A STUDY ON MATHEMATICAL MODEL OF DETERMINING THREE ISTANBUL FOOTBALL CLUBS WINNING OR LOSING[12]

Mehmet YILMAZ[1]

Hüseyin ÜNÖZKAN[2]

ABSTRACT

Todays, football becomes a very big industry and people's noticeable payings in their house or in stadiums to follow matches, takes an acceptable status. Important clubs uses stadium incomes, broadcast incomes, advertisemant sponsorship and some investment means for earning the money they need for their success, in order to gain these incomes they have to be succesfull in some certain rates. In recent terms, especially Europes important leagues trademark values declaring with billions of dollars, get the teams, which want to take the money in this sector, working more carefull, with using science, criteria like footballers health, during match analyses, following footballers living circumstances becomes to take attendance. The aim in this study is to create a healthfull mathematical model of gaining or losing matches of Beşiktaş, Fenerbahçe and Galatasaray who compete in Turkish Super League by using their matches published statistics. A logistic regression model is used to predict the outcome (winning or losing) of three valuable football club matches. We use some accessible 26 explanatory variables such as first goal time (home/away), number of scores (home/away), ball possession, correct pass number, etc. Furthermore, we added a weight variable that contains the total score of the last three weeks, both in order to prevent high correlation amongst variables and in order to obtain simpler model.

Key words: Logistic regression, Binary logistic regression, analyze of football matches, Turkish Super League.

ÜÇ İSTANBUL KULÜBÜNÜN KAZANMASI VEYA KAYBETMESİNİ BELİRLEYEN MATEMATİKSEL MODEL ÜZERİNE BİR ÇALIŞMA

ÖZET

Günümüzde futbol çok büyük bir endüstri haline gelmiş ve insanların maçları takip etmek için stadyumlarda ya da evlerinde belirli miktar parayı ödemeleri kabul edilir bir hal almıştır. Önemli takımlar başarıları için ihtiyaç duydukları parayı kazanmak maksadıyla stat gelirleri, yayın gelirleri, reklam sponsorları ve bazı yatırım araçlarını kullanmakta, bu gelirleri elde etmek için de belli oranlarda başarılı olmak durumunda kalmıştır. Son dönemlerde özellikle avrupadaki önemli liglerin marka değerinin artık milyar dolarla ifade edilmeleri, bu sektördeki parayı elde etmek isteyen takımları daha titiz çalışmaya itmiş, bilimden de yararlanarak futbolcu sağlığı, maç içi analizler, futbolcuların yaşam koşullarının takibi gibi kriterler dikkat edilir hale gelmiştir. Bu çalışmadaki amaç da Türkiye Süper Lig de mücadele eden Beşiktaş, Fenerbahçe ve Galatasaray futbol kulüplerinin maçlardaki yayınlanan istatistiklerinden faydalanarak, kazanma ve kaybetmeleri üzerinde etkili olan değişkenleri içeren sağlıklı bir matematiksel model oluşturabilmektir. Üç değerli futbol klubünün maç sonuçlarını (kazanma/kaybetme) kestirebilmek için lojistik regresyon modeli kullanılmıştır. İlk atılan gol zamanı, gol sayısı, topla oynama sayısı, doğru pas sayısı gibi ulaşılabilen 26 açıklayıcı değişken kullanılmıştır. Ayrıca, hem açıklayıcı değişkenler arasındaki yüksek korelasyonu engelleyebilmek hem de modelin daha sade olabilmesi için son üç haftadaki toplam puanı içeren bir ağırlık değişkeni eklendi.

Anahtar Kelimeler: Lojistik regresyon, İkili lojistik regresyon, Futbol maçlarının analizi, Türkiye Süper Ligi.

[1] Ankara University, Science Faculty, Department of Statistics, TURKEY.
E-mail: yilmazm@science.ankara.edu.tr
[2] Ankara University, Graduate School of Natural and Applied Science, Department of Statistics, TURKEY.
E-mail:hunozkan@gmail.com

INTRODUCTION

Because football becomes to absorb more money and lots of different sectors like bets has some activites relies on football industry, provides doing more scientific searchs and using every branch of sciences in football. Some decades before, in lots of football clubs only 3 or 4 workers were working, but todays analyze teams have at least the same number of employees. Therefore internet sites which are about football and bets, produce basic statistical data of past games before matches and publish some statistical data through matches.

In recent years these data have not been enough for success which football clubs are looking for. So clubs follow their footballers' exercises permanently through distinct health criters. But clubs apprehend that only doing good exercises do not clear for gaining and they begin to define some specific criters connect winning during matches to adjust their exercise programs. In these days every country's football leagues can be followed nearly in every country, so all teams may perform good strength and running much. On account of this, some important differences like qualified running distance, efficiency in penalty area, number of shots on target gets the match to gain or lose.

As clubs have only 90 minutes during matches and they have to use this time extremely efficient and prolific, things they have to do during matches be planned before games and primirally controlling these things during matches provides significant earnings.

At first times in collecting data the number of goals and assists, took attendance and goal keepers classified by goal numbers in their goal post. As time passed, criters adjusted to running distance for footballers and conformity with defence members for goal keepers. In these days, some other data are more important like footballers' stealing ball number, correct pass number, running distance with and without ball, how much time of this is with high speed, shot on target, correct pass on offense.

In 2013-2014 season at the beginning lots of statistical data were published and many writers considered the most important of these are running distances, team width and length, and ball possession rate. After some matches the Publisher release some data as well as total running distance of teams match by match. Because these data are only from three İstanbul clubs' matches and about the leader players, these data are not enough for analyze but their efficiency in match results are clear indeed.

Consequently, as which of variables during matches are influential is in debate, total distance with ball, total sprint number, total sprint number with ball, total sprint number without ball, total sprint distance with ball, total sprint distance without ball, total sprint number to rival post, sprints number of zones (left side,right side and center), total running distance of team, midfielders' total running distance, left and rigt sides players total running distance, first goal time, first goal time(rival), number of scores, rivals' number of scores, ball possession, correct pass, playing ball in offence area, correct pass in penalty area in offence, total shot, shot on target, total pass in penalty area in offence, using rate of offence area, using rate of defence area, distance to own post, long ball, centre, correct number of centre, percentage of playing with pass, ball in offence, correct ball in offence, free-kick, corner, distance to contact with rival, area using rates and field can consider as influential before

an analyze. According to thinking of some other variables can affect the result, teams last match results which show condition can be important, either.

Usually bets agencies decide rates through this approach using with markov chain in which the results of next match can takes place according to last three match results.

The originality of this study lies in the fact that it was tried to establish a model taking into account a large set of explanatory variables. Furthermore, to see the effect of the outcomes of the previous matches the team consecutively received in, a weight variable which is similar to Markovian structure has been added as an explanatory variable. In this way, adding weight variable has been effective way to overcome the potential multicollinearity problem probably occuring when the number of explanatory variables is too much in the model.

Previous works made by sport researchers based on match results have been generally related with The National Basketball Association (NBA). Mostly, logistic regression is used in sport studies. But there have not been done any work about football in Turkey. In addition to this, logistic regression was used in Turkey in many studies like deciding illness, guilt and guilty analyze, pleasure of passenger transportation and laboratory results. Logistic regression usage in world is also common and frequently seen in studies about sport injuries.

Some literature review is given as follows. Goddard and Asimakopoulos (2004) used ordered-probit model to predict the outcome of English soccer matches. Harville's approach relied on a much more complex system of linear models to predict the winner of an American Football match (Harville, 1980). To improve accuracy of logistic model Boulier and Stekler (2003) added power rankinks at the end of the each season. Willoughby (2002) used to predict winning in Canadian football games. Magel and Yana (2014) suggest a logistic regression model for predicting on soccer games played by three top European soccer leagues.

MATERIAL AND METHODS

The basic material of this study is data from 177 matches of Beşiktaş, Fenerbahçe and Galatasaray teams which were played in 2012-2013 and 2013-2014 seasons. This number is total of 59 matches for every team, 34 matches are from 2012-2013 season and the other 25 matches from 2013-2014. In these matches these variables were used: first goal time, first goal time(rival), number of scores, rivals' number of scores, ball possession, correct pass number, playing ball in offence area, correct pass in penalty area, total shot, shot on target, total pass in penalty area, using rate of offence area, using rate of defence area, distance to own post, long ball, number of centre, correct number of centre, percentage of playing with pass, ball in offence, correct ball number in offence, free-kick, corner, distance to contact with rival, area using rates, field. We also added a weight variable which shows team's total point in last three weeks.

We wanted to use some more variables in this study but because of some difficulties –especially some data needs paying to access – we could not. These variables are: total distance with ball, total sprint number, total sprint number with ball, total sprint number without ball, total sprint distance with ball, total sprint distance without ball, total sprint number to rival

post, sprints number of zones (left side, right side and center), total running distance of team, midfielders' total running distance, left and rigt sides players total running distance.

As we use 26 different variables and because statistical analyzing about match results has not studied yet, and since teams can gain only 0,1 or 3 points, make us to use logistic regression analyses.

In recent years, logistic regression becomes commonly popular. This method is an alternative to linear regression when two classed or multiple classed random variables disrupt normality hypothesis. Because of lacking of this hypothesis, flexible mathematical model and easily explainable, this method is usable (Özdamar, 1999).

In linear regression when Y (explained) defined as indicator variable taking two values (0,1), expected value of errors (*e i*) is zero, $E(e\ i) = 0$ and variance is constant, $Var(ei) = \sigma$.

As a result of this when we can not use hypothesis the estimates will not be best linear unbiassed estimator. This deficiency handles using linear regression in classifying analyses (Özdinç, 1999).

Through this, logistic regression considered as an appropriate method when other regression methods can not provides hypothesis and used frequently in classifying analyses. Logistic regression gives advantage in these applications because it does not require multivariate normal distribution hypothesis.

In logistic regression, like linear regression the aim is making a prediction about explained variable through some explanatory variables values. But there is 3 important differences between two methods (Elhan, 1977).

• In linear regression the explained variable is continuous, but this is discrete in logistic regression.

• In linear regression appropriate model is trying to predict explained variables value, but in logistic regression it is predicting the probability of the value which explained variable can take.

• In linear regression explanatory variable must be normally distributed, but there is not any necessity in logistic regression.

Logistic regression has 3 methods (Tatlıdil 1996). These are Binary Logistic Regression, Ordinary Regression, Nominal Logistic Regression. In this study, Binary Logistic Regression is used.

In Binary Logistic Regression if success probability to failure probability rate is expressed as odds rate, we reach;

$$p(X_1,\ldots,X_n) = p(X) = \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n) / \left(1 + \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)\right)$$

$$\text{Odds} = p(X) / 1 - p(X)$$

$$Odds = \frac{\exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n) / \left(1 + \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)\right)}{\left(1 / \left(1 + \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)\right)\right)}$$

$$= \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)$$

(Devore, 2004)

RESULTS

At first, data we need in this study were taken from MatchStudy.com. These data are 26 different data sets from 177 matches of Fenerbahçe, Galatasaray and Beşiktaş teams which were played in 2012-2013 and 2013-2014 seasons. This number is total of 59 matches for every team, 34 matches are from 2012-2013 season and the other 25 for every team are the first 25 matches from 2013-2014. Data are; first goal time(X1), first goal time(rival)(X2), number of scores(X3), rivals' number of scores(X4), ball possession(X5), correct pass(X6), playing ball in offence area(X7), correct pass in penalty area in offence(X8), total shot(X9), shot on target(X10), total pass in penalty area in offence(X11), using rate of offence area(X12), using rate of defence area(X13), distance to own post(X14), long ball(X15), centre(X16), correct number of centre(X17), percentage of playing with pass(X18), ball in offence(X19), correct ball in offence(X20), free-kick(X21), corner(X22), distance to contact rival(X23), weight(X24), area using rates(X25), field(X26) takes 1 if team plays home, 2 if plays out. Samples about some of data can be seen in Table 1 and Table 2.

Table 1. Some data samples

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
32	0	3	0	54	393	168	21	22	8	45	26,13	25
45	7	2	1	57	387	202	21	18	8	57	30,15	19,4
0	74	0	2	53	313	102	9	9	1	31	18,11	32,12
13	0	3	0	54	426	164	14	19	7	31	27,02	23,52
49	41	1	3	53	360	218	21	16	6	54	32,06	17,94
25	18	2	1	59	480	221	11	9	4	39	27,63	21,13
11	90	2	1	60	448	192	18	19	9	52	25,76	26,18
44	20	4	1	53	376	193	20	19	8	44	28,64	23,59
5	19	2	1	59	341	196	22	16	6	65	30,25	23,15
23	10	1	2	47	276	118	7	12	5	35	22,22	24,67
75	38	1	3	55	340	192	16	15	6	52	27,87	25,54
66	42	2	1	61	444	236	21	15	7	53	30,93	19,53
55	12	1	2	60	385	255	33	21	9	82	33,51	17,08
45	0	1	0	51	348	177	20	13	7	52	26,58	24,62
11	0	3	0	52	434	117	13	11	7	29	16,41	28,89
60	2	3	1	60	378	271	28	25	12	75	37,33	21,49
24	40	2	3	52	341	168	22	14	4	44	25,07	30
12	2	4	1	51	412	176	23	12	8	49	25,36	25,79
11	32	2	1	43	251	137	12	12	5	31	24,82	25,91
47	0	2	0	53	394	162	18	17	8	45	23,55	25,29
37	0	2	0	52	331	107	11	7	5	25	17,04	37,26
52	0	1	0	46	333	141	10	11	1	29	23,31	23,97
0	5	0	2	58	467	239	27	25	11	84	29,04	17,98
58	6	2	1	59	408	243	16	13	5	65	35,12	18,64

Table 2. Some data samples (continued)

X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26
57,15	32	12	5	68,9	94	80	4	5	40,6	5	1,08	1
60,74	35	18	2	65,07	95	75	7	12	49,56	5	1,38	1
50,41	36	5	1	60,7	59	45	2	6	38,68	5	0,81	2
55,49	33	20	4	65,42	90	78	2	5	44,54	4	1,08	1
61,84	44	29	5	58,53	100	91	5	4	50,51	4	1,38	1
57,04	31	21	4	66,25	141	123	4	2	44,2	4	1,22	2
56,43	57	18	3	67,5	112	88	2	9	45,21	4	1,08	1
59,37	38	19	4	62,61	108	98	2	6	46,58	7	1,17	1
60,23	41	31	5	58,8	99	84	1	16	45,48	5	1,27	1
54,19	26	10	0	59,51	54	42	3	3	42,54	7	1	2
57,06	27	24	7	56,89	85	69	5	4	49,18	4	1,08	1
59,97	33	20	3	65,4	147	126	3	4	49,6	1	1,5	2
63,37	33	34	11	57,82	111	86	4	17	52,82	4	1,5	1
57,13	41	17	4	61,26	103	77	2	6	45,37	4	1,27	2
52,47	48	10	2	67,32	69	60	3	6	40,72	6	0,92	2
61,98	34	31	5	58,54	132	110	4	11	52,56	6	1,5	1
54,72	22	23	8	59,4	88	70	3	8	40,02	9	1	2
57,59	42	21	4	65,99	98	84	0	3	42,36	6	1,12	1
54,76	28	5	4	54,71	70	53	2	5	41,51	6	1,12	2
55,7	23	12	4	64,1	96	79	0	3	42,17	6	0,92	1
49,25	43	11	3	61,31	56	45	1	3	36,81	9	0,66	2
56,16	37	14	2	64,63	77	58	2	2	46,23	9	1,12	1
59,35	38	31	7	64,28	118	91	1	12	48,75	9	2,33	2
63,2	37	39	4	68,5	130	104	7	10	51,32	6	1,5	1

In this study, we used one of the two methods in binary logistic regression in which adding variables one by one and repeat analyses for every new variable. Every combination of 26 variables were studied. Logistic regression highly influenced by multicollinearity and as we saw during the analyze that lots of variables had a linearly dependent with each other. To overcome multicollinearity problem, two methods are used. First one is that selection of explanatory variables including model is made by stepwise regression. After that some variables which show significantly large individiual p values (sig.) are excluded from model. Second is that suggested weight variable is being used.

Since the match outcomes are rarely tied this would make us to predicting matches easier through simply selecting the winning or losing. In this way, binary logistic regression will be used to improve on accuracy of outcome of next matches. According to this we put 1 for win and 0 for not win. The results of the most significant models we could find about 3 Istanbul teams' 177 matches are these:

Results of model 1

To measure goodness of the predicted outcomes we obtained the frequencies of winning or losing according to predicted probabilities depending on whether above or below the value of 0.5. Then we generated a table of observed versus predicted outcomes.

Table 3. Classification Table for Model 1

Observed			Predicted		
			POINT		Percentage Correct
			,00	1,00	
Step 1	POINT	,00	57	22	72,2
		1,00	18	80	81,6
	Overall Percentage				77,4

Model 1 produces a prediction accuracy of %77,4 (57+80 out of 177). Table 4 indicates the results of the logistic regression analysis. Variable coefficients are shown in second column (B), standart errors of the coefficient estimations in third column (S.E.), individual p values in fourth column (Sig.), the change in the log odds of team victory for a one unit increase in the individual explanatory variable (Exp(B)).

Table 4. Variables in Model 1

	Variable Names	B	S.E.	Sig.	Exp(B)
Step 1[a]	FIELD	-1,401	,369	,000	,246
	DISTANCE TO CONTACT RIVAL	,109	,033	,001	1,115
	FIRST GOAL TIME	,024	,009	,006	1,024
	FIRST GOAL TIME(RIVAL)	-,018	,007	,008	,982
	PLAYING BALL IN OFFENCE AREA	-,089	,017	,000	,915
	SHOT ON TARGET	,370	,083	,000	1,448
	CORRECT PASS IN PENALTY AREA IN OFFENCE	,133	,027	,000	1,142

As seen from Table 4, model 1 includes with above variables which predicts outcome with accuracy rate of %77,4 of every matches. Statistically significant (p values, Sig.<0.05) explanatory variable coefficients of model1 are field, distance to contact rival, first goal time, first goal time (rival), playing ball in offence area, shot on target, correct pass in penalty area in offence. Multicollinearity does not appear to be a problem for Model 1.

Results about model 2 included with a constant variable are followed:

Results of model 2

Table 5. Classification Table for Model 2

Observed			Predicted		
			POINT		Percentage Correct
			,00	1,00	
Step 1	POINT	,00	57	22	72,2
		1,00	17	81	82,7
	Overall Percentage				78,0

Model 2 produces a prediction accuracy of %78 (57+81 out of 177).

Table 6. Variables in model 2

		B	S.E.	Sig.	Exp(B)
Step 1[a]	FIELD	-1,615	,434	,000	,199
	FIRST GOAL TIME	,024	,009	,006	1,024
	FIRST GOAL TIME(RIVAL)	-,019	,007	,007	,981
	PLAYING BALL IN OFFENCE AREA	-,073	,014	,000	,930
	SHOT ON TARGET	,362	,083	,000	1,436
	CORRECT PASS IN PENALTY AREA IN OFFENCE	,118	,025	,000	1,125
	CONSTANT	3,706	1,248	,003	40,705

As seen from Table 6, the variables in model 2, which predicts with correct rate of %78 of every match, are field, first goal time, first goal time (rival), playing ball in offence area, shot on target, correct pass in penalty area in offence and constant variable. All coefficients in model 2 is statistically significant (p values, Sig.<0.05) and multicollinearity does not appear to be a problem for Model 2.

Because magnitude of correction rate of prediction was not enough for us, we considered that we had used 46 matches as lost but they actually had been drawn. So, we removed 46 matches and analyzed again. The results of model 3 are in Tables 7 and 8.

Results of model 3

Table 7. Classification Table for model 3

Observed			Predicted		
			POINT		Percentage Correct
			,00	1,00	
Step 1	POINT	,00	24	9	72,7
		1,00	5	93	94,9
	Overall Percentage				89,3

Model 3 produces a prediction accuracy of %89,3 (24+93 out of 131).

Table 8. Variables in model 3

		B	S.E.	Sig.	Exp(B)
Step 1[a]	DISTANCE TO OWN POST	,116	,039	,003	1,123
	FIELD	-1,940	,613	,002	,144
	CORRECT PASS IN PENALTY AREA IN OFFENCE	,156	,040	,000	1,168
	SHOT ON TARGET	,533	,136	,000	1,704
	PLAYING BALL IN OFFENCE AREA	-,103	,024	,000	,902
	FIRST GOAL TIME	,029	,013	,032	1,029
	FIRST GOAL TIME(RIVAL)	-,022	,011	,043	,978

As seen from Table 8, the variables in model 3, which predicts with correct rate of %89,3 of every matches, are field, first goal time, first goal time (rival), playing ball in offence area, shot on target, correct pass in penalty area in offence and distance to own post. All coefficients in model 3 is statistically significant (p values, Sig.<0.05) and multicollinearity does not appear to be a problem for Model 3.

Best results of possible models, which contains weight variable, are in Tables 9 and 10.

Results of model 4

Table 9. Classification Table for model 4

Observed			Predicted		
			POINT		Percentage Correct
			,00	1,00	
Step 1	POINT	,00	14	19	42,4
		1,00	5	93	94,9
	Overall Percentage				81,7

Model 4 produces a prediction accuracy of %81,7 (14+93 out of 131).

Table 10. Variables in model 4

		B	S.E.	Sig.	Exp(B)
Step 1a	WEIGHT	-,238	,076	,002	,788
	SHOT ON TARGET	,309	,072	,000	1,362
	FIRST GOAL TIME	,024	,011	,022	1,025

Model 4 predicts with correct rate of %81,7 of every match. Explanatory variables in model 4 are weight, first goal time and shot on target. All coefficients in model 4 is statistically significant (p values, Sig.<0.05) and multicollinearity does not appear to be a problem for Model 4. According to other models, this model is constructed with a smaller number of explanatory variables it is extremely compact and understable.

DISCUSSION AND CONCLUSIONS

After these applications most appropriate one for us is model 3 with %89,3 correction rate of 131 matches. Explanatory variables are field, first goal time, first goal time (rival), playing ball in offence area, shot on target, correct pass in penalty area in offence and distance to own post. If someone make a prediction on outcomes of next matches with using simple and compact model, then we shall suggest model 4 (includes only 3 explanatory variables) which is the second appropriate model according to it's accuracy rate.

Through results in model 3:

Distance to own post increases gaining 1,123 times in every meter. This variables' coefficient in model is 0,116 and exp(0,116) is 1,123 which means 1,123 times in every meter the team distance from own post get advantage.

Field variable get 1 if team plays home and 2 if plays out. Variables' coefficient in model is -1,94. According to this, if team plays home exp(-1,94) is 0,144 and if team plays out exp(-1,94*2) is 0,02. This means if team plays out gaining will be more difficult.

Correct pass in penalty area in offence variable increases gaining 1,168 times in every correct pass. This variables' coefficient in model is 0,156 and exp(0,156) is 1,168 which means 1,168 times in every correct pass the team do in rivals penalty area get advantage to win.

Shot on target variable increases gaining 1,704 times in every correct shot. This variables' coefficient in model is 0,533 and exp(0,533) is 1,704 which means 1,704 times in every correct shot the team get advantage to win.

Playing ball in offence area variable coefficient is -0,902. This means it influences gaining badly. This variables' coefficient is negative like field variable so make it difficult to win. As if playing offence were a good thing for team, this is a sign of two discrepancies. First, playing more in offence shows team can not enter to rivals' dangerous zone. Second, team is losing or can not score, so team plays in rival area but can not successful in scoring.

First goal time increases gaining 1,029 times in every minute. This variables' coefficient in model is 0,029 and exp(0,029) is 1,029 which means 1,029 times in every minutes get advantage to win. Although, someone thinks scoring early can get advantage, scoring early takes time to rival for reaction and changing score.

First goal time(rival) variable influences badly with -0,978. Like field and playing ball in offence area variables, this variables' coefficient is negative, so this influence winning badly. With this variable we can see that whenever the first goal of the rival scores is early, team has time to change score.

We can understand from this study that teams have to increase their shots on target and correct pass in dangerous area in offence if they want to increase winning probability.

According to model 3, we can predict coming weeks matches, which did not use in study, through during match data.

26. Week Beşiktaş 3-0 Akhisar Belediye
=e^(55,68*0,116+1*(-1,940)+97*0,156+9*0,533+175*(-0,103)+2*0,029+0*(-0,022))=e^6,48
Rate of gaining probability to losing probability is e^6,48=651,97, and gaining probability is 651,97/(1+651,97)=0,998468. Model estimates correctly.

26. Week Galatasaray 0-1 Kayseri
e^(59,57*0,116+1*(-1,940)+75*0,156+4*0,533+194*(-0,103)+0*0,029+90*(-0,022))=e^-3,16
Rate of gaining probability to losing probability is e^-3,16=0,0424, and gaining probability is 0,0424/(1+0,0424)=0,040675. Model estimates correctly.

27. Week Fenerbahçe 3-0 Bursaspor
e^(57,26*0,116+1*(-1,940)+76*0,156+14*0,533+174*(-0,103)+17*0,029+0*(-0,022))=e^6,59
Rate of gaining probability to losing probability is e^6,59=727,78, and gaining probability is 727,78/(1+727,78)=0,998627. Model estimates correctly.

28. Week Beşiktaş 2-1 Kayserispor
e^(51,83*0,116+1*(-1,940)+58*0,156+8*0,533+128*(-0,103)+43*0,029+90*(-0,022))=e^3,46
Rate of gaining probability to losing probability is e^3,46=31,8169, and gaining probability is 31,8169/(1+31,8169)=0,969527. Model estimates correctly.

27. Week Karabükspor 1-0 Beşiktaş
=e^(54,23*0,116+2*(-1,940)+62*0,156+4*0,533+150*(-0,103)+0*0,029+90*(-0,022))=e^-3,21
Rate of gaining probability to losing probability is e^-3,21=0,0403, and gaining probability is 0,0403/(1+0,0403)=0,038738. Model estimates correctly.
But how can Beşiktaş win Karabükspor in outside?

If Beşiktaş passed 70 instead of 62 correctly in penalty area in offence and shot 8 instead of 4 correctly,then rate of gaining probability to losing probability would be,
e^(54,23*0,116+2*(-1,940)+70*0,156+8*0,533+150*(-0,103)+0*0,029+90*(-0,022))=e^0,164 =1,179. Thus gaining probability would be 1,179/(1+1,179)=0,541 and probability of gain would be more than %50.

Finally we additionally say that we want to use more explanatory variables in modeling process such as current forms of most important players,

players injured, team value, fans, weather conditions etc but these data sets could not obtained without payments. Also, we say that working with more explanatory variables can lead to more serious multicollinearity problems.

Acknowledgment The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

1. Boulier, B.L. and Stekler, H.O. “Predicting the outcomes of national football league games”, International Journal of Forecasting, 19(2), 257–270, 2003.
2. Cook D., Dixon P., Duckworth W., Kaiser M.S., Koehler K., Meeker W.Q. and Stephenson W.R. “NFS/ILI project Beyond Traditional Statistical Methods”. Iowa State University, 2001.
3. Devore J.L. Probability and statistics for engineering and sciences. Sixth edition. California Polytechnic State University, 2004.
4. Elhan A.H. “Lojistik regresyon analizinin incelenmesi ve tıpta bir uygulama”, Yüksek Lisans tezi, Ankara Üniversitesi, 4-29, Ankara, 1997. [In Turkish]
5. Goddard, J. and Asimakopoulos, I. “Forecasting football results and the efficiency of fixed-odds betting”, Journal of Forecasting, 23 (1), 51–66, 2004.
6. Harville, D. “Predictions for national football league games via linear-model methodology”, Journal of the American Statistical Association, 75(371), 516–524, 1980.
7. Magel, R., and Yana M. "Examining Influential Factors and Predicting Outcomes in European Soccer Games." International Journal of Sports Science 4.3, 91-96, 2014.
8. Özdamar K. Paket programlar ile istatistiksel veri analizi. 1-2. Kaan Kitapevi, Eskişehir, 1999. [In Turkish]
9. Özdinç Ö. “Derecelendirme sürecinde ekonometrik bir değerlendirme”. SPK Yayını, No:130, Ankara, 1999. [In Turkish]
10. Tatlıdil H. Uygulamalı çok değişkenli istatistiksel analiz. Cem Web Ofset, Ankara, 1996. [In Turkish]
11. Willoughby, Keith A. "Winning games in Canadian football: a logistic regression analysis." College Mathematics Journal 33.3, 215-220, 2002.