GSJ: Volume 7, Issue 12, December 2019, Online: ISSN 2320-9186
www.globalscientificjournal.com

European Football Leagues Support among Undergraduate Students in Nigeria: A Multinomial Logistic Regression Approach

Babalola Bayowa Teniola

Department of Statistics, Ekiti State University, Ado Ekiti, Nigeria

teniola.babalola@eksu.edu.ng **+2348067525251**

ABSTRACT

Objective: This work researches into some popular European football leagues Nigerian students love watching. This includes the English Premier League, Serie A, Bundesliga and La-liga. Some factors that tend to influence their love for these leagues were examined.

Methodology: The data for this research were obtained using a questionnaire completed by some Nigerian undergraduate students. The Multinomial Logistic regression was employed.

Result: A multinomial logistic regression model was fitted . From the analysis, it was observed that factors like Viewing Access, Jersey, world class footballer, all have positive coefficients. The result revealed that English Premier League teams are the most supported by undergraduate students in the Nigeria.

Keywords: Football; Leagues; Multinomial logistic regression; Statistics.

INTRODUCTION

In recent years, the various leagues across Europe like English Premier League, Spanish La-Liga, French Ligue 1, German Bundesliga have spent a lot of money on football through investing on quality players, managers, stadia and other equipment.[1,2,3]. Consequently, this investment has

led to a major boost in economies of the different countries. For example, as a result of premier league and club activities, over 103,354 full time equivalent jobs were supported in the United Kingdom in 2013/14. In March 2015, Premier League clubs also committed to adopting the living wage for all their full time employees by the start of the 2016/17 season at the latest. According to statistics, the major European leagues countries have the highest number of settling immigrants; this automatically leads to positive economic and regional impacts[2].

Club competition between teams from different European countries trace back their origin as far back 1897[4]. The challenge cup was founded as a competition between clubs of the Austro-Hungarian Empire. Following World War II, the reduced standing of the Mitopa cup led to foundation of a new competition, the Latin cup, for teams from France, Italy, Spain and Portugal. This competition was played as mini tournament which helped in creating peace and unity among the countries.

A focus delivering the best possible football competition, with compelling matches played in world class stadiums full of passionate fans has established the European leagues as the most watched continuous annual globe sporting event in the world. Sporting success and fans interest allows clubs to generate massive income. Consequently, most clubs at the end of each season declare their profits to the tune of millions of dollars. Clubs are then able to invest in players, facilities, academies and their communities.

Since the interest of Nigerian youths is rapidly growing in the sport, it is therefore worthy of research. For the purpose of this research, we shall be focusing our attention on four major European leagues which are: The English Premier League, German Bundesliga, Italian Serie A, and the Spanish La liga. This Multinomial logistic regression model has been applied in data analysis in many areas, for example health, social, behavioral,and educational sectors[5]. This study is aimed at the application of Multinomial Logistic Regression model to sport entertainment.

MATERIALS AND METHOD

The data was obtained through a well-structured questionnaire that was framed for the purpose of this research. Since the response of interest is categorical (polytomous) response variables, the Multinomial Logistic Regression model (MLR) was employed in the analysis of the

questionnaire. The observations are statistically independent of each other. The observations are a random sample from a population where y_i has a multinomial distribution.

The MLR model is generally effective where the response variable is composed of more than two levels or categories and the categories cannot be ordered (are nominal in nature). The basic concept was generated from binary logistic regression. Continuous variables are not used as response variable in logistic regression, and only one response variable can be used. MLR can be used to predict a response variable on the basis of continuous and/or categorical explanatory variables to determine the percent of variance in the response variable explained by the explanatory variables; to rank the relative importance of independents to assess interaction effects; and to understand the impact of covariate control variables [6,7]. MLR allows the simultaneous comparison of more than one contrast, that is, the log odds of three or more contrasts are estimated simultaneously [8,9]. The logistic regression model assumes that the categorical response variable has only two values, in general, 1 for success and 0 for failure. The logistic regression model can be extended to situations where the response variable has more than two values, and there is no natural ordering of the categories. The logistic regression was explained comprehensively. See([8,10]).

Natural ordering can be treated as nominal scale, such data can be analyzed by slightly modified methods used in dichotomous outcomes, this method is called the multinomial logistic. If we have *n* independent observations with *p*-explanatory variables, and the qualitative response variable has *k* categories, to construct the logits in the multinomial case, one of the categories must be considered the base level and all the logits are constructed relative to it. Any category can be taken as the base level, so we will take category *k* as the base level. Since there is no ordering, it is apparent that any category may be labeled *k*. Let πj denote the multinomial probability of an observation falling in the *jth* category, to find the relationship between this probability and the *p*-explanatory variables, $X_1, X_2, \ldots, X_P$, the multinomial logistic regression model is given by:

$$log\left[\frac{\pi_j(X_i)}{\pi_k(X_I)}\right] = \alpha_{0i} + \beta_{ij}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi}$$

where $j = 1, 2, \ldots, (k - 1)$, $i = 1, 2, \ldots, n$. Since all the π's add to unity, this reduces to

$$log\left(\pi_j(X_i)\right) = \frac{exp(\alpha_{0i} + \beta_{ij}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi})}{1 + \sum_{j=1}^{k-1} exp(\alpha_{0i} + \beta_{ij}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi})}$$

For j = 1,2, . . , *(k*-1*)*, the model parameters are estimated by the method of Maximum Likelihood.[11,12] .

Estimating multinomial response probabilities

The equation that expresses multinomial logit models directly in terms of response probabilities

$$\pi(x) = \frac{exp(\alpha_j + \beta_j' x)}{1 + \sum_{h=1}^{j-1} exp(\alpha_h + \beta_h' x)}$$

α_j = and 0 $\beta_j = 0$.

The dependent variable was the choice of league the student supports while the independent variable were: viewing access, the time matches are played (in Nigerian time),the clubs' jersey, the number of world class footballers in the leagues, the coaches, the pattern or style of play and the number of African players in the league.

RESULTS AND DISCUSSION

Here, we are interested in extracting the regression models from the result in the Table 2.

Logit for support Serie A rather than EPL is given by:

$L^{(\text{Serie a})} = \log(\pi^{serie\ A}/\pi^{EPL})$

$= \beta_0(serie\ A) + \beta_1(serie\ A)x_1(access) + \beta_2(serie\ A)x_2(time) + \beta_3(serie\ A)x_3(jersey) + \beta_4(serie\ A)x_4(world\ class\ footballer) + \beta_5(serie\ A)x_5(coach) + \beta_6(serie\ A)x_6(pattern) + \beta_7(serie\ A)x_7(african\ players)$

$=-2.962 + 3.620x_1 - 0.451x_2 + 1.042x_3 + 1.984x_4 + 0.575x_5 - 0.425x_6 + 1.148x_7$

Logit for supporting la-liga rather than EPL.

$L^{(\text{La-liga})} = \log(\pi^{la-liga}/\pi^{Epl})$

$= \beta_0(la-liga) + \beta_1(la-liga)x_1(access) + \beta_2(la-liga)x_2(time) + \beta_3(la-liga)x_3(jersey) + \beta_4(la-liga)x_4(world\ class\ football) + \beta_5(la-liga)x_5(coach) + \beta_6(la-liga)x_6(pattern) + \beta_7(la-liga)x_7(african\ players)$

$=-0.927+ 1.060x_1+ 0.396x_2 – 0.367x_3 + 0.613x_4 + 0{,}213x_5+ 0.662x_6 + 0.260x_7$

$L^{(bundesliga)} = \log(\pi^{bundesliga}/\pi^{Epl})$

$= \beta_0(bundesliga) + \beta_1(bundesliga)x_1(access) + \beta_2(bundesliga)x_2(time) + \beta_3(bundesliga)x_3(jersey) + \beta_4(bundesliga)x_4(world\ class\ football) + \beta_5(bundsliga)x_5(coach) + \beta_6(bundesliga)x_6(pattern) + \beta_7(bundesliga)x_7(african\ players)$

$=-2.873+ 3.483x_1 +0.205x_2 + 1.146x_3 + 1.187x_4 + 0.883x_5 – 0.190x_6 - 1.148x_7$

Predicted Probability For The Full Model

EPL = 0, Serie A = 1, La-Liga = 2, Bundesliga = 3

Probability of response in category

For serie A

$\pi^{(Serie\ A)}=P(Y=1) = \dfrac{\exp(l^{(Serie\ A)})}{1+\exp(l^{La\text{-}Liga})+\exp(l^{(Bundesliga)})}$

For La-Liga

$\pi^{(La-Liga)}=P(Y=2) = \dfrac{\exp(l^{(La\text{-}Liga)})}{1+\exp(l^{Serie\ a})+\exp(l^{Bundesliga)})}$

For Bundesliga

$\pi^{(Bundesliga)}=P(Y=3) = \dfrac{\exp(l^{(Bundesliga)})}{1+\exp(l^{La\text{-}Liga})_{+}\exp(l^{(Serie\ A)})}$

For the reference EPL

$\pi^{(EPL)}=P(Y=0) = \dfrac{1}{1+\exp(l^{Serie\ A})+\exp(l^{La\text{-}Liga})+\exp(^{Bundesliga})}$

$L^{(1)} = -2.962(1) +3.620(1) – 0.451(1) + 1.042(1) + 1.984(1) + 0.573(1) – 0.425(1) + 1.148(1)$

$= 4.529$

$L^{(2)} = -0.927 + 1.060(1) + 0.396(1) – 0.367(1) + 0.613(1) + 0.213(1) + 0.662(1) + 0.260(1)$

$= 2.644$

$L^{(3)}$= -2.873(1) + 3.483(1) + 0.205(1) + 1.146(1) + 1.187(1) + 0.883(1) – 0.190(1) -0.923(1)

= 2.918

P(Serie A) = exp(4.529)/1+exp(4.529)+exp(2.644)+exp(2.918)=0.43

P(La-liga)=exp(2.644)/1+exp(4.529)+exp(2.644)+exp(2.918)=0.11

P(Bundesliga)=exp(2.918)/1+exp(4.529)+exp(2.644)+exp(2.918)=0.15

Reference category EPL

P(EPL)=1/1+exp(4.529)+exp(2.644)+exp(2.918)=0.02

Table 1: Table showing the probabilities of supporting the different leagues under a given situation.

EUROPEAN LEAGUES	EPL	SERIE A	LA-LIGA	BUNDESLIGA
PROBABLITY	0.02	0.43	0.11	0.15

As revealed in Table 1, for a student who says yes to all the predictors, the probability of supporting Serie A is 0.43, also the probability of supporting LA LIGA is 0.11, Bundesliga is 0.15 and EPL is 0.02

Table 2: Parameter Estimates

	B	STD ERROR	Wald	Df	Sig	Exp(B)
FAVORITE LEAGUES						
SERIE A Intercept	-2.962	.388	58.398	1	.000	
{ACCESS=1}	3.620	1.249	8.405	1	.004	37.388
{TIME=1}	-.451	.634	.507	1	.476	.637
{JERSEY=1}	1.042	.538	3.758	1	.053	2.835
	1.984	.651	9.276	1	.002	7.273

{WORLDCLASS=1}	.575	.557	1.065	1	.302	1.777
{COACH=1}	-.425	.985	.197	1	.657	.654
{PATTERN=1}						
{AFRICANPLAYER=1}	1.148	.510	5.065	1	.024	3.150
LA-LIGA Intercept	-.927	.178	27.076	1	.000	
{ACCESS=1}	1.060	1.261	.707	1	.401	2.885
{TIME=1}	.396	.407	.946	1	.331	1.486
{JERSEY=1}	-.367	.519	.820	1	.365	.693
{WORLDCLASS=1}	.613	.375	1.393	1	.238	1.845
{COACH=1}	.231	.584	.379	1	.538	1.259
{PATTERN=1}	.613	.336	1.283	1	.257	.617
{AFRICANPLAYER=1}	.260	.336	.599	1	.439	1.298
BUNDESLIGA Intercept	-2.873	.383	56.423	1	.000	
{ACCESS=1}	3.483	1.241	7.871	1	.005	371.032
{TIME=1}	.205	.655	.097	1	.075	4.433
{JERSEY=1}	1.146	.597	3.678	1	.055	10.139
{WORLDCLASS=1}	1.187	.718	2.727	1	.099	13.392
{COACH=1}	.883	.596	2.200	1	.138	7.774
{PATTERN=1}	-.190	.873	.047	1	.828	4.575
{AFRICANPLAYER=1}	-.923	.811	1.297	1	.255	1.946

a. The reference category is EPL

b. This parameter is set to zero because it is redundant.

Holding other variables constant, the odds for someone who has an access to viewing Serie A constantly choosing Serie rather than EPL are 37.388 and the odds of someone who loves seeing African players play choosing Serie A rather than EPL are 3.150.

Holding other variables constant, the odds for someone who has an access to viewing Bundesliga constantly supporting a club in Bundesliga are 371.032 while World class players pulled the odds of supporting a club in Bundesliga to 13.392.

SUMMARY

As stated earlier, the data used in this research is a primary data gotten from the undergraduate's student of Ekiti State University, Ado-Ekiti, Nigeria. Multinomial logistic regression was applied on the data and some interesting facts were discovered. European football leagues which involve English Premier League, German Bundesliga, Spanish Primera division (LA-LIGA) and the Italian league (Serie A) were modeled based on seven (7) factors which are; Access to information, time of play, jersey, world class footballer, coach's charisma, pattern of play, and numbers of African player.

According to the questionnaire distributed, out of 300 students, 171 students support EPL, 84 students support LA LIGA, 25 students goes for SERIE A, while 9 students support Bundesliga. This shows that EPL is the most supported league among the undergraduates. The likelihood ratio test was to test the hypothesis that a subset of some explanatory variables can be dropped that is, the associated regression coefficient is equal zero or not. From the analysis, it was observed that factors like Viewing Access, Jersey, world class footballer, all have positive coefficients. This indicates that they all tend to favour the higher ranked categories which is EPL followed by LA LIGA.

According to the parameter estimates, based on the *p*-value the predictors variable that is significant in choosing SERIE A in relative to ENGLISH PREMIER LEAGUE are access, world class footballer and African players. Students see almost all the factors in La-Liga as well in EPL. While in Bundesliga in relative to EPL, access is significant. The odd ratio indicates that "a unit" increase in access, jersey, world class footballer, coach charisma, African footballer will increase the odd of supporting SERIE A rather than EPL. The odd ratio also indicates that "a unit" increase in time will increase the odds of supporting EPL rather than SERIE A.

The odd ratio indicates that a unit increase in access, time, world class footballer, coach charisma, Pattern, African footballer will increase the odd of supporting La-Liga rather than EPL. The odd ratio also indicates that a unit increase in Jersey will increase the odds of supporting EPL rather than La-liga. The odd ratio indicates that a unit increase in access, time, jersey, world class footballer, coach charisma, will increase the odd of supporting Bundesliga rather than EPL. The odd ratio also indicates that a unit increase in pattern, African player will increase the odds of supporting Bundesliga rather than EPL.

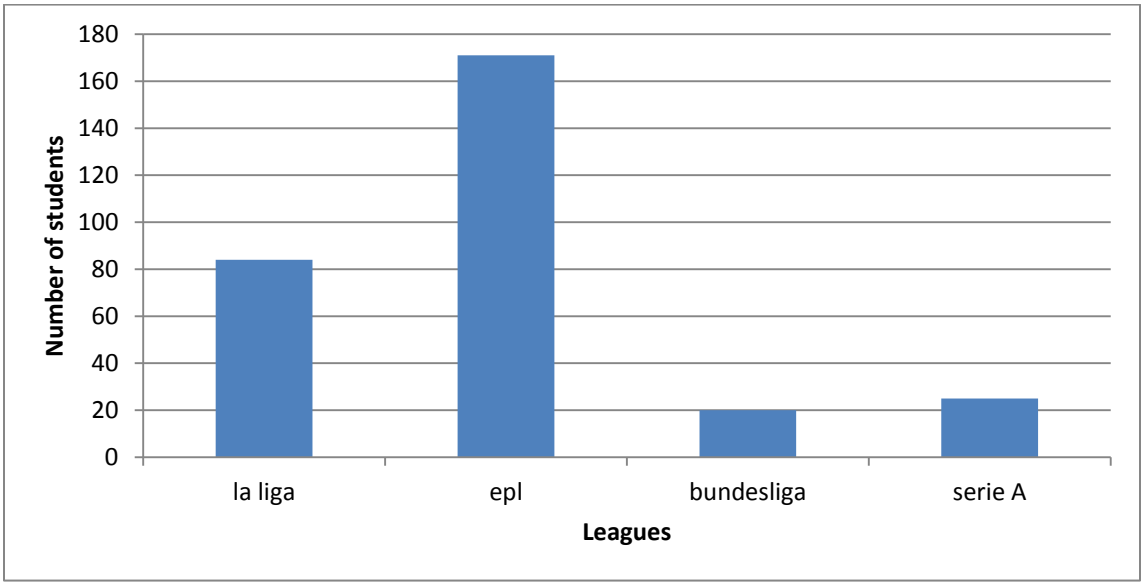

Fig 1.0: A bar chart of leagues supported by students of some European leagues.

This represents the bar chart of students supporting their favorite leagues. It shows that EPL is the most supported league, followed by La-Liga, Serie A and Bundesliga.

CONCLUSION

The fitted model above can be used to predict the different probabilities of supporting European Leagues Clubs to a certain degree given the conditions of the students involved. As discovered during the course of this study, some of the explanatory variables actually affect the support of a European Football league.

RECOMMENDATIONS

We recommend that Nigerian Government and Nigerian Professional Football League officials should look into these factors and try to improve all the factors especially access to viewing Nigerian Professional Football League, improve the stadia, and invite more elite coaches into the

leagues. This will make Nigerian students support Nigerian Professional Football League and this will consequently have a positive impact on the economy of the Country.

ACKNOWLEDGEMENT

We thank the Department of Statistics, Ekiti State University, Ado- Ekiti for the support received during the course of this research.

REFERENCES

1. Premier League Economic and Social impact(2019). https://www.ey.com/Publication/vwLUAssets/EY-Premier-League-economic-and-social-impact-January-2019/$FILE/EY-Premier-League-economic-and-social-impact-January-2019.pdf. Ernst & Young LLP. UK.
2. English Premier League season review (2015) http://review.premierleague.com/2014-15/the-premier-league/economic-impact.html
3. German soccer rules:50+1 explained (2018). https://www.bundesliga.com/en/news/Bundesliga/german-soccer-rules-50-1-fifty-plus-one-explained-466583.jsp
4. European Cup and Uefa Champions league History(2019). https://en.wikipedia.org/wiki/European_Cup_and_UEFA_Champions_League_history
5. Abdalla M.E. (2012). An application on Multinomial logistic regression model. Pak.j.stat.oper.res.8(2):271-291
6. Cho.Y.P and Rebecca.N.N (2002), "Using multinomial logit models to predict adolescent behavioral risk", Journal of Modern Applied Statistical Methods.2(1):1-12.
7. Mullender L (2005), "Tutorial multinomial logit analysis", www.mktgeng.com/.../tutorials/S005%20, %20Multinomial%20Logit%20Analysis%20Tutorial.p.
8. Garson,D.(2009).Logistic Regression with SPSS.North Carolina StateUniversity, Public administration Program.
9. Schafer J.L.(2006).Multinomial logistic regression models.STAT 544-Lecture 19.
10. Agresti, A (2007), "An Introduction to Categorical Data Analysis", John Wiley & Sons, Inc.
11. SchwabJ.,(2007).Multinomial Logistic Regression BasicRelationships.www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/Analyzi.
12. Benoit K.(2012). Multinomial and Ordinal Logistic Regression. ME104: Linear Regreesion Analysis.