MAT022: Foundations of Statistics and Data Science

Bailey Millns - C2098580

09 February 2021

Abstract

After every NBA regular season, a most valuable player (MVP) award is given to the best player that season. In the 2014-15 season Stephen Curry was awarded the MVP award. This report aims to give an insight into some of the reasons why Curry was awarded the MVP award, as we try to quantify his MVP season. We look at Curry's three-point shooting percentage and find that, using a Z test, he appears to make a higher proportion of three-point shots compared to the rest of the NBA. We perform a T test to compare the top three in MVP voting, and the test suggests Curry is a more efficient player than LeBron James and James Harden. Ultimately, we conclude with a non-parametric Wilcoxon signed-rank test which suggests the result that Curry may not be affected by high pressure situations.

Contents

1 Introduction

The national basketball association (NBA) is a professional basketball league in North America. The league is comprised of 30 wonderfully named teams, such as the Denver Nuggets and the Utah Jazz. The teams are divided into two conferences: The Western Conference and The Eastern Conference and are divided (roughly) based on geographical location. A Basketball season is split into two parts the regular season and the playoffs, the playoffs are held after the regular season to determine the league champion (Wayne, 2020).

Every regular season the Most Valuable Player (MVP) award is given to the best regular season player. In the 2014-15 season Stephen Curry was awarded the MVP award, with James Harden and LeBron James coming in second and third, respectively. Stephen Curry is a six-foot three point guard, who plays for the Golden State Warriors. Curry was the leader of the team with the best record of the season (a franchise record of sixty-three wins), he was the first player in fifty-five years to win the MVP award playing for the Warriors. The Warriors scored nine-hundred and twenty more points than they allowed with Curry on the count, the highest of any player that season, all of this with the lowest minutes played ever by an MVP. However, his most impressive season achievement is he managed to break the record for the most three-point shots made in a single regular season (noa, 2015). Stephen Curry revolutionized the game of basketball with his three-point shot. Never in the history of the NBA has the three-point shot been utilised as much as the modern NBA, and this is, in large parts, thanks to Curry.

In this study we will focus on Stephen Curry and his MVP season. We will look at his three-point shooting in more depth, as we compare his three-point percentage against the rest of the NBA. This is conducted using a Z test. A lower tail two-sample Z test tests the null hypothesis that the proportion of group A is greater than or equal to the proportion of group B.

Next, we will investigate the top three player in the MVP voting, where we will compare the points per shot (PPS) of Stephen Curry vs the combined PPS of LeBron James and James Harden. PPS is a player efficiency metric calculated by taking the total points from two-point shots and three-point shots and dividing that by the number of shots attempted. This measure favours players who are efficient not just great scorers (Tolnick, 2012). We will compare the results using a two-sample T test. A lower tailed two-sample T test tests the null hypothesis that the mean of sample one is greater than or equal to the mean of sample two.

Finally, we try to answer the question: is Stephen Curry's shot affected by pressure? We compare the mean shot percentage of Curry in high pressure moments against regular moments to see if they are equal. The data obtained did not fit the assumptions of a T test so therefore we will conduct the test with the non-parametric Wilcoxon signed-rank test. The two-tailed Wilcoxon signed-rank test tests the null hypothesis that the difference in the means of the two groups is equal to zero.

The data used for this analysis is a data set of the shots taken by players in the NBA between October 2014 and March 2015, and originally contained 128069 observations on 23 variables. A description of the data set and a summary on how the data was cleaned can be found in this notebook.

2 Three-Point Shooting Percentage: Curry and the NBA

The three-point shot is the most valuable shot in basketball so that is where we will begin the analysis. Shown below in figure 1 is each player's three-point percentage against their three-point attempts, while filtering for the players who qualify. To qualify for three-point percentage the player must have at least 82 three-point field goals made in the season. There appears to be an overall positive correlation between percentage and attempts, and we can see the top three-point shooters of the season. The size and colour of the bubble is based upon the number of three-point shots made by each player. We can see that Stephen Curry has the most three-point shots made in the season.

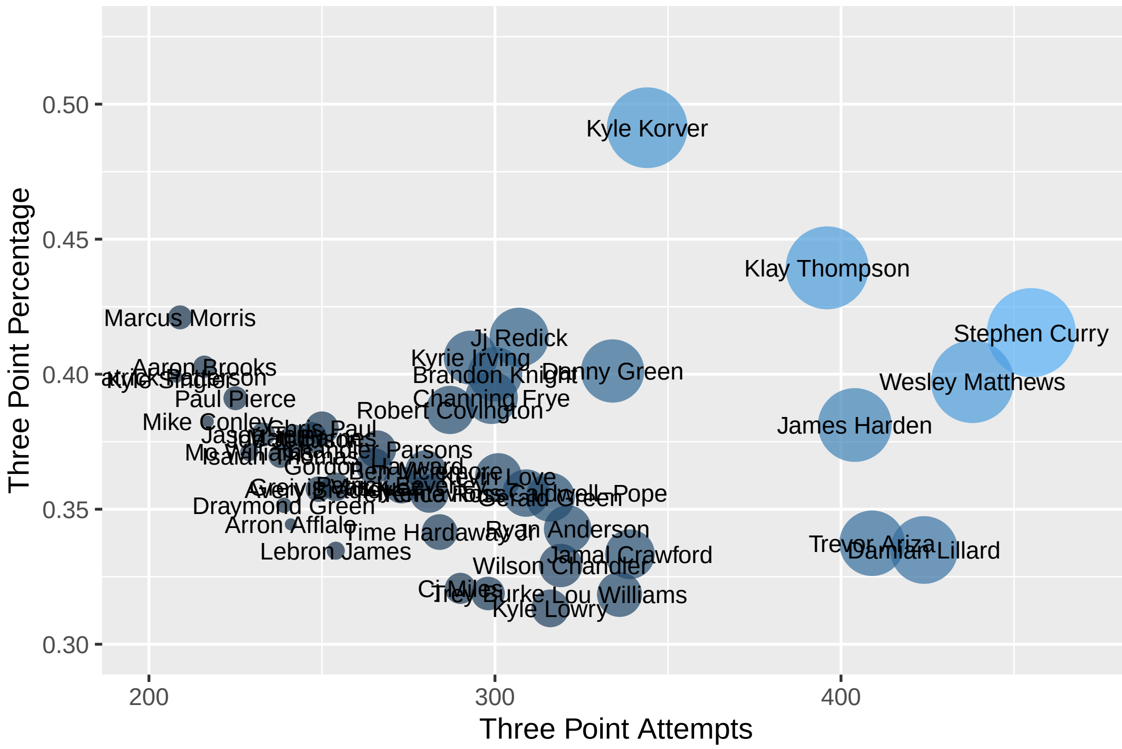

Figure 1: NBA three-point shooting visualisation

2.1 Z Test to Compare Proportions

We will now perform inferential analysis tests to try and quantify Stephen Curry's MVP season. We will first conduct a Z test to test whether the proportion of made three-point shots are larger for Stephen Curry compared to the rest of the NBA. The null hypothesis is:

$$H_0 : p_{SC} \geqslant p_{NBA}. \tag{1}$$

and the alternative hypothesis is:

$$H_1 : p_{SC} < p_{NBA}. \tag{2}$$

Where p_{SC} is the proportion of three point shots made by Curry and p_{NBA} is the proportion of three point shots made by the rest of the NBA.

However, to conduct a Z test both samples must be normally distributed. Since the player either makes or misses the shot the data is described by a binomial distribution, and when the samples are large the binomial distribution is well approximated by the normal distribution, due to the central limit theorem. The central limit theorem states that "the distribution of the sample approximates a normal distribution as the sample becomes larger assuming that all samples are identical in size, and regardless of the population distribution shape" (Ganti, 2019). The Z test was conducted and the results are found below, in table 1.

Table 1: Results from the Z test.

Measurement	Results
Estimated proportion of Curry's three-point made shots	0.4153846
Estimated proportion of the NBA's three-point made shots	0.3509162
P-Value	0.9978815

Table 2: Results from the F test between Curry's three-point data and the NBA three-point data.

Measurement	Results
Estimate for the ratio of variances	1.0684612
P-Value	0.3062384

We obtain a p-value of 0.9979 which is considerably above the significance level of $\alpha = 0.05$, for instance, we therefore retain H_0. We thus conclude that the data strongly suggests that the proportion of Stephen Curry making a three-point shot is greater than the rest of the NBA making a three-point shot. We can support this result by conducting a T test.

2.2 T Test to Compare Proportions

To conduct a T test the variance of both samples must be equal, this can be checked with an F test to compare variance. The result can be found in table 2. We obtain a p-value of 0.3062 which is greater than the significance level of $\alpha = 0.05$. In conclusion, the F test suggests there is no significant difference between the two variances. We are able to now conduct a T test, where the null and alternative hypotheses are still equations 1 and 2. The T test result is found in table 3

Table 3: Results from the T test to compare Curry's three-point data and the NBA three-point data

Measurement	Results
Estimated mean of Curry's three-point percentage	0.4153846
Estimated mean of the NBA's three-point percentage	0.3509162
P-Value	0.9970815

We obtain a p value of 0.9971, which again at significance level of $\alpha = 0.05$, we retain H_0. From this result we might conclude that the proportion of Stephen Curry making a three-pointer is higher than the rest of the NBA.

3 Points per Shot: MVP Finalists

Basketball fans love to look at the post game statistics and see how many points their favourite player gets in a game or a string of games. Points per game and total points are great statistics to measure a player's performance but can often be misleading. Would you rather a player has 50 points but missed 80% of his shots or a player with 40 points but only missed 50% of his shots? Efficiency is important, so efficiency metrics like true shooting percentage and points per shot can paint a broader picture of a player's offensive impact. As the data set given does not include free throws, we will focus on the points per shot metric. Seen below in figure 2 is a histogram of all the NBA players points per shot. Stephen Curry is 13th on the list of the top points per shot players. This is even more impressive when you consider that 13 of the top 15 are centres, who do most of their work around the basket, whilst Stephen Curry's shots are primarily jump shots.

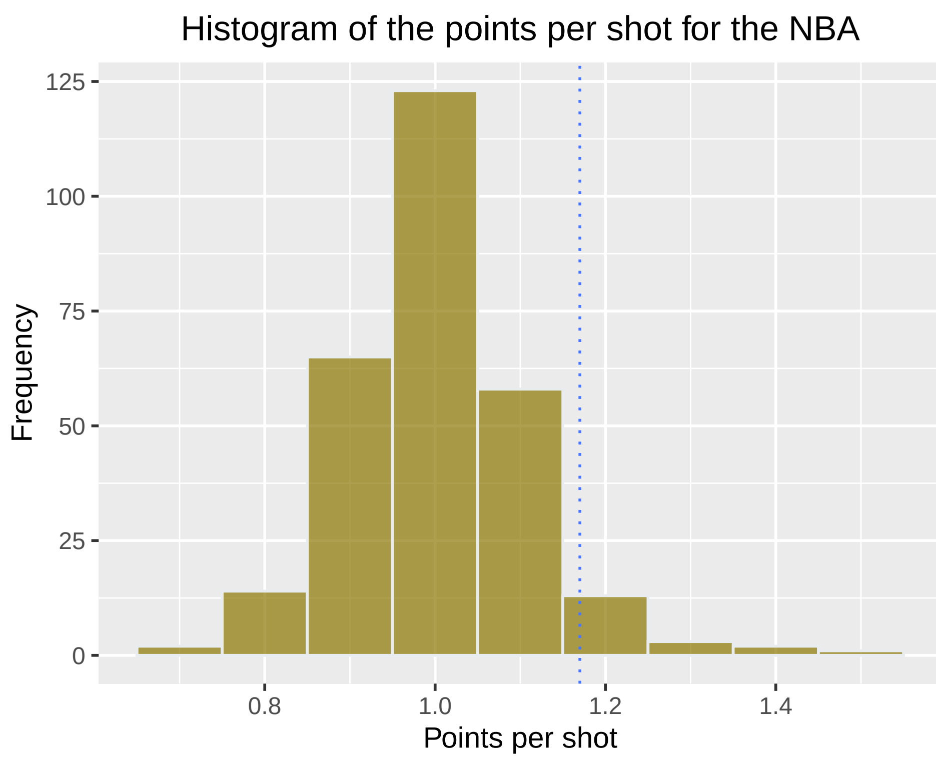

Figure 2: Histogram of the points per shot for all NBA players.

3.1 T test to Compare MVP Finalists

We will now compare the top three in MVP voting to see if Curry has significant advantage in PPS. We will use a T test to compare Stephen Curry's PPS per game vs LeBron James and James Harden's combined PPS per game. The null hypothesis of this test is that Stephen Curry's PPS per game is greater than or equal to LeBron James and James Harden's PPS per game, and so alternative hypothesis is that Curry's PPS per game is less than James and Harden's PPS per game.

The two-sample T test has the assumptions that both sample data are normally distributed and the variances of the two samples are equal. The normality assumption will first be tested by looking at the histograms of both samples and then a Shapiro-Wilk normality test will be conducted. To test if the samples have an equal variance, the F test to compare the variance will be used to test the assumption. The histogram of Stephen Curry's PPS per game and LeBron James and James Harden's PPS per game is plotted in figure 3. Both histograms appear approximately normal, however, to support this claim a Shapiro-Wilk normality test has been conducted for both samples. Both results can be found in table 4.

Table 4: Results from the Shapiro-Wilk test.

Measurement	Curry_PPS	James_And_Harden_PPS
P-Value	0.1640065	0.7809302

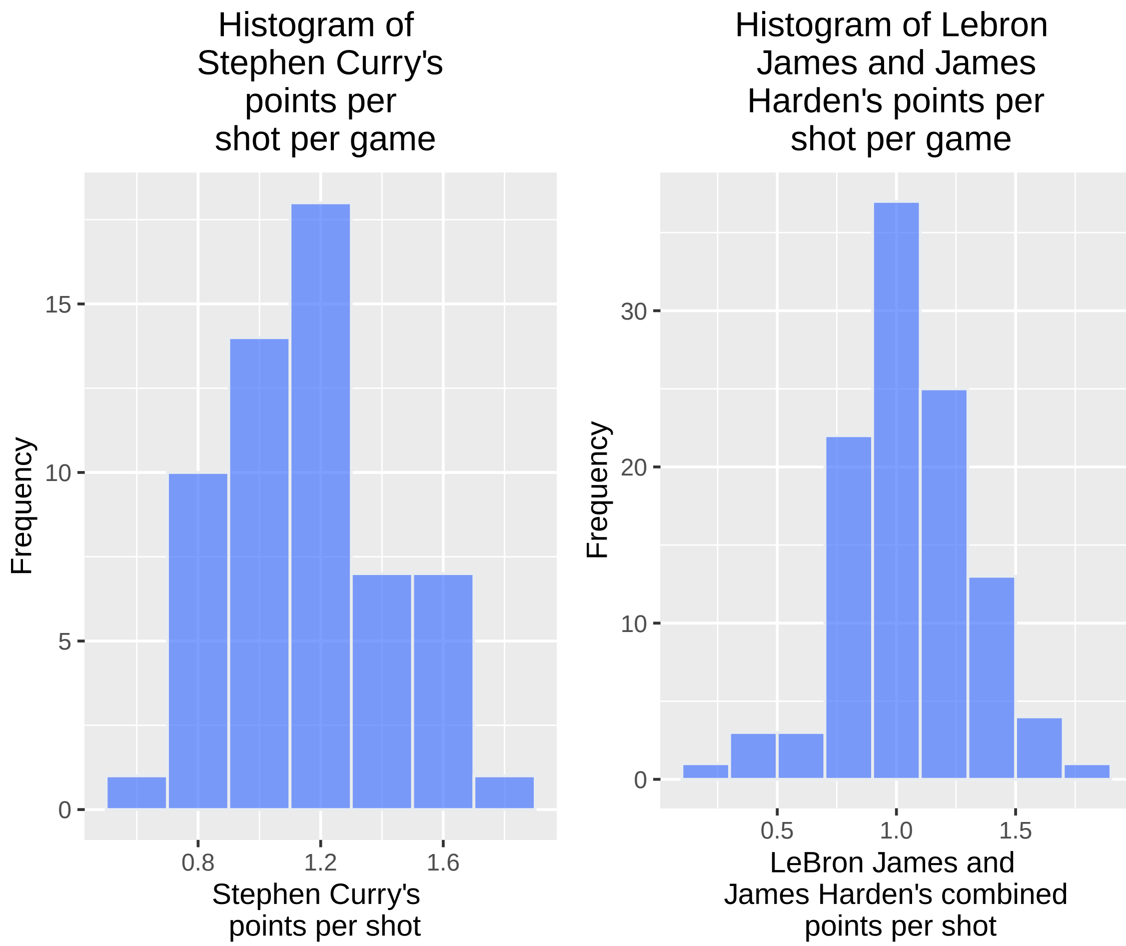

Figure 3: Histogram of Stephen Curry's PPS and histogram of LeBron James and James Harden's combined PPS.

We obtain a p-value of 0.164 for the Stephen Curry PPS per game data. At the significance level $\alpha = 0.05$ for instance, we retain H_0 and we thus conclude that you cannot reject the hypothesis that the sample comes from a population which has a normal distribution. Likewise, for the LeBron James and James Harden's PPS per game data. We obtain a p-value of 0.7809. At significance level $\alpha = 0.05$ we retain the null hypothesis, that we cannot conclude that the sample data does not come from a normal distribution. In addition, both have large sample sizes of 58 and 109, respectively. Therefore, the central limit theorem is in effect. Now we need to check if both sample have equal variance. The results of the F test to compare the variances is found in table 5.

We obtain a p-value of 0.7381 which is greater than the significance level of $\alpha = 0.05$. In conclusion, the F test suggests there is no significant difference between the two variances. We are now able to conduct the T

Table 5: Results from the F test between Curry's PPS and James and Harden's PPS.

Measurement	Results
Estimate for the ratio of variances	1.0744817
P-Value	0.7381446

Table 6: Results from the T test to compare Curry's PPS and James and Harden's PPS.

Measurement	Results
Estimated mean of x	1.1649235
Estimated mean of y	1.0526254
P-Value	0.9920107

test, the result can be located in table 6.

We obtain a p value of the two-sample T test equal to 0.9920. This is much greater than the significance level of $\alpha = 0.05$. This strongly suggests that the PPS of Stephen Curry is greater than the combined PPS of James Harden and LeBron James.

4 The Clutch

"The clutch" in basketball is the time in the game where the pressure is at its highest, usually in the last few minutes of a close game or the last second on the shot clock (Zhang, 2019). Clutch moments often define a player's legacy, for instance Michael Jordan hitting "The Last Shot" to win his 6th NBA championship which cemented his place as one of the greatest players of all time. Seen below, in figure 4, is a histogram of the field goal percentage of NBA players in the clutch. We have defined the clutch as being last five minutes of the last period in a game that finishes with a final margin less than eleven or a shot taken in the last second of the shot clock. Stephen Curry is ranked 194 out of 281 NBA players in field goal percentage in the clutch. Does this suggest Curry is a worse player in clutch moments? We will conduct a two-sample T test to test whether Curry's field goal percentage is the same in the clutch as his percentage not the clutch. The null hypothesis is:

$$H_0 : \mu_{clutch} = \mu_{regular}. \tag{3}$$

and the alternative hypothesis is:

$$H_1 : \mu_{clutch} \neq \mu_{regular}. \tag{4}$$

Where μ_{clutch} is the mean field goal percentage of Curry in clutch moments and $\mu_{regular}$ is the mean field goal percentage of Curry in regular moments.

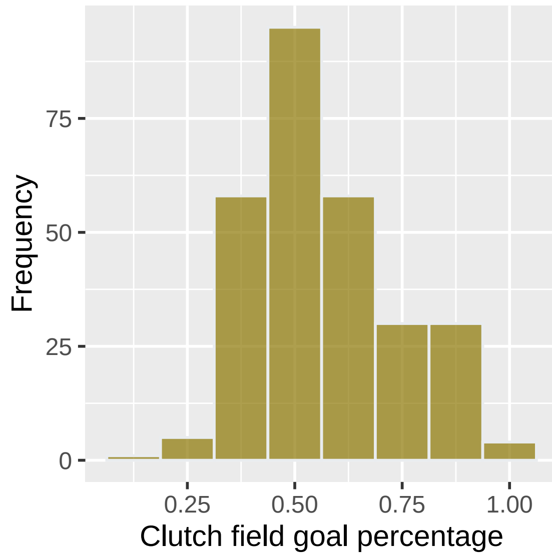

Figure 4: Histogram of the field goal percentage in the clutch for the whole NBA.

To conduct a T test the data has to be normally distributed, we can check this by plotting a histogram. The histograms of Curry's field goal percentage in the clutch and in regular time can be found in figure 5. The histogram for the clutch data does not appear to resemble a normal distribution. We can conduct a Shapiro-Wilk test to further test if we have normally distributed data.

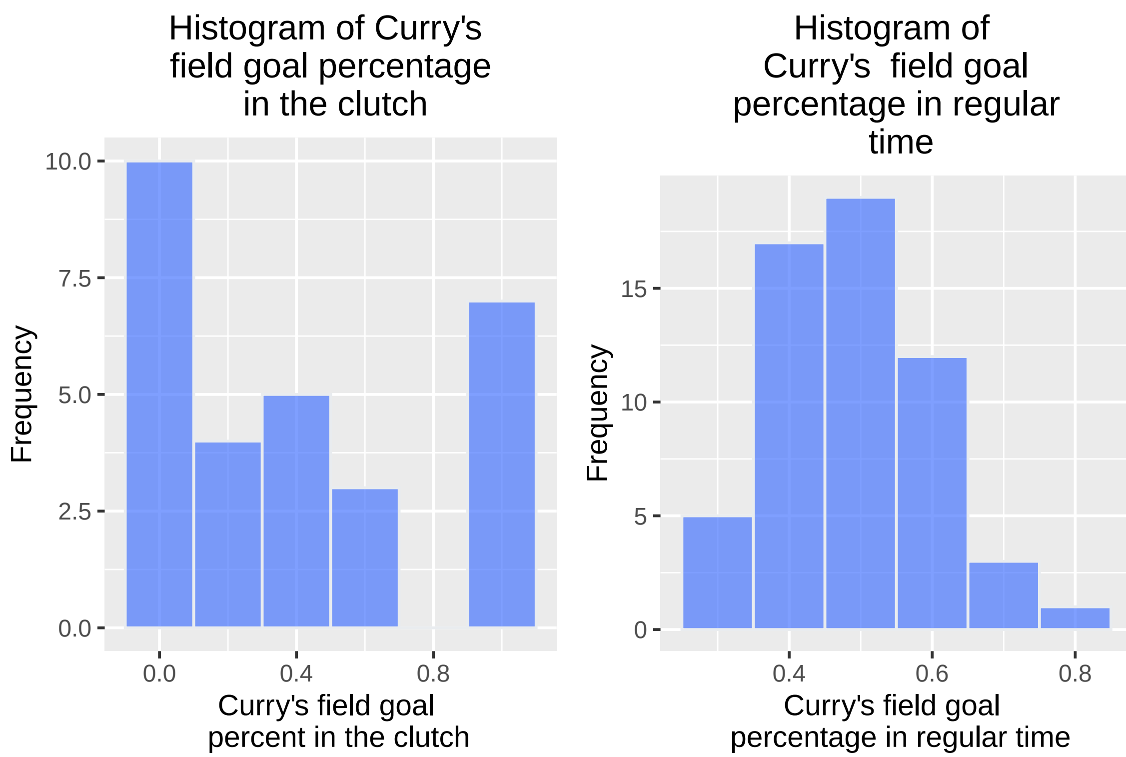

Figure 5: Histogram of the field goal percentage in the clutch and in regular time for Stephen Curry.

Table 7: Results from the Shapiro-Wilk test.

Measurement	Clutch	Regular
P-Value	0.000277	0.8096102

Table 8: Results from the Wilcoxon signed-rank test to compare Curry in high pressure and regular moments.

Measurement	Results
P-Value	0.071463

We obtain a p- value of 0.000277 for the clutch data which, at significance level of $\alpha = 0.05$, we reject the null hypothesis that the data is approximately normally distributed. Also since the clutch data is a relatively small sample size of 22 is not sufficiently large enough for the central limit theorem to apply.

4.1 Wilcoxon Signed-Rank Test to Compare Curry in High Pressure and Regular moments

Therefore, we cannot conduct a two sample T test but we can conduct the non-parametric Wilcoxon signed-rank test. The result of the Wilcoxon signed-rank test is located in table 8.

We obtain a p value of 0.07146 which, at $\alpha = 0.05$, is greater than the significance level so we therefore retain H_0. There is evidence to indicate that we cannot conclude that a significant difference exists between Stephen Curry's shooting percentage in clutch moments and in the regular time.

5 Conclusion

We have conducted several inferential statistical tests to give insight into why Stephen Curry was awarded the MVP award. The results of the Z and T test strongly suggested that the proportion of Curry making a three-point shot is, strongly suggested to be, higher than the rest of the NBA. Since Curry is quite often regarded as the greatest shooter of all time I would like to explore this notion. I would analyse Curry and other historically great three point shooters, over their careers, to see if I could obtain a significant result to suggest the greatest shooter of all time.

We have found that the result of a T test strongly suggested that the points per game for Stephen Curry is greater than the combined PPS of LeBron James and James Harden. Which tells us that Curry may be a more efficient offensive player compared to the runners-up in the MVP voting and this, probably, is one of the many reasons Curry was awarded the MVP above the other two players. Although, as stated above, the data set does not include free throws which I know is a large part of James Harden's game. In further work I would obtain the data of the free throws attempted for this season so that I am able to calculate the true shooting percentage, which is a more complete efficiency metric.

Finally, we tried to conduct a two-sample T test, but the sample data was not normally distributed, so we conducted its non-parametric counterpart the Wilcoxon signed-rank test. The result suggested that we cannot conclude that there is a significant different in Curry's field goal percentage in clutch moments and regular moments. We did obtain a quite small p value ($p = 0.07146$) which at $\alpha = 0.01$, for instance, we would reject H_0 in favour of the alternative hypothesis. Maybe we could obtain a more convincing result if the sample size increased to include the clutch data in the playoffs. Also, I would like to obtain the data that shows the score of the game at the time the shot was taken, so that I can more accurately test Curry's "clutchness", instead of looking at the final margin of the game.

Our conclusions should be checked with data from other sources, as inconsistencies and errors were found within the data set.For instance, the data set is approximately missing around 20 games, which could possibly alter the conclusions of my results. More details are again found in this notebook.

References

(2015). Warriors Guard Stephen Curry Named 2014-15 Kia NBA Most Valuable Player.

Ganti, A. (2019). Central Limit Theorem (CLT).

Tolnick, M. (2012). Why not Points Per Shot?

Wayne, J. (2020). NBA 101: Beginner's Guide to How the NBA Works.

Zhang, L. (2019). What is "Clutch Shooting" in Basketball?