Will Stephen Curry Hit a Double-Double?

Keywords

Double-Double - A record of double digit numbers in at two offensive categories (Points, Blocks, Steals, Assists, Rebounds).

Triple-Double - A record of triple digit numbers in at two offensive categories (Points, Blocks, Steals, Assists, Rebounds).

Performance Stats - Referring to PTS, REB, STL, AST, BLK

If a player achieves a Triple-Double or a Triple-Triple, we will consider these as in the same category as achieving a Double-Double.

Introduction

A Double-Double in the NBA is a difficult task to achieve and greatly increases the probability of winning if a player hits a Double-Double in a game. This statistics is of high importance to many sports betters out there because it is such a simple statistic to bet on. Either a player achieves a Double-Double or he doesn't. However, since hitting a Double-Double is uncommon in most games, betting that a specific player will achieve a Double-Double and finding success in this bet is a sure way to gain a big sum of money. In general, the probability that a player will hit a Double-Double is fairly low and so the reward for successfully picking a player hitting a Double-Double is significantly greater than not hitting a Double-Double.

The goal of this experiment is to calculate the probability of Stephen Curry hitting a Double-Double in the 2018 NBA Finals, Game 1 against Cleveland. The given data provides Curry's statistics for all of his games from his debut on 10/28/2009 to 10/29/2018. Although the outcome has already been observed, the focus of this experiment is to predict his performance as if the 2018 NBA Finals did not already happen. We will pretend that the 2018 Conference Finals has just finished and the next game played will be the 2018 NBA Finals, Game 1. After we calculate Curry's probability on achieving a Double-Double, we will compare our result with the actual outcome and check the accuracy of our model.

There are two side questions that should also be considered and those are the *(i)* average count of PTS, AST, REB, STL and BLK per minutes played and *(ii)* the probability of achieving a Double-Double against each opponent. The graphs representing these statistics are provided below.

The model that will be used is the logistic regression model with the response variable outputting 1 if Curry has achieved a Double-Double and 0 otherwise. The predictor variables that will be taken into account are Points, Rebounds, Assists, Blocks, Steals, Opponent, and the number of minutes played in each game.

Methodology

The raw data provided consists of several variables (Dates, Opponent, Score, Minutes, Successful Shots, Total Shots, 3 Points Successful, Total 3 Points, Successful FT, Total FT, REB, AST, BLK, STL, PF, TO, PTS, Type, Result, Score GS, and Score Opponent). However, the only variables that will be taken into consideration are Opponent, Minutes, PTS, REB, AST, BLK, and STL. Additionally, all games recorded on and after 5/31/2018 will be removed as well. Lastly, the data has been split into six different seasons (as there are games between the Finals and the first season game) and has been merged so that our data only consists of these six seasons. Data cleaning is complete at this point.

(i) Average Count per Minutes Played
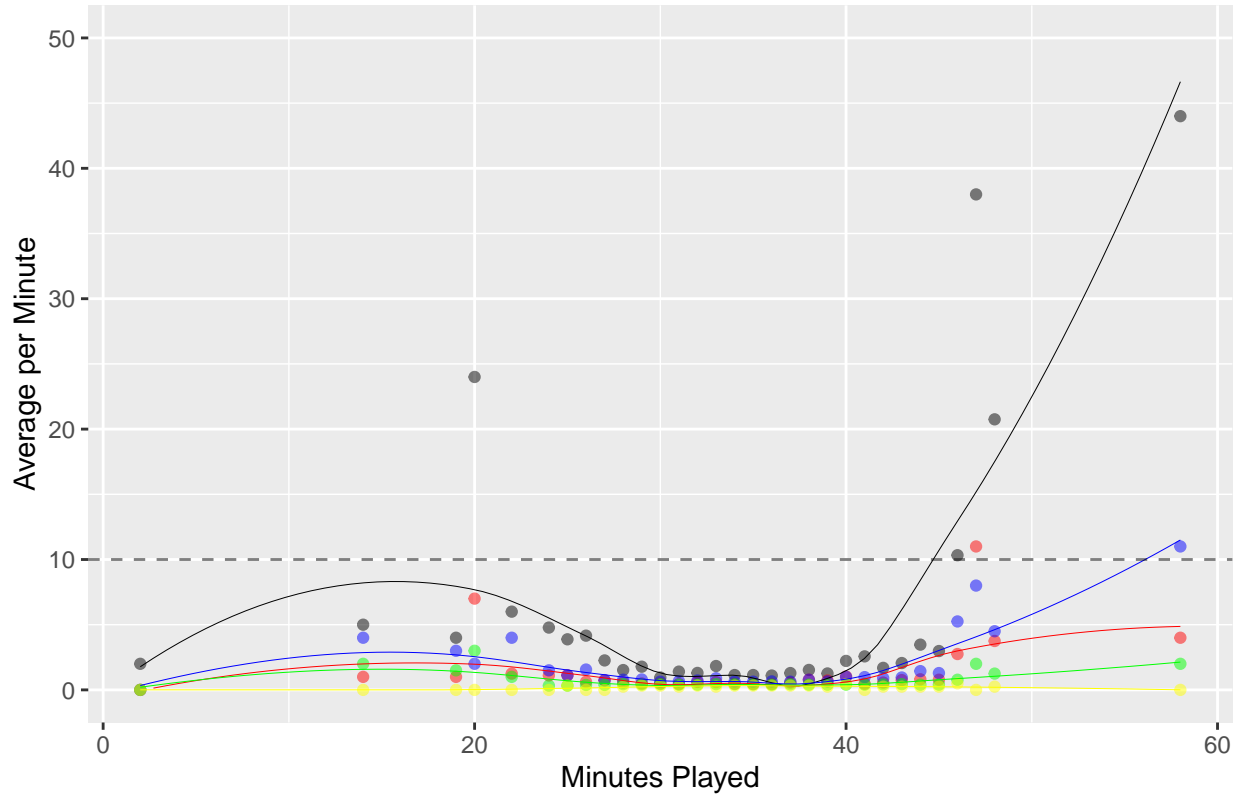
The importance of *(i)* is that it shows the expected mean for each variable at different numbers of minutes played. As we can see from the graph, it shows that regardless of the minutes played, Curry's performance stats average is below double-digits. This suggests that the probability of Curry hitting a Double-Double is lower than the probability of not hitting a Double-Double.

(ii) Probability of Double−Double Against each Opponent
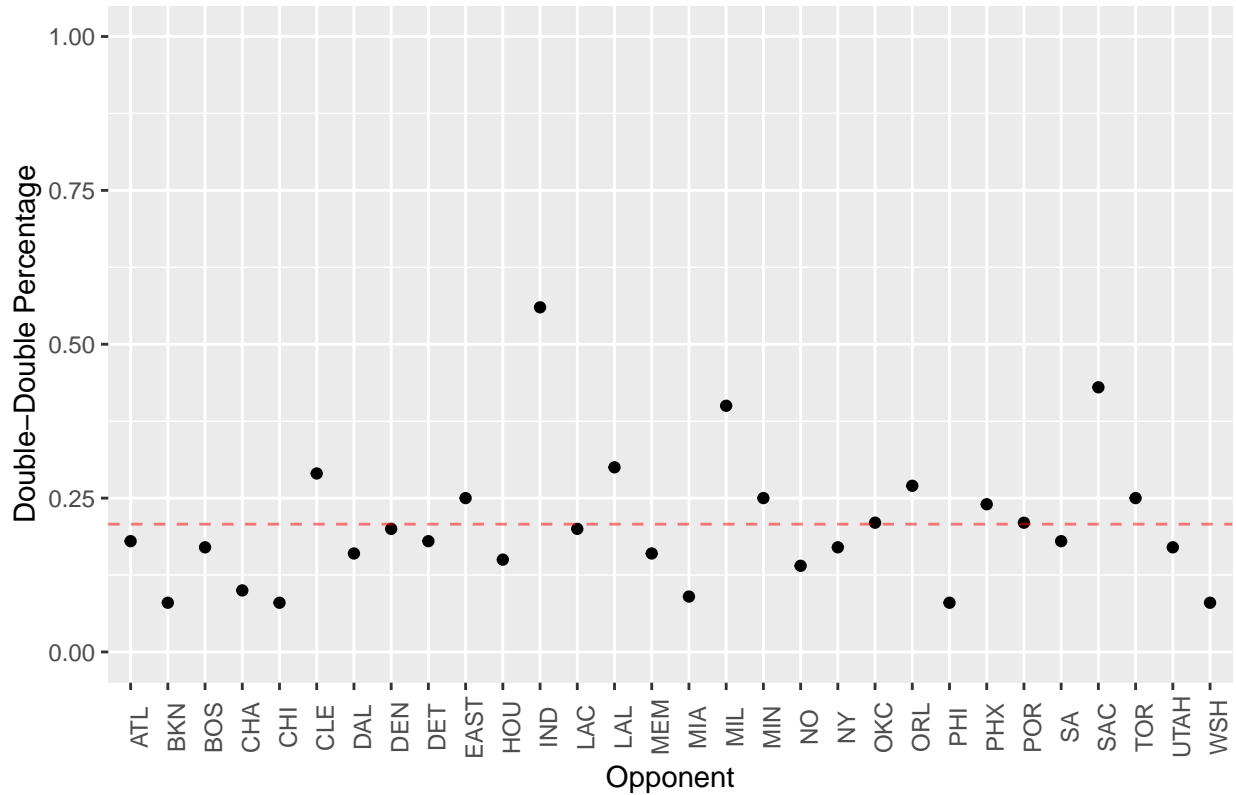
On the other hand, *(ii)* shows the probability of hitting a Double-Double against each team based on games prior to the 2018 NBA Finals, game 1. The dotted red line is the average probability of hitting a Double-Double against all teams. Observe that most probabilities lie below the average with only one being greater than 0.5; the probability of hitting a Double-Double against Indiana is roughly 0.55. Since the 2018 NBA Finals, Game 1 is against Cleveland, the probability against Indiana is irrelevant. However, this graph shows that the probability of hitting a Double-Double against any opponent is fairly low.

Combining observations of *(i)* and *(ii)*, it suggests that achieving a Double-Double in any game is low, including the 2018 NBA Finals, Game 1. This is consistent with the logistic regression model we have constructed and will be discussed in the following paragraph.

Logistic Regression Model
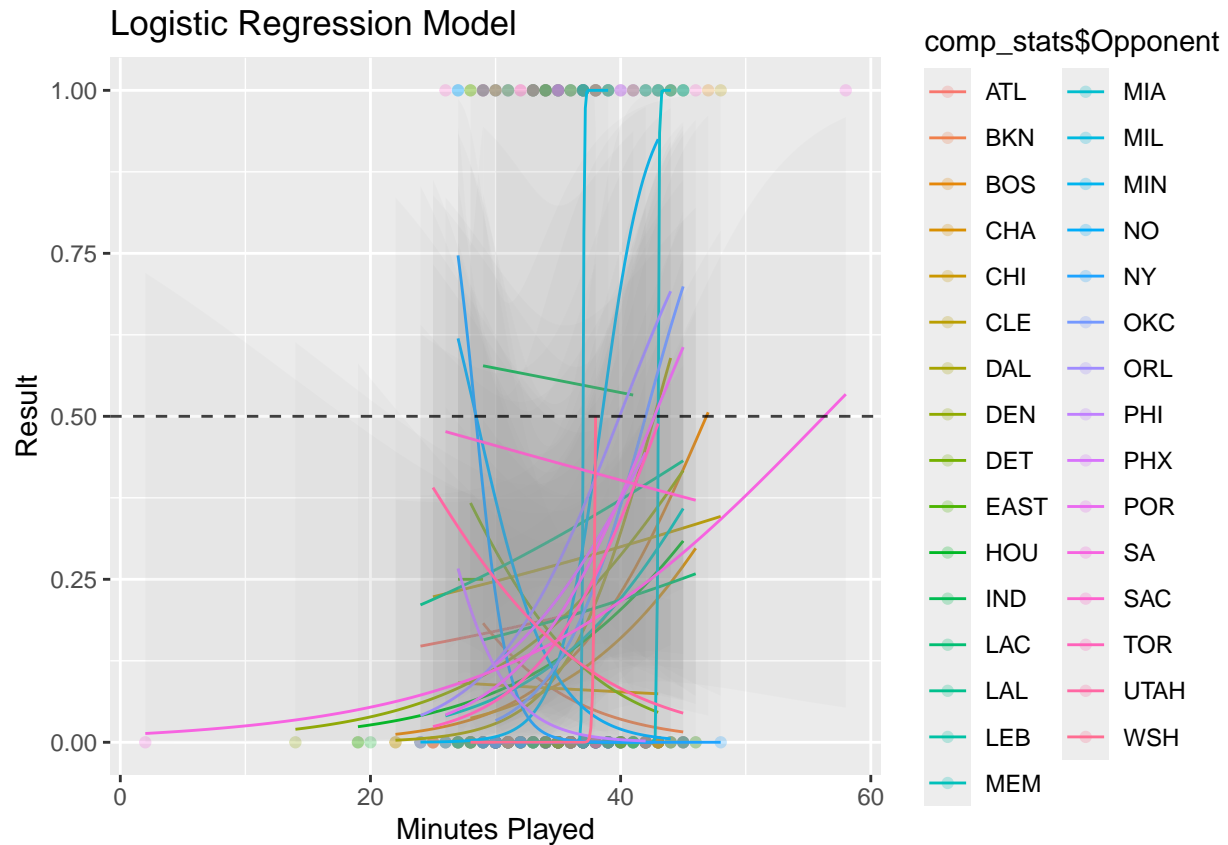
The last step in our study is to formulate our logistic regression model where our response is a binary output with predictor variables Minutes, Opponent, PTS, REB, AST, STL, and BLK. The graph below shows the logistic regression probability where the x-axis is the number of minutes played and the different colours represent all of the different opponents. For some opponents, very few games were recorded and and can be distinguished by the straightness of each line. The straighter the line, the less observations against that particular team were recorded. The dotted black line is the threshold in which we decide whether or not a Double-Double will be achieved. In short, if the observation is above this line, then Curry will achieve a Double-Double.

Results

To make our prediction, it will be assumed that Curry will be playing an average number of minutes with average performance stats against the Cleveland Cavaliers. Based on our logistic model, we estimate that the probability Stephen Curry will achieve a Double-Double in the 2018 NBA Finals, Game 1 against Cleveland is 0.0227. This is consistent with our other observations in that the probability of Curry achieving a Double-Double is fairly low in most aspects.

Discussion

Weaknesses:

One weakness this logistic regression model has is that there are many factors that affect Stephen Curry's performance outside of the variables mentioned. For example, a change in diet can affect his energy levels in a game, causing him to play better or worse. Or another example is Curry's mental state; if his head is in the game, it may cause him to perform better than if his mind was elsewhere. Strictly speaking, there are many uncontrollable factors outside the game of basketball that affect Curry's performance and taking these factors into account makes our model significantly more difficult to be accurate. As a "band-aid" solution to this problem, we assumed that Curry will perform at an average level with average stats to make our prediction.

Strengths:

One strength this logistic regression model has is that it is consistent with other observations such that the calculated probability of Curry achieving a Double-Double is low. This is also consistent with the actual recorded observation of the 2018 NBA Finals, Game 1 against Cleveland in which Curry did not achieve a Double-Double.

References

- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr

- Stephen Curry Stats: https://data.world/datatouille/stephen-curry-stats