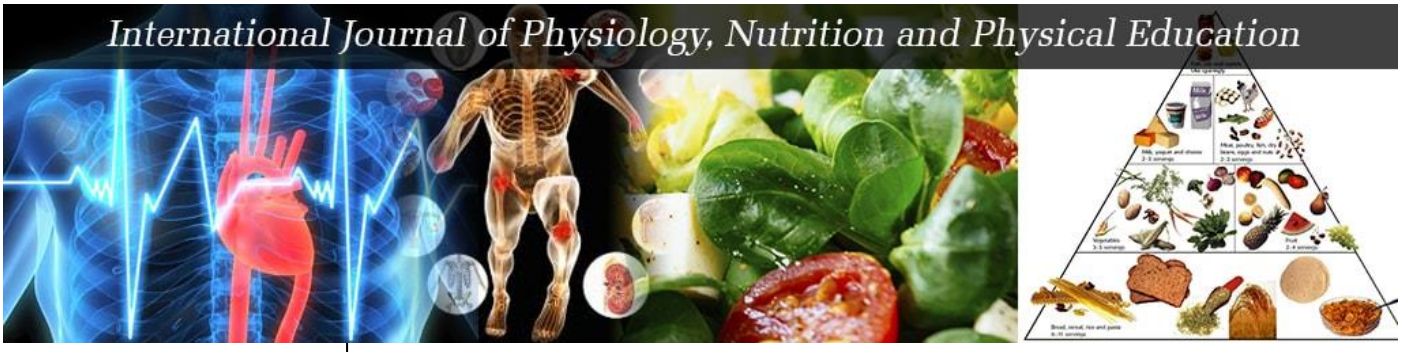

ISSN: 2456-0057
IJPNPE 2019; 4(1): 383-386

www.journalofsports.com
Received: 08-11-2018
Accepted: 12-12-2018

Amritashish Bagchi
Symbiosis School of Sports Sciences, Symbiosis International (Deemed University), Pune, Maharashtra, India

Shiny Raizada
Symbiosis School of Sports Sciences, Symbiosis International (Deemed University), Pune, Maharashtra, India

Aniket Mhatre
Symbiosis School of Sports Sciences, Symbiosis International (Deemed University), Pune, Maharashtra, India

Anthony Augustine
University Sports Board, Symbiosis International (Deemed University), Pune, Maharashtra, India

Correspondence
Amritashish Bagchi
Symbiosis School of Sports Sciences, Symbiosis International (Deemed University), Pune, Maharashtra, India

Forecasting the winner of pro kabaddi league matches

Amritashish Bagchi, Shiny Raizada, Aniket Mhatre and Anthony Augustine

Abstract
The purpose of the study was to develop a prediction model to forecast the outcome of pro kabaddi league matches. These probabilities can assist a coach, team captain or manager in considering a certain tactics for the other half. The data was collected from 2017 season of Pro Kabaddi League (PKL). A total data of 272 matches were recorded, out of which 32 matches were draw and therefore not included in the study. The dependent variable selected for this study was Match Outcome (Win/Loss). Raid Points, Tackle points, All Out Points and Extra Points were selected as the predictor variables. For the purpose of this study only the first half data was used and in statistical technique Binary Logistic regression was used to predict the outcome of a match (Win/Loss). The result indicates that the developed Logistic regression Model was significant. According to the statistical significance of the predictor variables, they were numerically weighted and can be used to predict the match outcome. All the predictor variables except Extra Points were included in the prediction model with coefficient of determination (R^2) of.203 (Cox & Snell) and .271 (Nagelkerke). The classification matrix shows that 68.8% of match results were correctly classified by the model.

Keywords: Kabaddi, pro kabaddi league (PKL), prediction model, win and loss

Introduction
Kabaddi is one of the traditional sports of India, which requires a combination of Agility, Speed, Power, Strength and Co-ordination. Kabaddi is a contact team sport, where 14 players (7 on each side) take part at time and needs no equipment which makes it a very popular sport in the developing countries [3]. It requires a small playing ground; time span of the match shall be a minimum of 40 minutes divided equally in two halves of 20 minutes each along with a rest period of 5 minutes between halves. Although it is an outdoor sport that is being played on clay court, but now a days with its great success it is being played in indoors on synthetic surface.

In today's high-tech world, technology in sport plays a vital role for not only recording data in many ways but also analyse it and present it in such a format (graph or tables) that the coaches could easily interpret and make some conclusion for the individual or team performance. These data are the match statistics such as possession percentage, number of successful tackles, interceptions, Percentage of successful shot passes/long passes in football, percentage of Aces, Number of unforced errors, Percentage of successful service return, Second serve win percentage in Tennis and so on. These match statistics or data vary from sport to sport and is the one that viewers want to know above all [5].

Many researches were done to predict the future events on the basis of present or past data. These researches had proved that the future performance of an individual or team could be predicted through the analysis of certain variables, which are found to be the basis for total performance [3, 10]. Researches in Kabaddi show the anthropometric, physical, physiological and psychological needs or characterstics of Kabaddi players [1, 6, 7, 8]. But to the best of my knowledge none of the studies in Kabaddi were focused in predicting the match winner on the basis of match statistics.

Logistic regression is one of the statistical methods that can be one of the solutions of many prediction models in sports. The main purpose of this statistical technique is to predict the outcome (binary or multinominal) on the basis of predictor variables selected by the researcher

[10, 11]. Many prediction models have been developed by researchers in predicting the match outcome in several sports. The purpose of this study is to develop a prediction model to predict the outcome of a Kabaddi match on the basis of match statistics of first half.

Methodology

A total data of 272 matches were recorded, out of which 32 matches were draw and therefore not included in the study. All the data were collected from the website flashscore.com [9]. The dependent variable selected for this study was Match Outcome (Win/Loss). Raid Points, Tackle points, All Out Points and Extra Points were selected as the predictor variables. For the purpose of this study only the first half data was used. Data is presented as mean with standard deviations. The statistical technique Binary Logistic Regression was used to develop the prediction model. For this purpose Statistical Package for Social Science (SPSS) version 24.0 was used. The level of significance was set at 0.05.

Results and Discussion

Unlike linear regression and general linear models, logistic regression does not make many of the key assumptions that are based on ordinary least squares algorithms, such as linearity, normality, homoscedasticity, and measurement level. Therefore, only the descriptive statistics (i.e. mean, standard error of mean, standard deviation, skewness, kurtosis etc.) was used to see the nature of data and the correlation matrix was used to check the assumption of high multicollinearity among the variables.

Table 1: Descriptive statistics

	Raid points	Tackle Points	All Out Points	Extra Points
Mean	8.9125	4.1667	1.0000	.9042
Std. Error of Mean	.23147	.14253	.08516	.06091
Std. Deviation	3.58586	2.20814	1.31931	.94359
Skewness	.515	1.219	1.058	.795
Std. Error of Skewness	.157	.157	.157	.157
Kurtosis	.154	4.034	.338	-.051
Std. Error of Kurtosis	.313	.313	.313	.313
Range	20.00	15.00	6.00	4.00

Table 2: Correlation Matrix

		Raid points	Tackle Points	All Out Points	Extra Points
Raid points	Pearson Correlation	1	.067	.570**	.007
Tackle Points	Pearson Correlation	.067	1	.471**	.012
All Out Points	Pearson Correlation	.570**	.471**	1	.212**
Extra Points	Pearson Correlation	.007	.012	.212**	1
**. Correlation is significant at the 0.01 level (2-tailed).					

The correlation coefficient between sets of variables is shown in the above table. The correlation matrix was used to check the assumption of multicollinearity. Although there is a significant correlation between variables but none of the variable were found to be highly correlated. And this was checked by calculating Variance Inflation Factor (VIF) using SPSS. Variance Inflation Factor quantifies the severity of multicollinearity. For all the variables the VIF value was near by 1, which means the multicollinearity between the independent variables was low.

Table 3: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	54.462	3	.000
	Block	54.462	3	.000
	Model	54.462	3	.000

As compared to -2 Log Likelihood value (i.e. 332.711) of the null model, the omnibus test of model coefficients shows a significant decrease in the -2 Log Likelihood value (i.e. 278.249), it means the developed model is significantly better fit than the null model.

Table 4: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	278.249[a]	.203	.271
a. Estimation terminated at iteration number 5 because parameter estimates changed by less than.001.			

From the above table it can be seen that the value of Nagelkerke R^2 is.271 in the third model developed in binary logistic regression, but the value of Cox & Snell R-square is found to be.203. The Nagelkerke R^2 value was considered for the developed model because the Cox & Snell R-square is based on the log likelihood for the model compared to the log likelihood for a baseline model. However, even for a "perfect" model with categorical outcomes, it has a theoretical maximum value of less than 1. Nagelkerke R^2 is the adjusted version of the Cox & Snell R-square that adjusts the scale of the statistic to cover the full range from 0 to 1 [4]. The value of Nagelkerke R^2 is.271 which means 27.1% of the variability in the dependent variable is explained by the selected independent variables.

Table 5: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.011	8	.432

The Hosmer-Lemeshow test (HL test) is a goodness of fit test for developed logistic regression model. The null hypothesis of Hosmer-Lemeshow test is that the fitted model is correct, which means that higher the p – value better is the model. In the above table, the p – value of Hosmer and Lemeshow test is.432 which is insignificant. Hence the model fit is good, in other words the observed event rates match the expected event rates in population subgroups.

Table 6: Classification Table

Steps	Observed		Predicted		
			Match Results		Percentage Correct
			LOSS	WIN	
Step 1	Match Results	LOSS	91	29	75.8
		WIN	46	74	61.7
	Overall Percentage				68.8
a. The cut value is.500					

The above table shows the summary of correct and wrong classification of the subjects in match Outcome (i.e. Loss or Win) on the basis of the developed regression model. It can be seen from the table that in loss category out of 120 matches 91 matches were correctly classified and similarly, 74 out of 120 matches were correctly classified in the category of win. Overall 68.8% of matches were correctly classified on the basis of selected independent variables.

Table 7: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Raid Point	.096	.055	3.086	1	.079	1.101
	Tackle Point	.236	.090	6.844	1	.009	1.267
	All Out Point	.522	.163	10.225	1	.001	1.685
	Constant	-2.270	.646	12.369	1	.000	.103
a. Variable(s) entered on step 1: raid point, tackle point, all out point.							

The above table provides the regression coefficient (B), the Wald statistic (used to test the significance of individual coefficients in the model) and the all-important Odds Ratio (Exp (B)). "B" coefficients are also known as unstandardized coefficients and are used to develop the regression equation [2]. Three variables out of four independent variables were selected by the model i.e. Raid Point, Tackle Point and All Out Point. All the independent variables in the model were found to be significantly contributing. All these variables are important in predicting the match outcome of a Kabaddi match. It may help the team captain, coaches or managers to change the tactics accordingly for the second half. But it should be taken into consideration that although the variables included in the model is highly significant and it can classify upto 68.8% of cases correctly, still it only explain 27.1% of the variability in the dependent variable. It means 72.9% of the variability is explained by some other variables which were not included in the study.

All the four variables selected for the study were important to win a match. To understand the importance one needs to understand the meaning of these terms. Whenever the raider enters the opposition half he has two means to score a point (Raid point). One is Bonus point which is awarded when six or more defenders are in the court and if a raider crosses the bonus line with trailing foot in the air he gets the bonus point. The other one is Touch point, it is awarded when the raider initiates touch on one or more defender and returns to mid line, he get the touch points equivalent to defender he has touch and those defenders are force out of play and equivalent number of players from raiding team who are out get revived. When the defender prevents the raider from returning to midline within 30sec of raid the tackle point awarded to defender side. Similarly, all out point is when all players of the team are forced out of play and all out is set to be inflicted upon them with the opposition team being awarded two additional points the play is started by reinstating all the seven players of the team back on court. The last variable which is not selected by the model is Extra Points, it includes the error or technical point given by the referee to the opponent team (Such as T-shirt tackle, argument with referee, delay in raid, coaching while game is on, wrong tackles which are not allowed to be done etc.). It is not included in the model may be because as compare to other variables it happens less number of times.

Regression Equation

Using regression coefficients (B) of the model shown in the table 7, the regression equation was developed which is as follows:

Logit = -2.270 +.522 (All Out Point) +.236 (Tackle Point) +.096 (Raid Point)

Odds = e^{logit} = -2.270 +.522 (All Out Point) +.236 (Tackle Point) +.096 (Raid Point)

$$P(Y) = \frac{odds}{1+odds}$$

The above regression equation can be used to predict the match outcome (i.e. Win/Loss) of the future Pro Kabaddi League matches on the basis of three predictor/ independent variables (i.e. All Out Point, Tackle Point and Raid Point) of the first half data.

Conclusion

The purpose of the study was to develop a model to predict the outcome of Pro Kabaddi League matches on the basis of first set data. The developed Logistic regression Model was found to be significant. According to the statistical significance of the predictor variables, they were numerically weighted and were used to predict the match outcome. Out of four predictor variables three were included in the prediction model with coefficient of determination (R^2) of.203 (Cox & Snell) and .271 (Nagelkerke). 68.8% of match results were correctly classified by the model. Further study could be done by including few more variables that significantly contribute to the match outcome. So that the remaining variability can be explained and the model fit can be improved for more correct prediction along with high probability.

References

1. Ali S, Adhikari D. Physical and Anthropometric Characteristics of Kabaddi Players. Indian Journal of Applied Research. 2011; 4(1):464-465.
2. Bewick V, Cheek L, Ball J. Statistics review 14: logistic regression. Critical Care. 2005; 9(1):112-8.
3. Devaraju K, Kalidasan R. Prediction of Kabaddi Playing Ability from Selected Anthropometrical and Physical Variables among College Level Players. Asian Journal of Information Technology. 2012; 11(4):131-134.

4. IBM Knowledge Center. Ibm.com. https://www.ibm.com/support/knowledgecenter/en/SSLVMB_sub/spss/tutorials/plum_germcr_rsquare.html. 8 January 2018.
5. Klaassen FJ, Magnus JR. Forecasting the winner of a tennis match. European Journal of Operational Research. 2003; 148(2):257-267.
6. Nizama K. Comparison of Mental Toughness Variables among Kho-Kho and Kabaddi Players of Junagadh District. International Journal of Scientific Research. 2012; 2(3):343-344.
7. Palanisamy D, Dhanaraj S, Subbiah S. Influence of Circadian Rhythms on Selected Physiological Variables among College Trained and Untrained Male Kabaddi Players. International Journal of Scientific Research. 2012; 1(4):128-129.
8. Patel S, Pandey U, Saxena S. Comparative Study of Mental Toughness among Kabaddi Players of Different Level. Indian Journal of Applied Research. 2011; 1(3):201-202.
9. Pro Kabaddi League. Results - Kabaddi/India. Flashscore.com, 2017. https://www.flashscore.com/kabaddi/india/pro-kabaddi/results/. 12 February 2018.
10. Raizada S, Bagchi A, Menon H, Nimkar N. Predicting the outcome of ICC cricket world cup matches. International Journal of Physiology, Nutrition and Physical Education. 2019; 4(1):119-122
11. Schoonjans F. Logistic regression. MedCalc. https://www.medcalc.org/manual/logistic_regression.php. 3 February 2018