Predicting Results of Basketball Games with Linear Regression

Duc Le, M.S

Mathematical Modeling

Computational & Data Sciences

Chapman University

DISCLAIMER

The following mathematical model is strictly for experimental and academic purposes. I am not a betting expert nor a financial advisor. If you choose to utilize the following model for alternative purposes, please do so at your own risk.

Abstract

The following report summarizes the technique used to predict an outcome of a basketball game between the Toronto Raptors and the Houston Rockets, taken place on December 5th, 2019. This mathematical modeling program could be applied to predict an outcome of any basketball game, given certain required statistical data and under specific assumptions. Based on these statistical data, which are past-game logs, the expected final score for each participating team can be computed using linear regression performed on the offensive categories on past-game logs.

Keywords*: Basketball. Mathematical Modeling. Linear Regression.*

Introduction

Although there are many factors that affect the outcome of a basketball game, the winner of any given game is determined by only one thing: total points scored. The following model will disregard other statistical categories recorded throughout a basketball game and will only focus on these specific offensive and defensive statistical data such as: field goals attempts, field goals made, three-point shots attempts, three-point shots made, free-throw attempts, and free-throws made.

The game logs used for this experiment were recorded through each team's first 17 games of the 2019-2020 NBA Season. Each team's first 17 games are consisted of games played on their home arenas (home games) and games played on the road (away games). Professional sports team perform differently depending on the environment of the competition. Some teams perform better at home, others play better on the road.

Before proceeding with the model, we must address the following assumptions:

1) We assume most of the players (especially the star/superstar players) from past games participate in the game.
2) We assume the participating players are healthy/injury-free, prior and throughout the game, and play their usual allocated minutes.
3) We assume the game will be held on one of the participating team's home court and is decided within the regulation period (no overtime).
4) The following prediction is made through the observed trend of past games. The method cannot measure nor predict anomalies/outliers. For example, the model cannot foresee a historically efficient or inefficient shooting night from either participating team.

The Model

From Basketball Reference, we collect the necessary data using the previous game logs. For better accuracy, it is important to separate the data collected from games at home and the games on the road. This will provide us with a better understanding on the difference between the teams' performances depending on where they play. For instance, some teams perform significantly better at home so when mixing in road games data, it might affect our result.

The data is split into six columns: field goals made, attempts, three-pointers made, attempts, free-throws made, attempts. We also want the same statistics for the defensive end, so we collect the statistics of field goals made allowed, field goal attempts allowed, etc.

Here is an example of the Toronto Raptors' home offensive stats:

FGM	FGA	3PM	3PA	FTM	FTA
42	103	14	40	32	38
33	85	10	39	28	30
51	86	13	27	10	12
43	78	14	29	24	29
50	97	20	43	12	15
45	87	9	27	14	19
38	84	11	34	14	22
46	94	21	41	13	17

Using MATLAB, we convert the table's columns into six different vectors containing the needed statistics. The process is done for the both team's offensive and defensive stats. Next, we calculate the Estimated Shot/Field Goal Attempts (EFGA) for both teams by taking the mean of the average field goals attempts (FGA) and the field goal attempts allowed. For example, if **Team A** has the average of **x field goal attempts** and **Team B** has the average of **y field goal attempts allowed**, then the EFGA for **Team A is** $\frac{x+y}{2}$**.** The same step is repeated for Expected Three-Pointers Attempts (E3PA) and Expected Free-Throws Attempts (EFTA).

Linear Regression

With MATLAB's built-in linear regression function, we apply it to the vectors extracted from the data. For this procedure, the shot attempts is the explanatory (independent) variable and the shots made is the scalar response (dependent). For instance, applying linear regression to the Field Goals Attempts (FGA) and Field Goals Made (FGM) will result in three outputs: the slope, y-intercept and residuals for the line of best-fit.

With this function, the Expected Field Goals Made (EFGM) can be calculated by inputting the EFGA. This procedure is repeated for all three chosen categories: Field Goals, Three-Pointers and Free-Throws for both teams. Once we have the expected shots made for each category, the final score can be computed by a simple formula:

Final score = 2*EFGM + ETM + EFTM. (With ETM = Expected Three-Pointers Made and EFTM = Expected Free-Throws Made).

Results

Mavericks @ Rockets 11/24

Model's Predicted Score

Mavericks 126 – Rockets 111 (Mavericks +15).

Mavericks: FG 47/93. 3PT 14/39. FT 18/23.

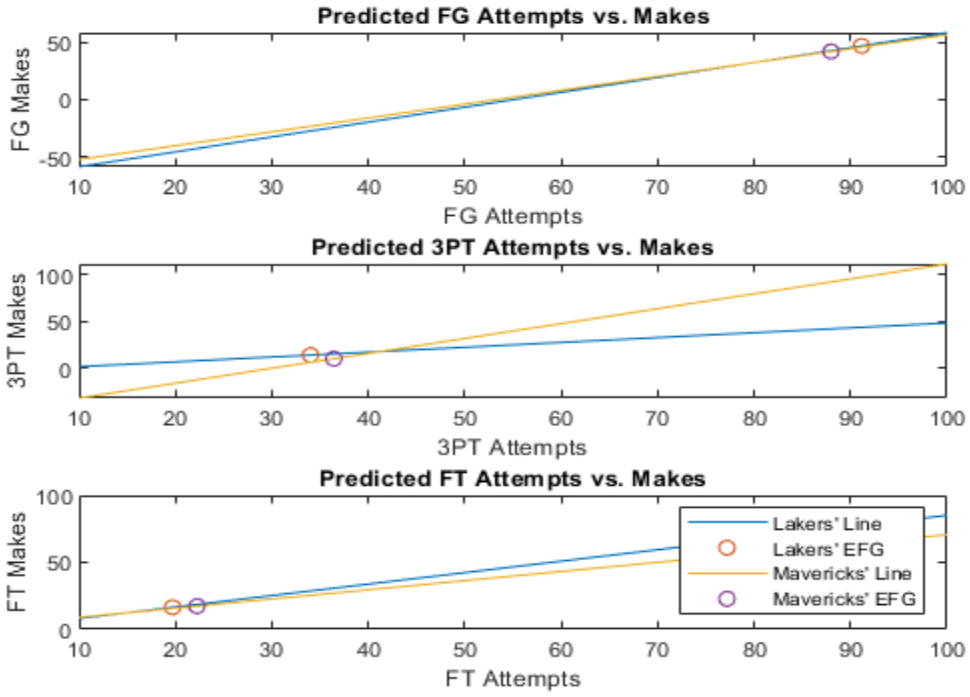
Rockets: FG 42/92. 3PT 7/40. FT 18/26.

Actual Score

Mavericks 137 - Rockets 123 (Mavericks +14).

Mavericks: FG 49/102. 3PT 17/44. FT 22/26.

Rockets: FG 45/98. 3PT 10/44. FT 23/31



Mavericks @ Lakers 12/1

Model's Predicted Score

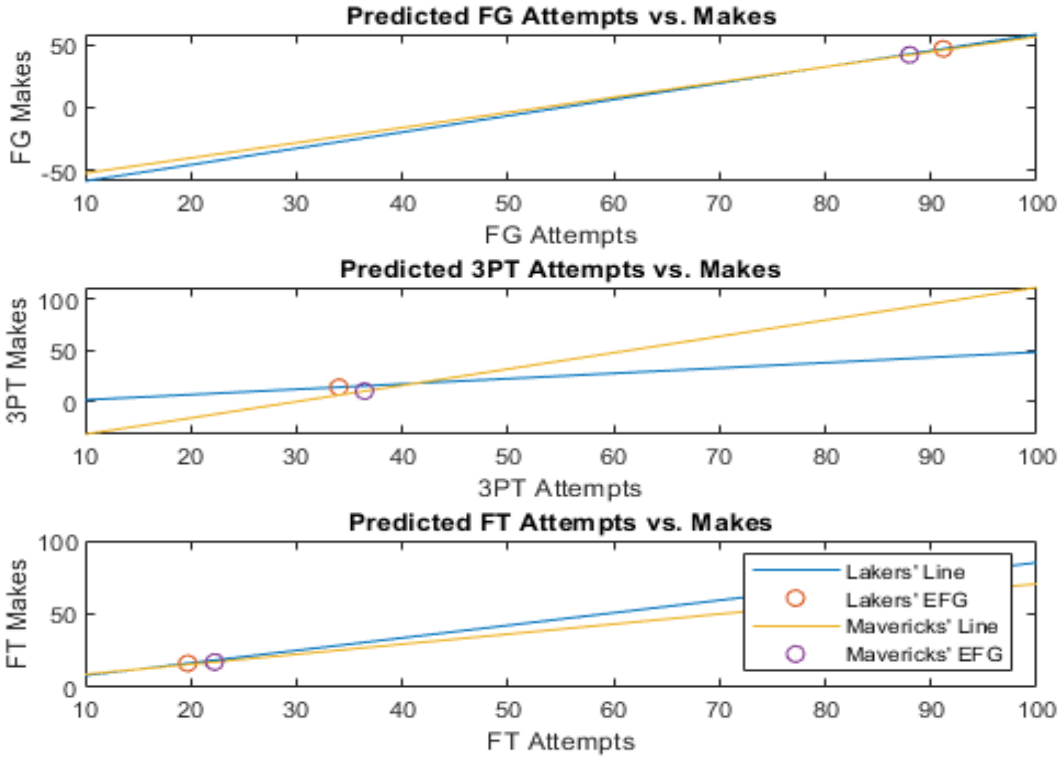
Lakers 123 – Mavericks 111 (Lakers +12)

Actual Final Score

Lakers 100 – Mavericks 114 (Mavericks +14)

*Mavericks went on a 28-3 run in the 3rd quarter to go up by 20.

*Statistical data from the Lakers is slightly inflated.



Rockets @ Spurs 12/3*

Model's Predicted Scores

When rounding the final score with "round" (MATLAB):

Spurs 117 – Rockets 117 (Draw)

When rounding the final score with "ceil" (MATLAB):

Spurs 118 – Rockets 117 (Spurs +1)

Actual Score

Spurs 135 – Rockets 133 in 2OT. (Spurs +2)

Through 4Q's, Spurs 115 – Rockets 115. (Spurs were down 22 in the 3rd quarter & 14 with 6 minutes left).

Elaboration for Rockets @ Spurs Game

This simulation was the first time a game resulted in a draw, each team with a score of 117. As mentioned above, one of the assumptions states that the result is assumed to be decided within the regulation period (through 4 quarters). To maneuver around this issue, I went ahead and experimented this the model by changing the "round" function previously used to acquire integer digits for the score to "ceil". This resulted in the Spurs winning by 1, Spurs 118 to Rockets 117. It just so happened that the actual game was tied at 115 before going into double overtime, resulting in a Spurs' victory by 2 points. Here is a deeper analysis of the statistics through regulation:

Model's Simulated Statistics in Regulation

Spurs 117 – Rockets 117

Spurs: FG 43/91. 3PT 12/30. FT 20/23.

Rockets: FG 42/90. 3PT 12/39. FT 21/25.

Actual Statistics in Regulation

Spurs 115 – Rockets 115

Spurs: FG 43/88. 3PT 11/24. FT 18/26.

Rockets: FG 36/96. 3PT 15/47. FT 24/26.

A coincidence is completely possible since the model did not account for games that are undecided by the end of regulation, but the result of this particular simulation is interesting, nonetheless.

There could be more variables added to the model to provide a more accurate prediction such as: days of rest for each team, length of their travels, and by simply obtaining more data throughout the course of the season

References

All data is retrieved from

www.basketballreference.com

Footnotes

[1]Add footnotes, if any, on their own page following references. The body of a footnote, such as this example, uses the Normal text style. *(Note: If you delete this sample footnote, don't forget to delete its in-text reference as well. That's at the end of the sample Heading 2 paragraph on the first page of body content in this template.)*

Tables

Table 1

Table Title

Column Head	Column Head	Column Head	Column Head	Column Head
Row Head	123	123	123	123
Row Head	456	456	456	456
Row Head	789	789	789	789
Row Head	123	123	123	123
Row Head	456	456	456	456
Row Head	789	789	789	789

Note: Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style, available on the Home tab, in the Styles gallery, to get the spacing between table and note. Tables in APA format can use single or 1.5-line spacing. Include a heading for every row and column, even if the content seems obvious. A table style has been setup for this template that fits APA guidelines. To insert a table, on the Insert tab, click Table.

Figures

Figure 1. Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.

For more information about all elements of APA formatting, please consult the *APA Style Manual, 6th Edition*.