ENGS 93
Statistical Methods of Engineering

Factors Impacting Kevin Durant's Scoring

November 3, 2023
Kiran Dsouza

Table of Contents

1. Introduction ……..2

2. Methodology ……3

3. Statistical Analysis ……...4

4. Conclusions ……..7

5. Future Steps ……..7

References ……8

Appendix ……..9

Introduction

Kevin Durant, also known simply as Durant, is one of the best players in the NBA right now. He has two NBA titles to his name and has also won an MVP award. However, even more interesting statistically is that Durant is arguably one of the most prolific scorers in NBA history. He averages 27.3 points per game and currently sits at 12th on the all-time scoring list. He is a hot topic among basketball fans to talk about since he brings drama to any organization he goes to. In particular, there has been much speculation about various factors that may impact how well he performs in any given game, and some of these deserve to be investigated further. More specifically, we will examine three factors to see if these have a statistical significance for Durant's scoring

The first factor is if Durant is playing in the same game as Stephen Curry. Steph Curry is considered the greatest shooter of all time in the NBA, and his resume is just as impressive as Durant's. These two players were on the same team for a while. Both of these players are very efficient in scoring, so it brings up the question: does Durant see a difference in the points he scores when playing alongside an elite shooter like Curry since there is only one basketball.

The next factor explored is whether Durant is playing against a "good", "bad", or "average" team. Durant has always been on great teams. When he plays, his team has yet to miss the playoffs. Some players, when they are on good teams, tend to play fewer minutes or not try as hard against teams who aren't as good, especially superstars since they have nothing to prove. On the other hand, other players who are trying to get personal awards such as the Most Valuable Player (MVP) like to stat pads. Stat padding is a term used in sports where a player plays in a way such that their statistics improve (such as PPG) despite these actions not helping, not helping said players' team win. An example of this is a player still trying to score in a blowout game.

The third and last factor this paper will explore is whether or not Rihanna is present at the game. ESPN had a segment a few years ago about how NBA players perform differently when the pop star Rihanna attends their games. For example, when Dwayne Wade played with Rihanna in the stands, his shot percentage went from a respectable 52% to a below-average 42%. In this segment, ESPN revealed Durant's average went from 32 points per game (PPG) to 36 PPG with Rihanna in attendance., This paper explores if there was any statistical significance to this ESPN story.

Methodology

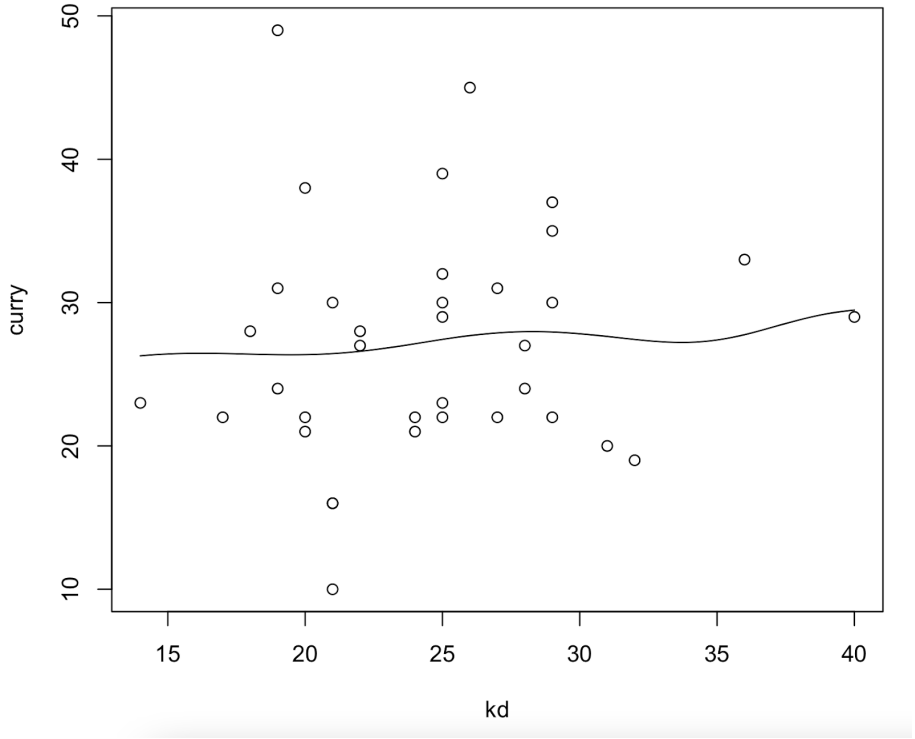
To perform this analysis, several data groups will be needed. This paper will be looking at the 2018-2019 season as it was the last season Durant was playing for the Golden State Warriors. ESPN has game logs for all players across their careers, so Durant and Currys' points and if they played with each other will be taken from there. ESPN also provides records of each team during this season, which will be used to classify a team as "average", "good", or "bad". For determining whether or not Rihanna was present at a game during this season, news outlets, and her Instagram feed will be used.

For analysis, hypothesis testing will be used. Graphs of the data will be used to see if the assumptions needed for their respective test are met. For example, data shouldn't be heavily skewed. In addition, we will assume that all games played are independent of each other.

Statistical Analysis

Curry & Durant's scoring

For this test, we will see if there is a relationship between the number of points Durant and Curry score in a game. For a visual, rather than go with a box plot as that wouldn't make sense for this inference procedure, a kernel regression will be used. In the plot below, each point represents a game both Curry and Durant (KD) played. Looking at all the points, it appears like there is no relationship between Curry and Durant's scoring at a glance. The only notable observation is there isn't any game where both Durant and Curry scored 35 or more points which makes sense as there are 48 minutes per game and only one basketball to go around. The regression line looks to be slightly positive, but it's very weak and doesn't look like any obvious function.



To test if there is a relationship between Durant and Currys' scoring I ran Kendall's test. Since looking at the data, it doesn't appear to have a monotonic relationship, and I am using a continuous variable, the assumptions are met, and I had no problem running this test. My two hypotheses are:

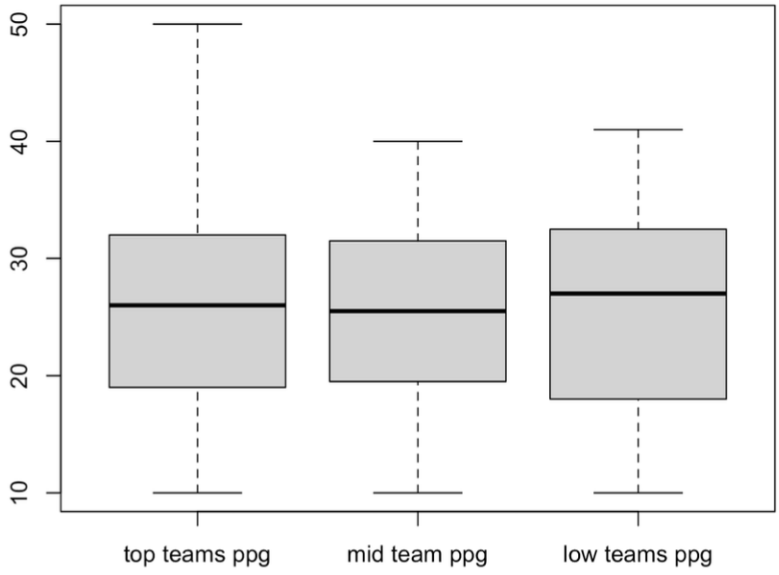
- the points Curry scores in a game and the points Durant scores in a game are independent
- the points Curry scores in a game and the points Durant scores in a game are associated with the null and alternative hypotheses respectively.

Running the test I got a p-value of .458 which is greater than the alpha level of .1, so can't reject null. Thus, there isn't enough evidence to conclude there is an association (positive or negative) between the points Curry scores in a game and the points Durant scores in a game.

Different teams influence on Durant

For the second inference procedure, we want to find out if the difficulty of the team Durant plays against has any relationship with how well he scores. To do this the NBA teams will be partitioned into three groups depending on how they did this season. The first group consists of the top teams, which are teams who landed in the top 4 in their respective conference. The second group, mid-teams, consists of teams who landed in spots 4-8 in their respective conference. The third group, low teams, are the teams who did not make the playoffs, meaning they placed outside of the top 8 in their conference.

The ANOVA F test will be used with its null and alternative hypothesis being that all 3 populations have the same mean and the 3 population mean aren't the same respectively. The plots below contain the three groups' box plots. Just at a glance, we can see the median looks about the same. Unfortunately, the top teams group and low teams group look right skewed and right skewed respectively, so for the ANOVA F test the results should be taken with a grain of salt.

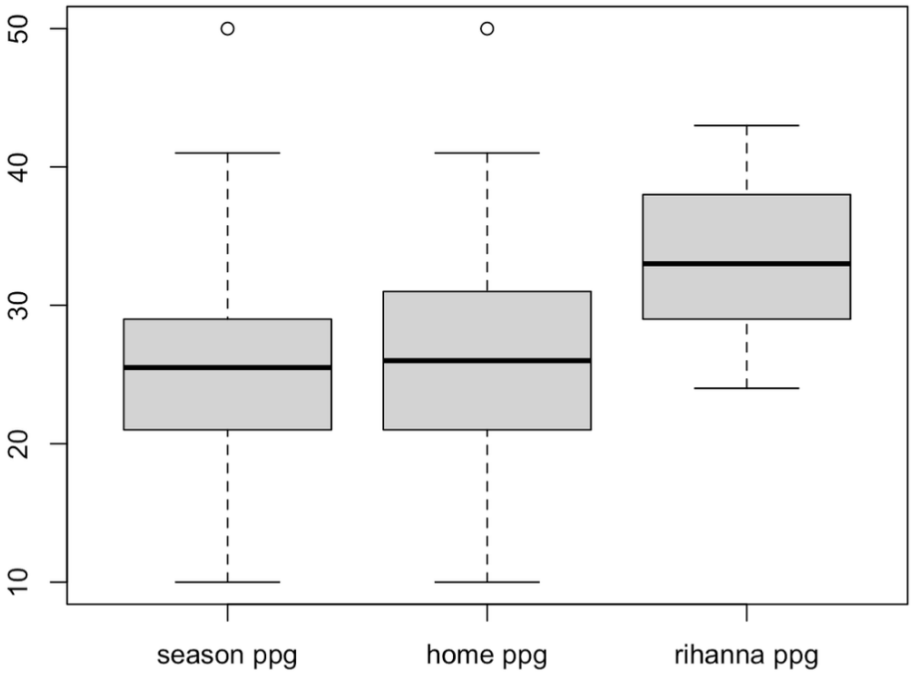


The ANOVA F test gave a p-value of 0.965, which is very high, and since they are all greater than the alpha level of 0.1, I can't reject the null, so I cannot conclude the mean or median of all the groups aren't different.

The Rihanna Effect

To test if Rihanna's attendance does help Durant score more, a simple hypothesis test with an alpha level of 0.1 will be performed. My null hypothesis is that Durant averages the same PPG when Rihanna attends the game, and my alternative hypothesis is Durant averages more PPG when Rihanna attends the game.

I have three samples: The points Durant scored at all the games he played during the season, home games in the season (since Rihanna only attended home games when Durant played for GSW) and when Rihanna participated in the home games. For the data, I'm assuming each game is independent of other games. I have summarized all the data in the box plot.



At a glance, we can see that all three samples seem to have a normal symmetric distribution, and the games with Rihanna in attendance do have a higher median. I want to know if these results are meaningful. For my tests, I decided to do a two-sample t-test.

Since the data is independent, normal, and symmetric, the assumptions for both tests are met. The result when comparing the games with Rihanna and the season games is that I got a p-value of 0.0037 for the t-test. The result when comparing the games with Rihanna and the home games is that I got a p-value of 0.0037 for the t-test. Since all the p-values are lower than my alpha level, for all tests I can reject the null and conclude Durant averages more PPG when Rihanna attends the game.

Conclusions

In this study, I statistically analyzed three factors that might impact Durant's scoring in NBA games, using data from ESPN, news outlets and Instagram. I found that I could not conclude that playing with Curry affects Durant's scoring, and neither could I conclude that from the strength of his opponent. However, I could conclude that having Rihanna present does improve Durant's scoring.

A big assumption I made is that games are independent of each other, but that may not be the case. For example, once a player "catches fire" he tends to stay on fire. An example of that is James Harden once went on a 30-game streak of scoring 30+ points. I also assumed normality for these hypothesis tests. Of course, it is impossible to have a controlled environment when running these tests since there are too many uncontrollable variables. For example, when comparing Curry and Durnats' scoring I would want them to play the same amount of minutes, against the same team, in the same stadium, and alongside the same player. This is very unrealistic. Since I cannot use a controlled environment, I only can conclude correlation and not causation.

Future Steps

Many factors influence how a player scores in the NBA. I only looked at three variables in my case study, but there are many confounding variables in basketball. A few examples include minutes played, coaching, injuries, rule changes, and play style. Some of these variables are very hard to quantize such as coaching and play style, so it can take a lot of work to get accurate results when working with these factors.

If we continue analyzing Durant's scoring, we could look across his entire career to get more sample points, and how his scoring changes across different teammates, different stages in his career and with different coaches. Of course, these ideas still face the same challenges as the ideas discussed in this paper as the NBA is constantly changing, so trying to hold variables constant is impossible.

References

1. ESPN. “Stephen Curry Career Stats - NBA.” *ESPN*, ESPN Internet Ventures, www.espn.com/nba/player/stats/_/id/3975/stephen-curry. Accessed 6 Nov. 2023.
2. ESPN. “Kevin Durant Career Stats - NBA.” *ESPN*, ESPN Internet Ventures, www.espn.com/nba/player/stats/_/id/3202/kevin-durant. Accessed 6 Nov. 2023.
3. ESPN. “James Harden Career Stats - NBA.” *ESPN*, ESPN Internet Ventures, www.espn.com/nba/player/stats/_/id/3992/james-harden. Accessed 6 Nov. 2023.
4. ESPN “NBA 2018-19 Regular Season Standings.” *NBA Team Standings & Stats | NBA.Com*, www.nba.com/standings?Season=2018-19. Accessed 6 Nov. 2023.

Appendix

```
#R code
season_points = c(20, 22, 29, 25, 29, 31, 28, 19, 24, 25, 21, 29, 21, 24, 27, 21, 29, 25, 19, 36, 22, 18, 25, 21, 27, 20, 25, 40, 26, 25, 32, 19, 14, 28, 20, 17, 34, 33, 31, 10, 17, 24, 28, 22, 32, 28, 19, 37, 39, 26, 10, 27, 29, 27, 41, 17, 13, 40, 29, 24, 26, 33, 24, 50, 35, 36)
 home_points = c(20, 29, 31, 28, 21, 29, 21, 28, 26, 22, 33, 18, 25, 21, 27, 20, 40, 14, 28, 20, 33, 24, 10, 17, 24, 28, 19, 37, 26, 10, 29, 41, 40, 29, 24, 26, 33, 24, 50, 35, 36)

Rihanna = c(39, 29, 24, 38, 33, 27, 43, 32, 34)

boxplot(season_points, home_points, Rihanna, names = c("season ppg", "home ppg", "Rihanna ppg"))
 t.test(rihanna, season_points, alternative="greater", var.equal = FALSE)


t.test(rihanna, home_points,alternative="greater", var.equal = FALSE)


library(jmuOutlier)
 library(correlation)
 top_teams = c(32, 19, 24, 14, 28, 20, 17, 34, 33, 31, 10, 17, 24, 28, 22, 32, 28, 19, 37, 50)
mid_teams = c(20, 25, 27, 40, 26, 25, 32, 19, 14, 28, 20, 17, 34, 33, 31, 10, 17, 24, 28, 22, 32, 28, 19, 37)
 low_teams = c(20, 17, 34, 33, 31, 10, 17, 24, 28, 22, 32, 28, 19, 37, 39, 26, 10, 27, 29, 27, 41, 17, 13, 40)
 boxplot(top_teams, mid_teams, low_teams, names = c("top teams ppg", "mid team ppg", "low teams ppg"))

data=c(top_teams, mid_teams, low_teams) groups= rep(1:3, c(20, 24, 24))
perm.f.test(data,groups)

%ANOVA F test was performed using Wolfram online

Durant = c(20, 22, 29, 25, 29, 31, 28, 19, 24, 25, 21, 29, 21, 24, 27, 21, 29, 25, 19, 36, 22, 18, 25, 21, 27, 20, 25, 40, 26, 25, 32, 19, 14, 28, 20, 17)
 curry = c(22, 28, 37, 29, 30, 20, 27, 31, 21, 22, 16, 22, 16, 22, 22, 10, 35, 39, 24, 33, 27, 28, 23, 30, 31, 38, 32, 29, 45, 30, 19, 49, 23, 24, 21, 22)

cor.test(Durant,curry,method='kendall')
 plot(Durant, curry)
 bw = 10 lines(ksmooth(Durant,curry,kernel='normal',bandwidth=bw))
```