Crunch Time: How NBA shooting changes late in games

Tyler Gorecki

Abstract

The National Basketball Association (NBA) is the American professional basketball league and is known as the best in the world. The top players all play in this league and there are many millions of fans who follow the game. A lot of games end up being decided by the shots at the end of the game when the scores between the two teams are close and one big play or mistake can be the difference between winning and losing. I researched whether, when the intensity is higher at the end of games, the shooting percentage of NBA players decreases or stays the same and the added pressure has minimal effect on players. I defined late in games as the last minute of game time in the fourth quarter and overtime. With my research and utilizing a two-sample proportion test, I am able to determine that the shooting percentage of all NBA players as a whole does decrease late in games.

Introduction

NBA players are always critiqued on their ability to perform in the clutch. There are many arguments that being clutch isn't a real idea and that it is just due to a smaller sample size for players late in games. I wanted to analyze if the shooting percentage of the league overall decreases late or if it stays relatively the same. In this paper, I plot NBA play-by-play data I acquire to analyze any specific trends and use statistical testing to compare the shooting percentages. I feel confident that my results will provide proper insight into answering my research question.

Data Description

This dataset is a collection of every play from the 2019-20 NBA season[1]. It was collected from a sports statistics database and the dataset itself contains over 500,000 rows and 40 columns of data. Prior to analyzing the data, the important shooting columns we identified are described below:

Variable	Description
Quarter	The quarter the play was made in
SecLeft	How many seconds are left in the quarter at the time of the play
Shooter	Player who shot during shooting plays
ShotType	Type of shot taken (five categories)
ShotOutcome	Outcome of the shot (make, miss, or nan)
ShotDist	Distance from the basket of the shot taken

Data Processing Methodology

Because this dataset was so large, we wanted to remove a lot of the data that wasn't of importance to us in answering our research question. The first thing we did was remove all plays that weren't 'shots'. This meant removing all rows that contained an 'nan' value in the 'ShotOutcome' column. This column contained values 'make', 'miss', or 'nan' so this row removal would leave us with only shooting plays. After that, we wanted to define what it meant to be 'late' or 'not late' in the game. We created a new column / variable called 'time' that took the value of 'late' if the shot was taken within the last minute of the fourth quarter or overtime and 'not late' otherwise. This created a binary categorical variable that will help us in finding an answer to our research question. The final data processing we performed was to narrow the columns down to only four that were of interest to us: 'ShotType', 'ShotOutcome', 'ShotDist', and our newly created column 'time'. With this resulting dataset, we now feel confident that we

have all of the information needed to analyze the difference between ‘not late’ and ‘late’ shooting percentages.

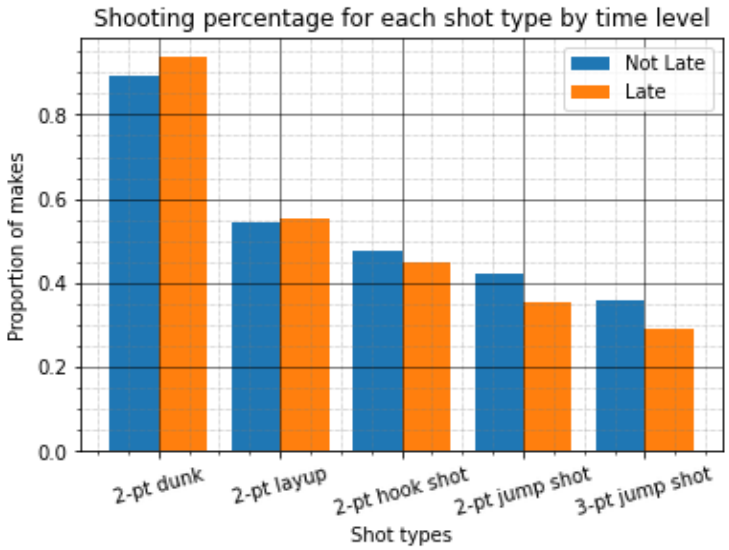
Results

Two-way Table

Shot Percentages	Not Late	Late
Make	46.10%	41.44%
Miss	53.90%	58.56%

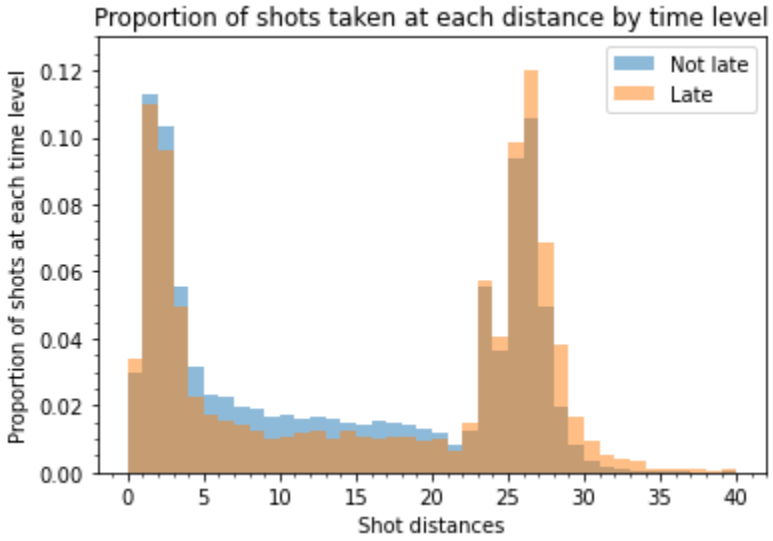
This two-way table provides a basic understanding of our data before doing any further visualizations or testing. We appear to see a difference between the percentage of shots made ‘not late’ in games and those shots made ‘late’ in games. This is something we will keep in mind as we continue our research.

Plot 1



From this plot, it is clear that there is a higher shooting percentage ‘late’ in games on dunks and layups, but a lower shooting percentage ‘late’ in games on hook shots and 2- and 3-point jump shots, in comparison with ‘not late’ shots.- When late in games, teams, especially if the game is really close, are more likely to play extra aggressive on defense. This means limiting any shot they can, resulting in harder defense played further from the basket. As a result, jump shots become better contested and harder to make, but this leaves opportunities for easier shots to be made on the inside (closer to the hoop) when there is a defensive breakdown. Our next plot provides more context to the analysis.

Plot 2



Plot two is a graph of two overlapping histograms showing the distribution of the distance of shots at each of our personally-assigned time levels. A sample way to interpret this is to understand from the plot that about one-tenth or 10% of the shots taken late in the game from this NBA season were taken from 25 feet from the basket.

This plot shows a clear difference between the distribution of shots taken not late and late in games. Deeper shots are taken later in the game more often, with the point of separation between the two distributions occurring just short of where the three point line stands, at 23 feet, nine inches. This is likely due to the fact that when teams are losing late in the game, they are more inclined to shoot three pointers to cut into the lead. Plot one, however, shows that the shot percentage decreases as the distance from the basket increases.

Two Sample Proportion Test:

From the two-way table and the visualizations, it appears that the shooting percentage is lower late in games. We will use a two-sample proportion test to analyze if this difference is significant enough to support that claim. The two samples will be all of the 'not late' and all of the 'late' shots from the 2019-20 NBA season. These will serve as samples of the overall population of all NBA basketball shots in the 2010s decade. In this decade, teams increased their volume of three-point shots as a result of modern analytics showing the increasing importance of threes (the average number of three point jumpers per game increased consistently year-to-year throughout the decade[2]).

Proportion one in this test comes from the sample of 'not late' shots and proportion two comes from the sample of 'late' shots. The proportions represent the percentage of made shots and can be found in the two-way table shown earlier in the report. I used the proportions_ztest function from the statsmodels package in python to run this test. In terms of assumptions, the data comes from a binomial distribution and the number of successes and failures are all much larger than five. The one holdup may be that all samples aren't completely independent from each other. There is no way to have a scenario where every shot measured in an NBA season is

exactly the same. Sometimes the defense is covering one player extra well, sometimes a player is left wide open, sometimes a player is shooting really well that specific game, etc., but I feel the randomness in this is relatively balanced enough that we can proceed with the testing procedure without worrying about illegitimate results.

Our null hypothesis states that the proportion of shots made ‘not late’ in games is equal to the proportion of shots made ‘late’ in games. Our alternative hypothesis states that the proportion of shots made ‘not late’ in games is greater than the proportion of shots made ‘late’ in games.

Proportion Test Code

```
# null, H0: p1 = p2
# alternative, Ha: p1 > p2
stat, p_value = proportions_ztest(count=successes, nobs=samples,  alternative='larger')
p_value # 3.437*e-11
# there is enough evidence to suggest that the proportion of successes in the
# sample of not late shots is greater than the proportion of successes in the
# sample of late shots
```

The resulting test statistic of this test was 6.523 and the p-value was thus very small (nearly zero). Therefore, there is enough evidence to reject the null hypothesis that the proportion of ‘not late’ and ‘late’ shots made are equal, and instead conclude that the overall shooting percentage ‘not late’ in games is greater than the overall shooting percentage ‘late’ in games.

Possible Changes

Before concluding, we would like to discuss three ways to improve this project in the future if we had more time and we were to do it again. First, we would filter our data to have a maximum point differential between the teams at the time of the shot to make sure the game is still meaningful and both teams are within reach. This could remove the situations where the game is already decided and teams are letting their opponent shoot wide open shots because it won’t have an effect on the final score. Second, we would analyze the differing proportion of made shots and differing distribution of shot distances between winning and losing teams. It would be interesting to measure whether there might be a clear difference between the two sides of the lead. Finally, we would perform a two-sample proportion test separately for each shot type. From plot one, there appeared to be some shot types where the shooting percentage was greater in ‘not late’ situations, and some where the shooting percentage was greater in ‘late’ situations, so that would also be an interesting analysis.

Conclusion

After proceeding with our proportion test and our visualizations, we feel confident concluding that the proportion of shots made ‘late’ in games is less than that of shots made 'not late’ in games. This means that if a random shot is taken ‘not late’ in the game, it would be expected that it would result in a make more often than when a random ‘late’ shot.

There are a few possible explanations for this difference, thinking from a basketball

perspective. Late in games, especially if the game is close, players tend to play harder on defense and thus make it tougher to get open shots (or good shots at all). Another consideration is that when teams are down late in games, they are more likely to shoot earlier in the shot clock and thus might not get the best possible shot for their team to save time in the game altogether. Finally, late game shots are always higher pressure when the score is close and one shot could be the difference between winning and losing the game. With these ideas in mind and the research done, this result is understandable.

This data can be used by NBA teams to analyze what shots are most effective late in games, when they need a shot to win. Obviously, there is a chance taken with everything but if a player can get a higher percentage shot according to the data provided, it puts the teams in a better position to win. This not only helps offensively, because, defensively, teams can use the same data to try to counter what the offense is trying to do and force them into tougher shots. With deeper analysis as described above, teams could use this data to win more games. That could end up being the difference between winning and losing a championship game, as many of them come down to shots being made or missed in the last minute of regulation and/or overtime.

Resources

[1]*https://sports-statistics.com/sports-data/nba-basketball-datasets-csv-files/*

[2]*https://shottracker.com/articles/the-3-point-revolution* - The number of three pointers taken per game has increased almost every year since the three-point line was created, but a clear spike began in the 2010s with the increased use of basketball analytics.