (%)	81.27	2.244	81.88	2.323	82.44	1.464	0.241	-
KPG	7.77	1.215	7.272	1.496	6.738	1.325	0.079	-
PB (%)	49.94	4.345	49.88	4.561	49.83	4.018	0.997	-

DISCUSSION

This study's findings are discussed in comparison with similar studies in the literature. Soccer emerges as a sports branch where competition is felt more and more every day. There may be very small details that make teams successful in top leagues[14-15]. Based on the statistical analysis, it was concluded that the effect of technical features on end-of-season scores for all three seasons was statistically more significant than that of the conditional parameters (p < 0.05). Lago-Peñas et al. (2010) compared the statistical data of the winning, losing, and tied teams in the Spanish League and stated that the winning teams had more ball possession than the losing teams and that the percentage of passes and total assists, as well as the number of shots taken towards the goal and those resulting in a goal, were statistically significant[16]. Grant et al. (1999) pointed out that the teams that made it to the semi-finals in the 1998 World Cup had a significantly higher percentage of passes with possession of the ball according to statistical parameters compared to other teams[17]. Hook and Hughes (2001) examined the European Soccer Championship using a method similar to that used in a previous study and reported that the ball possession rate differed statistically significantly in unsuccessful teams[18]. Jones et al. (2004), in a similar study on the English Premier League, found that successful teams had a statistically significantly higher rate of ball possession than unsuccessful teams[19]. Yue et al. (2014) analysed the statistics of the teams that won the

Bundesliga matches in Germany and listed the statistical parameters that were effective in winning the match in order of importance. They stated that by far the most important parameter is the effective goal attempt (goal/shot) ratio, which is defined as the ratio of shots taken towards the goal to the shots that turned into a goal. Other important parameters that affect winning a match in soccer are, in descending order of importance, the total number of shots, the total number of passes, and the ball possession rate[20]. Hughes and Franks (2005) examined the total number of passes made before the goals scored in the World Cup final matches and reported that the number of passes and the ball possession rate prior to scoring the goal were higher among the winning teams than the losing teams. They reported a statistically significant relationship between the number of goals scored and the rate of possession and total passes[21]. Lago and Martin (2007) attempted to understand the differences depending on whether a team is the home team, ahead in terms of score, drawn, or behind in terms of score. They reported that the home team possession rate was higher than that of the guest team and that the teams had a statistically significant higher percentage of ball possession than the teams that were ahead in terms of score when losing[22]. Lago (2007) analysed the performances of winning and losing teams in the 2006 World Cup and reported that the performance parameters of the first division games, played according to the league method, were significantly better for the winning

teams than the losing teams. On the other hand, he stated that the performance statistics of the winning teams and the losing teams in the next round matches, played according to the single match elimination method, did not make a statistically significant difference[23]. Rein and Memmert (2016) emphasised the importance of statistical data and tactical analysis in elite soccer and stated that having different field experts evaluate these data using a multidisciplinary approach could yield useful results[24]. Mao et al. (2016) analysed the matches played in the Chinese Professional Soccer League in the 2014–2015 season and found that the teams that were successful at the end of the season had statistically significantly higher parameters than the unsuccessful teams, specifically for shot rate and shot percentage when a goal was scored[25].

Similar studies in the literature support the results obtained in this study. Our study investigated the impact of conditional properties with special specifications on success in the Turkey Super League. It was concluded that for the last three seasons, the effect of technical characteristics on success was high, whereas conditional characteristics were not a determining factor in achieving success. Despite these results, in soccer, conditional characteristics cannot be ignored; the fact that the teams' conditional characteristics are similar, i.e. no statistically significant difference was found, affects the results of the study. Considering that soccer is of great importance in our country, as in the rest of the world, it has been observed that there are not enough scientific studies in the field of soccer, specifically pertaining to Turkey. It is thought that providing the necessary support to both sports scientists and all individuals, institutions, and organisations related to soccer, so that they may understand the sport's scientific foundations, will contribute to soccer's development in Turkey, especially with regard to the Turkey Soccer Federation, including soccer clubs, club managers, and soccer coaches. Furthermore, conducting scientific studies on soccer players will provide targeted support pertaining to specific aspects and is ultimately expected to contribute to the development of soccer in Turkey.

CONCLUSION
Our study is far from drawing the general framework for success in soccer. The reason for this is the limited statistical data used, as well as the existence of many other parameters that affect success in soccer, such as the number of club fans, the home team advantage, weather conditions, club finances, etc.

In order to be successful in soccer, both as a club and as a country, detailed research and analysis of the abovementioned features using scientific methods can provide useful results to understand the barriers to the development of soccer in Turkey. It is thought that researching the leagues in countries that are successful in soccer using similar methods and comparing them with Turkish soccer may provide useful information.

The Turkey Super League analogs in this study also apply to the lower leagues, contributing to the understanding of the importance hierarchy with regard to statistical parameters

It can therefore be concluded that the Turkey Super League teams who want to succeed should work on enhancing their technical features, as it is these features that are thought to have contributed to the last three seasons' success.

REFERENCES
1. Güngör, A. (2014). "The Relationship Between Financial Performance and Sporting Success in the Soccer Industry Analysis and Application Turkey." Journal of Social Science(1): 16-36.
2. Karamatov, O., et al. (2019). "soccer Economy: Example of Kastamonu University." Journal of Bartın University Faculty of Economics and Administrative Sciences **10**(20): 238-264.
3. Marchiori, M. and M. de Vecchi (2020). "Secrets of soccer: Neural network flows and game performance." Computers & Electrical Engineering **81**: 106505.
4. Reep, C. and B. Benjamin (1968). "Skill and chance in association football." Journal of the Royal Statistical Society. Series A (General) **131**(4): 581-585.
5. Armatas, V., et al. (2007). "Relationship between time and goal scoring in soccer games: Analysis of three World Cups." International Journal of Performance Analysis in Sport **7**(2): 48-58.
6. Armatas, V., et al. (2007). "Relationship between time and goal scoring in soccer games: Analysis of three World Cups." International Journal of Performance Analysis in Sport **7**(2): 48-58.
7. Bate, R. (1988). "Football chance: tactics and strategy." Science and football.
8. Gómez, M.-Á., et al. (2018). "Analysis of playing styles according to team quality and match location in Greek professional soccer." International Journal of Performance Analysis in Sport **18**(6): 986-997.
9. Heuer, A. and O. Rubner (2012). "How does the past of a soccer match influence its future? Concepts and statistical analysis." PloS one **7**(11): e47678.
10. Lago-Peñas, C., et al. (2017). "Styles of play in professional soccer: an approach of the Chinese Soccer Super League." International Journal of Performance Analysis in Sport **17**(6): 1073-1084.
11. Pappalardo, L. and P. Cintia (2018). "Quantifying the relation between performance and success in soccer." Advances in Complex Systems **21**(03n04): 1750014.
12. Yue, Z., et al. (2014). "Statistical analysis for the soccer matches of the first Bundesliga." International Journal of Sports Science & Coaching **9**(3): 553-560.
13. Liu, H., et al. (2013). "Inter-operator reliability of live football match statistics from OPTA Sportsdata." International Journal of Performance Analysis in Sport **13**(3): 803-821.
14. Günay, M. and A. İ. Yüce (2008). The Basis of Soccer Sciences, Gazi Kitabevi
15. Giulianotti, R. (2012). "Football." The Wiley-Blackwell Encyclopedia of Globalization.
16. Lago-Peñas, C., et al. (2010). "Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league." Journal of sports science & medicine **9**(2): 288.
17. Grant, A., et al. (1999). "An analysis of the successful and unsuccessful teams in the 1998 World Cup." Journal of Sports Sciences **17**(10): 827.
18. Hook, C. and M. Hughes (2001). "Patterns of play leading to shots in Euro 2000." Pass. com **1**: 295-302.
19. Jones, P., et al. (2004). "Possession as a performance indicator in soccer." International Journal of Performance Analysis in Sport **4**(1): 98-102.
20. Yue, Z., et al. (2014). "Statistical analysis for the soccer matches of the first Bundesliga." International Journal of Sports Science & Coaching **9**(3): 553-560.
21. Hughes, M. and I. Franks (2005). "Analysis of passing sequences, shots and goals in soccer." Journal of Sports Sciences **23**(5): 509-514.
22. Lago, C. and R. Martín (2007). "Determinants of possession of the ball in soccer." Journal of Sports Sciences **25**(9): 969-974.
23. Lago, C. (2007). "Are winners different from losers? Performance and chance in the FIFA World Cup Germany 2006." International Journal of Performance Analysis in Sport **7**(2): 36-47.
24. Rein, R. and D. Memmert (2016). "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science." SpringerPlus **5**(1): 1-13.
25. Mao, L., et al. (2016). "Identifying keys to win in the Chinese professional soccer league." International Journal of Performance Analysis in Sport **16**(3): 935-947.