International Journal of Physical Education, Sports and Health 2022; 9(6): 134-137

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (RJIF): 5.38
IJPESH 2022; 9(6): 134-137

www.kheljournal.com
Received: 18-08-2022
Accepted: 23-09-2022

Prashant Kumar Choudhary
PhD Scholar, Department of Sports Management, Lakshmibai National Institute of Physical Education, Gwalior, Madhya Pradesh, India

Suchishrava Dubey
Ph.D. Scholar Department of Sports Psychology, Lakshmibai National Institute of Physical Education, Gwalior, Madhya Pradesh, India

Dipendra Singh
Bachelor student, Lakshmibai National Institute of Physical Education, Gwalior, Madhya Pradesh, India

Corresponding Author:
Prashant Kumar Choudhary
PhD Scholar, Department of Sports Management, Lakshmibai National Institute of Physical Education, Gwalior, Madhya Pradesh, India

A statistical model to forecast the outcome of the golden state warriors through NBA season matches

Prashant Kumar Choudhary, Suchishrava Dubey and Dipendra Singh

DOI: https://doi.org/10.22271/kheljournal.2022.v9.i6b.2685

Abstract
The purpose of the study was to develop a Statistical model to predict the outcome of one of the teams in the NBA i.e. the Golden State Warriors. The last eight seasons' data were taken into consideration for the statistical analysis. These probabilities can assist a team captain or management in considering a certain offensive, defensive, or playing strategy for the next season. The data was collected from the last eight seasons of the NBA i.e. from 2014 to 2022. Data from 656 matches were recorded, and the dependent variable selected for this study was Match Outcome (Win/Loss). Field Goals Made (FGM), Field Goals Attempted (FGA), Three-Point Field Goals Made (3 PM), Three-Point Field Goals Attempted (3PA), Free Throws Made (FTM), (FTA) Free Throws Attempted, and Rebound (REB), these variables were chosen to serve as the predictor variables. For this study, the data was taken from the last eight season matches of the NBA and further, it was interpreted through a statistical technique, binary logistic regression which was used to predict the outcome of a match (Win/Loss). It was found 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. Out of seven predictor variables, only the variable Team score was included in the prediction model with a coefficient of determination (R2) of.750 (Cox & Snell) and .1.000 (Nagelkerke). 95.87% of match results were correctly classified by the model.

Keywords: NBA season, prediction model, binary logistic regression, total free throw attempt, win/loss

Introduction
The National Basketball Association, often known simply as the NBA, is a professional basketball league played throughout the continent of North America. The league is one of the most important professional sports leagues in both the United States and Canada. It has a total of 30 clubs, 29 of which are based in the United States, and one team that plays in Canada. The league was established in New York City on June 6, 1946, under the name Basketball Association of America (BAA). The Golden State Warriors are a professional basketball team from the United States that plays its home games in San Francisco. The Warriors are a professional basketball team that competes in the National Basketball Association (NBA). They are a part of the Western Conference Pacific Division of the NBA. The Warriors were able to repeat as champions of the Basketball Association of America (BAA) in 1956 because of the efforts of Hall of Famers Paul Arizin, Tom Gola, and Neil Johnston. The Warriors won the BAA title for the first time in 1947. The franchise concluded the 1964–1965 NBA season with the poorest record in the league after trading away their best player Wilt Chamberlain in January of that year (17-63). Their time spent rebuilding was cut short in large part because the Warriors made the selection of Rick Barry in the first round of the draught four months after the transaction. The Warriors pulled off what is generally regarded as one of the most shocking victories in the annals of NBA history when they won their third championship in 1975, led by great players Barry and Jamaal Wilkes.
The Golden State Warriors are a professional basketball club from the United States that is located in San Francisco and competes in the National Basketball Association's Western Conference (NBA). The Warriors have triumphed to seven titles overall, including six in the NBA and one in the Basketball Association of America (BAA).

The Warriors have been around since 1946 and had their beginnings in Philadelphia, where they were initially situated. Because of the performance of future Hall of Fame forward Joe Fulks, who was the BAA's first scoring leader, this club, which was one of the initial members of the BAA, was able to win the league's first title. The next year, the Warriors fell short of winning the BAA championship, and the following year after that, the BAA amalgamated with the National Basketball League, which is when the Warriors joined the NBA (NBL). During their first six seasons in the new league, the Warriors only had one season in which they completed their division's race in a position higher than fourth place. However, during the 1955–1956 season, the Warriors not only had the best record in the league but also won their first NBA championship thanks to the efforts of forward Paul Arizin and center Neil Johnston. In recent years, very few studies have concentrated their attention on the performance evaluation of individual NBA players or whole NBA teams during games. However, to the best of my knowledge, none of the research has concentrated on developing a prediction model to anticipate the outcome of the match based on the results of the past matches from the previous six seasons. The creation of prediction models in the sporting world might be one of the potential answers to the problem of predicting the result of a match. During the halftime break, it will be easier for the team captain, coaches, and team management to come up with fresh strategies. In the field of statistics, logistical regression is a well-liked approach for forecasting a result (binary or multinominal) based on a dataset that contains one or more independent variables. This method may be used to make predictions about a range of outcomes. These variables are also referred to as predictor variables, and their underlying nature may either be scaled or categorical. When the total number of covariates is either very big or strongly correlated, there is a possibility that the parameters may become unstable. Logistic regression necessitates that there be minimal to no multicollinearity among the variables that are considered independent. This entails ensuring that the independent variables do not have an excessively high degree of correlation with one another [3]. Checking that the assumption of multicollinearity holds among the independent variables is one way to address this concern. Additionally, using the model will assist in determining the proportion of times that the model correctly classified data. The goal of the research was to construct a model that could forecast the results of forthcoming NBA season matches for the Golden State Warriors (GSW) based on data from the previous eight seasons. These matches include those that will take place in 2023 and beyond.

Methodology

The information was compiled using statistics from the most recent eight NBA seasons, which span the years 2014 through 2022. There were a total of 656 matches reported in the data. Because logistic regression relies on a limited number of assumptions, one of those assumptions is that the dependent variable must have a binary data type. Therefore, the outcome of the match (whether it was won or lost) was chosen to be the dependent variable for this investigation. Field Goals Made (FGM), Field Goals Attempted (FGA), Three-Point Field Goals Made (3 PM), Three-Point Field Goals Attempted (3PA), Free Throws Made (FTM), (FTA) Free Throws Attempted and Rebound (REB). These variables were chosen to serve as predictors. The whole of the information was obtained from the NBA.com website. The researcher took into account a total of eight different seasons for the aim of carrying out this investigation. In the field of statistical methodology, The prediction model was developed using a technique called binary logistic regression. To better understand the nature of the data, descriptive statistics were used. Before beginning the analysis, all of the presumptions were addressed and taken care of. Statistical Package for the Social Science (SPSS) version 20.0 was used to accomplish this goal. A value of 0.05 was chosen to represent the level of significance.

Result and Discussion

Table 1: Descriptive statistics of all scaled variables

Variables	N	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
TFGM	8	41.4000	.73485	2.07846	-.434	.752	-1.486	1.481
TFGA	8	87.3875	.49260	1.39329	.148	.752	.842	1.481
THREEPM	8	12.4750	.56117	1.58723	.053	.752	-1.634	1.481
THREEPA	8	32.8125	1.55855	4.40825	.507	.752	-.832	1.481
TFTM	8	16.7875	.35428	1.00205	1.087	.752	.834	1.481
TFTA	8	21.3125	.39299	1.11155	.831	.752	-.776	1.481
TREB	8	44.5375	.48105	1.36061	.007	.752	-1.699	1.481
Valid N (List wise)	8							

In contrast to linear regression and other broad linear models that are founded on ordinary least squares algorithms, logistic regression does not make nearly as many of the same fundamental assumptions. These assumptions include linearity, normalcy, homoscedasticity, and measurement level. Therefore, only descriptive statistics (such as mean, standard error of the mean, standard deviation, skewness, kurtosis, etc.) were used to see the nature of the data, and the correlation matrix was used to check the assumption of high multicollinearity among the variables. This is one of the few assumptions that need to be fulfilled, and therefore only descriptive statistics were used.

Table 2: Correlations

		TFGM	TFGA	THREEPM	THREEPA	TFTM	TFTA	TREB
TFGM	Pearson Correlation	1	-.064	-.165	-.474	-.250	-.323	.574
	Sig. (2-tailed)		.880	.695	.235	.551	.436	.137
	N	8	8	8	8	8	8	8

TFGA	Pearson Correlation	-.064	1	.236	.294	.189	.209	.196
	Sig. (2-tailed)	.880		.574	.479	.654	.620	.641
	N	8	8	8	8	8	8	8
THREEPM	Pearson Correlation	-.165	.236	1	.885**	-.533	-.423	.365
	Sig. (2-tailed)	.695	.574		.004	.174	.296	.374
	N	8	8	8	8	8	8	8
THREEPA	Pearson Correlation	-.474	.294	.885**	1	-.285	-.232	.092
	Sig. (2-tailed)	.235	.479	.004		.494	.581	.828
	N	8	8	8	8	8	8	8
TFTM	Pearson Correlation	-.250	.189	-.533	-.285	1	.917**	-.546
	Sig. (2-tailed)	.551	.654	.174	.494		.001	.162
	N	8	8	8	8	8	8	8
TFTA	Pearson Correlation	-.323	.209	-.423	-.232	.917**	1	-.363
	Sig. (2-tailed)	.436	.620	.296	.581	.001		.377
	N	8	8	8	8	8	8	8
TREB	Pearson Correlation	.574	.196	.365	.092	-.546	-.363	1
	Sig. (2-tailed)	.137	.641	.374	.828	.162	.377	
	N	8	8	8	8	8	8	8

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

In logistic regression, one of the assumptions is that there should not be a substantial degree of multicollinearity among the variables that are being analyzed independently. The multicollinearity assumption was validated with the help of the correlation matrix table that can be seen above. This table displays the correlation coefficient that was found between each set of variables. Even though there is a considerable connection between the variables, none of the variables were determined to be significantly associated. This is even though there is a correlation between the variables. This was validated by using SPSS to perform a calculation known as the Variance Inflation Factor (VIF). The VIF value was 1, which indicates that there was no multicollinearity between the independent variables. This was the case for all of the variables. As a result, we can proceed with the analysis of the logistic regression.

Table 3: Omnibus tests of model coefficients

		Chi-Square	DF	Sig.
Step 1	Step	11.090	1	.001
	Block	11.090	1	.001
	Model	11.090	1	.001

2LL is a measure of how well the estimated model fits the likelihood. A good model results in a high likelihood of the observed results. This translates to a small number for -2LL (If a model fits perfectly, the likelihood is 1, and -2 times the log-likelihood is 0). The omnibus test of model coefficients shows a significant decrease in the -2 Log Likelihood value (i.e. .000[a]), which means the developed model is a significantly better fit than the null model.

Table 4: Model summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.000[a]	.750	1.000

Estimation terminated at iteration number 20 because parameter estimates changed by less than.001.

Unlike linear regression in logistic regression, there is no actual (Coefficient of Determination) value, which summarizes the proportion of variance in the dependent variable, explained by the independent variable selected by the model. The higher the proportion better will be the model. It can be seen from the above table that in the second model the value of Nagelkerke is.1.000 and the value of Cox & Snell R-square is found to be.750. Both Nagelkerke and Cox & Snell R-square values are the approximation of the actual value. The Nagelkerke value was considered for the developed model because in Cox & Snell R-square even for a "perfect" model with categorical outcomes, it has a theoretical maximum value of less than 1. Nagelkerke 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 [10]. The value of Nagelkerke is 1.000 which means 100% of the variability in the dependent variable is explained by the selected independent variables.

Table 5: Hosmer and lemeshow test

Step	CHIN-Square	DF	Sig.
1	.000	4	1.000

The Hosmer-Lemeshow test, often known as the HL test, is a test that determines whether or not a constructed logistic regression model is accurate. It is a test of the null hypothesis, which states that the fitted model is accurate; hence, the p-value has to be negligible for it to be possible to reject the null hypothesis. The p-value in the table that you just looked at is 1.000, which is much higher than.05. As a result, the model fit is satisfactory.

Table 6: Classification Table [a,b]

	Observed		Predicted		
			Outcome		Percentage correct
			Loss	Win	
Step 0	Outcome	Loss	200		30.48
		win	0	429	65.39
	Overall Percentage				95.87

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) based on the developed regression model. It unveils the number of wins predicted by the logistic regression model compared to the number actually observed and similarly the

number of losses predicted by the logistic regression model compared to the number observed. Overall 72.2% of matches were correctly classified based on selected independent variables.

Table 7: Variables in the equation

TFGA		B	S.E.	Wald	DF	Sig.	Exp(B)
Step 1[a]	Constant	.000	.707	.000	1	1.000	1.000

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 (Bewick, Cheek & Ball, 2005). Only the variable Field goal Attempted is selected by the model. Basketball is an unpredictable game where fortunes can change in a matter of time. The result depends on many factors which work together and make this game unpredictable. These factors are falls in crucial situations/times, miss of a shoot from a good basketballer, injury to a cardinal player, wrong decisions by referees, whether condition change suddenly, indoor courts lights and the surface area also gets affected by constant sweat of players. And many more. Most of the paramount variables selected by the model are which contribute to a significant difference in the game of basketball, furthermore, NBA is a league where every team stands equal when concerning playing ability and performance. What matters is the subtle difference incurred due to the experience and matches played a factor.

Regression Equation Using regression coefficients (B) of the model shown in table 7, the regression equation was developed which is as follows: Logit = .762 +.000 (TGFA). The above regression equation can be used to predict the match outcome (i.e. Win/Loss) of the future NBA Matches for Golden State Warriors (GSW) based on one predictor/ independent variable (i.e. TGFA) of the last eight season data.

Conclusion

The purpose of the study was to develop a Statistical model to predict the outcome of the Golden State Warriors through NBA season matches, based on data from the last eight seasons i.e. from 2014 to 2022. 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. Only one variable i.e. Team Score out of seven variables is selected by the model with a coefficient of determination (R^2) of.750 (Cox & Snell) and .1.000 (Nagelkerke). 95.87% of match results were correctly classified by the model. Further study could be done by including more predictor 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 accurate prediction along with high probability.

Reference

1. This Date in the NBA: June. National Basketball Association. Retrieved March 4, June 6, 1946. The National Basketball Association was founded at the Commodore Hotel in New York. Maurice Podoloff was the league's first president, a title later changed to commissioner; c2022.
2. NBA Directories (PDF). National Basketball Association. October 17, 2019. Retrieved February 24; c2022.
3. Rathborn, Jack (November 18, 2020). "NBA Draft 2020: What time does it start in the UK, who has the No 1 pick and how can I watch it? The Independent. Archived from the original on June 18, 2022. Retrieved December 10, 2020. The 2020 NBA Draft is here after days of juicy gossip surrounding trades as the world's greatest basketball league dominates the headlines during its offseason.
4. This Date in the NBA: August. National Basketball Association. Retrieved June 14, 2020.
5. The World's Highest-Paid Athletes. Forbes; c2020.
6. Revealed. The world's best paid teams, Man City close in on Barca and Real Madrid. SportingIntelligence.com. May 1, 2012. Retrieved June 11; c2012.
7. Gaines, Cork. The NBA is the highest-paying sports league in the world. Business Insider. Retrieved May 20, 2015.
8. Members of USA Basketball. USAB.com. Retrieved June 14; c2020.
9. Mathewson, TJ (March 7, 2019). TV is biggest driver in global sport league revenue. GlobalSportMatters.com. Retrieved December 25, 2020.
10. Raizada S, Bagchi A, Menon H, Nimkar N. Predicting the outcome of ICC cricket world cup matches. Indian Journal of Physical Education, Sports Medicine & Exercise Science. 2018;18(2):60-65.