Investigation of NCAA Basketball's Three Point Strategy Using Logistic Mixed Effects Regression Model

Che Hoon, Jeong

Denison University

Dr. Sarah Supp

Data Analytics Seminar

December 12, 2022

Introduction

I. Previous Research

Modern basketball has undergone drastic changes in the way teams strategize their plays. Notably, the National Basketball Association (NBA) has observed a surge in three point shot attempts. Teams such as the Golden State Warriors attempt high volumes of three point attempts (3PA), in which they ultimately broke the 3PA record during the 2018-2019 season (Freitas, 2021). Leveraging the three point shot, the Golden State Warriors displayed a dominant performance in the league. For the past 8 seasons since 2015, the team advanced to the NBA finals 6 times, and won the championship 4 times. The adoption of the three point shot strategy is also reflected in other teams such as the Houston Rockets. The team frequently used a “small ball” strategy, by utilizing athletes who are shorter but better shooters. Subsequently, they attempted more three point field goals than two point field goals during the 2017-2018 season (Freitas, 2021). Researchers have conducted studies aggregating and analyzing NBA shot records to investigate the efficacy of the three point shot and its overall impact in winning games.

One study conducted the Friedman's ANOVA test that revealed the increase in the coefficient of the 3PA over field goal attempts were statistically significant at the 5% level, after filtering out playoff games from the most recent 15 seasons until 2020 (Jaguszewski, 2020). In addition, the study presents that since the 2004/05 season, every team with the highest 3PA/FGA coefficient had at least a 50% win percentage, with an average winning rate of 66% (Jaguszewski 2020). Moreover, the greatest increase in the three point shot frequency in the league was between the 2056/16 and 2016/17 season, in which the Golden State Warriors attained the greatest number of regular season wins record (Jaguszewski, 2020). In other words, the NBA has observed a significant increase in three point shot attempts, and an increase in the importance of three point shots.

Taking note of the success of the three point strategy, sports media companies such as ESPN reported case studies of how collegiate basketball followed suit (Medcalf, 2018). In the first two rounds of the 2018 NCAA tournament, players have made a record-breaking number of 716 three pointers (Medcalf, 2018). Moreover, Villanova has capitalized on the warriors' "small-ball" three point strategy, in which they won the 2016 and 2018 NCAA championships (Medcalf, 2018).

II. Gap in the Literature

Despite the abundance of research on the three point strategy in professional basketball, notably the NBA, there is limited research conducted at the collegiate level. Most analysis on collegiate basketball has been focused on case studies of individual teams. Thus, this research conducts a quantitative analysis on the three point strategy in the NCAA Division 1 Basketball league.

The NBA has observed a new trend, in which sharp-shooting players such as Stephen Curry attempt three point shots at greater distances (Powell, 2022). It is not uncommon for players to shoot shots thirty feet away from the basket or near the center circle (Powell, 2022). Despite this phenomenon, there are no studies exploring whether college basketball players are also attempting three point shots at greater distances throughout the years. Therefore, this paper tests whether there is an increase in average three point shot distance in the NCAA.

III. Research Question

This research investigates whether NCAA college teams' three point shot attempts and three point made shots significantly increases the likelihood of winning. Moreover, it tests whether collegiate players are increasingly taking three point shots at greater distances over time.

Methods

I. Data

The NCAA Division 1 data was accessed via SQL commands from the Google BigQuery cloud database uploaded by the NCAA and Sports Radar (*NCAA Basketball*). Play-by–play information was queried from the 2009-2010 season to the 2017-2018 season. Game box score seasons are available from 2013 to the 2017 season. The Google BigQuery cloud database has a Google Drive file download size limit of 1GB and a local CSV file download size limit of 10 MB. The NCAA shot log data has a total file size of 3.43GB. Therefore, the queried results were saved as a BigQuery table, which were then exported to Google Cloud Storage. Details of this process could be found via the Google Cloud Big Query online documentation (Google). All data is publicly available and can be accessed by anyone. Details of SQL data retrieval and raw dataset can be accessed via *appendix I*.

II. Exploratory Data Analysis

A. Shot Chart Visualization

One of the most revolutionary advancements in NBA basketball analytics was created in 2012 by Kirk Goldsberry, who introduced the basketball shot chart, a heat map visualization that presents shot performance of players in a court (Beshai). The shot chart can be utilized to identify average shot distributions across each position, team, and individual player. In addition, shot charts may be used to identify a player's shooting performance against specific teams, which players have the highest or lowest efficiency in different areas of the court, and compare a player's shot distribution and performance against the league's or position's average (Beshai). Using this information, technical staff and coaches are able to better define player behaviors and devise future game strategies, and management teams may base future decisions on the metrics gained from such analysis (Sarlis, 2020).

Basketball shot charts are prominent visualizations that are utilized to analyze spatial distribution of shots across the court (Reich et al., 2006). Therefore, shot charts were used in this study to conduct exploratory data analysis and gain visual insight into shot distributions in the NCAA. The shots were created by modifying python code that generates shot charts for NBA Data (Tjortjoglou, 2015). The python code utilizes the Matplotlib package to produce the plots (Hunter, 2007). Shot chart visualizations present half-court representation instead of a full basketball court. The NCAA data presents shot locations through x-coordinates and y-coordinates of a full court. Therefore, NCAA shot coordinates were transformed such that it represents the half-court specification via the following formula:

$$x\,coordinate = 1128 - x\,coordinate\ ,\ y\,coordinate = 600 - y\,coordinate$$

Where 1128 is the width of the court and 600 is the height of the court.

B. *Three Point Shot Distance Trend Across Seasons*

NCAA games generally take place from November through March, with some games taking place during April, May, and October on rare occasions. Specifically, in the NCAA shot data, there were at least 10,000 shot observations from the months of November through March.

However, during seasons where there were games during April, May, and October, all months recorded less than 500 shots during each month. Thus, the dataset was filtered such that only games from November through March were used.

In investigating the increase in three point shot distance, the dataset filtered out for corner threes. Although there is no formal definition of long-range, or “deep”, three point shots, it is generally referred to as shots taken further beyond the arc and not from the corner line. For instance, an article from ESPN treats “deep” three pointers as shots taken at least 3 feet beyond the three point arc (Goldsberry 2019). Moreover, the corner three is bounded by the sideline, which prevents players from shooting at greater distances as they would step out of bounds. Therefore, corner three point shots, with x-coordinate values less than 119, were filtered out.

The three point shot distance was generated by using the Euclidean distance between the shot coordinates and the coordinates of the hoop:

$$distance = \sqrt{(shot_{xcoord} - 63)^2 + (shot_{ycoord} - 300)^2}$$

$$where\ (63, 300) = hoop\ coordinates$$

III. Statistical Model

A. Model Choice

The statistical model that was used to investigate the three point strategy’s impact on win likelihood is the Mixed Effects Logistic Regression. The dependent variable, win, takes a binary value of 1 or 0, indicating a win or loss, respectively. Therefore, a logistic regression was utilized to predict the binary outcome instead of a linear regression model. Considering the NCAA box-score data contain repeated measures from the same team across different years, a mixed effects model was utilized to take into consideration the random variability that occurs throughout different years (UCLA Advanced).

B. Variables

The sixteen variables of interest in the model are key performance metrics used in basketball: three point shot attempts, three point shots made, two point shot attempts, two point shots made, free throw attempts, free throws made, offensive rebounds, defensive rebounds, assists, turnovers, steals, blocks, personal fouls, fast break points, second chance points, and points off turnovers.

C. Logistic Mixed Effects Regression Assumptions

1) Outliers

The model is sensitive to unusually large or small values. Summary statistics and boxplots of all relevant variables were produced to determine whether observation needed to be dropped. True outliers were kept since outstanding performances are nevertheless representative of basketball games. Only outliers that appear as data input errors were dropped.

2) Multicollinearity

Issues of multicollinearity were tested using Variance Inflation Factors (VIF) values, which measures the severity of multicollinearity. Logistic Mixed Effects models may fail to converge with the presence of severely correlated variables. Thus, variables with VIF values greater than 5, which indicate issues of multicollinearity, were removed from the model.

3) Linearity

The assumption of linearity was tested by observing whether the nature of log of probabilities and the predictor variables are linear. Scatter plots with overlaying regression lines were generated to observe this relationship.

IV. Statistical Tests

A. Seasonal Mann Kendall Test

Seasonal Mann Kendall Test was conducted to observe whether a statistically significant upward trend in three point shot distance exists. The assumptions of the seasonal Mann Kendall Test were met. The data does not contain covariates, factors that influence the data that is being plotted. Moreover, there are only one data point per time period. The seasonal Mann Kendall Test was used instead of the Mann-Kendall test, because the data was collected seasonally from the months of November through March.

B. Wald Test

The wald statistic was used to evaluate the joint significance of the variables of interest - three point attempts and made three points. The Wald test approximates the Likelihood Ratio test, which compares the fit of one model to another (UCLA). Specifically, it tests the null hypothesis that a set of parameters, the coefficients of interest, are simultaneously equal to zero. In other words, it tests whether the set of variables are jointly significant in the model. If the test fails to reject the null hypothesis, it suggests that the removal of the variables does not harm the fit of the model (UCLA).

V. Model Evaluation

A. Train-Test-Split

Model performance was evaluated using the train-test-split validation procedure. The train-test-split tests how accurately a model can predict a new dataset. 70% of the data were used in the training set and 30% were used in the test set.

B. Confusion Matrix

The confusion matrix was used to evaluate how accurately the model classifies data in a two-by-two grid. It displays the number of observations that were correctly classified as a win, falsely classified as a win, correctly classified as a loss, and falsely classified as a loss.

C. Receiver Operating Characteristic (ROC) Curve

ROC curve was used to evaluate the model's accuracy. The ROC curve helps determine how accurately a model classifies data, in which a greater area under the curve indicates better model performance (Zou, 2007). The area under the curve (AUC), ranges from 0 to 1. A perfect model will have an AUC value of 1, whereas a model that fails to predict every outcome is 0. A model with an AUC value of 0.5 performs equally as random guessing.

Results

I. Exploratory Data Analysis

A. Shot Chart Visualization

Section *A* and *B* of *Figure 1* below are shot charts of Duke University's basketball team during the 2013 and 2017 seasons. Duke University is one of the powerhouse basketball programs in the NCAA, with a total of 11 national championship appearances. Exploratory analysis was conducted with Duke's shot data considering its consistent performance. As seen in the shot charts, Duke attempted more three point shots that are further behind the three point arc during the 2017 season than the 2013 season. Moreover, we can observe greater volume of three point shots taken during the 2017 season as shown by the greater number of data points. In addition, it can be observed that the team took much more shots near the basket and behind the three point line, compared to mid-range shots.

The University of North Carolina (UNC), who were the 2017 NCAA Division 1 champions, display similar patterns. As shown in sections *C* and *D* of *Figure 1* below, UNC shot a much higher volume of three point shots in 2017 than 2013. Moreover, more three pointers are shot further behind the arc in 2017 than in 2013. Therefore, it may be expected that teams may favor taking more three point shots.

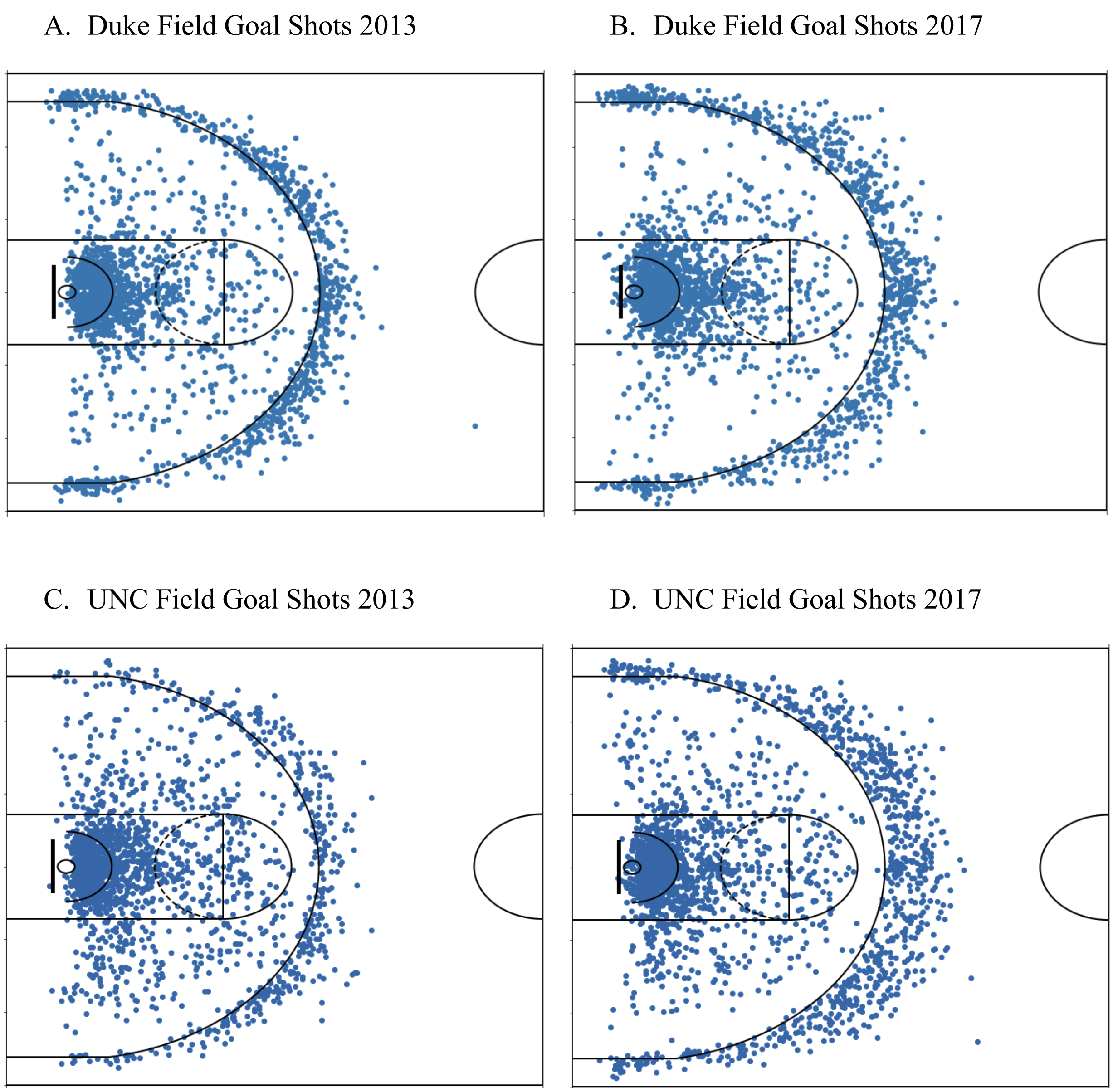

Figure 1. Multi-Panel Figure of Shot Charts

Figure 2 below supports the observation from the shot charts of *Figure 1*, in which the three point attempts (3PA) in the NCAA continuously increased from 2013 and 2017. Specifically, the median value increases over time, as well as the upper quartile. Similarly, *Figure 3* shows that the three point made shots (3PM) increase from 2013 to 2017. The lower quartile, median, and upper quartile, increase respectively throughout the seasons. In other words, NCAA Division I basketball teams are increasingly attempting and making more three point shots. Further analysis is required to determine if three point shots statistically contribute significantly to winning games.

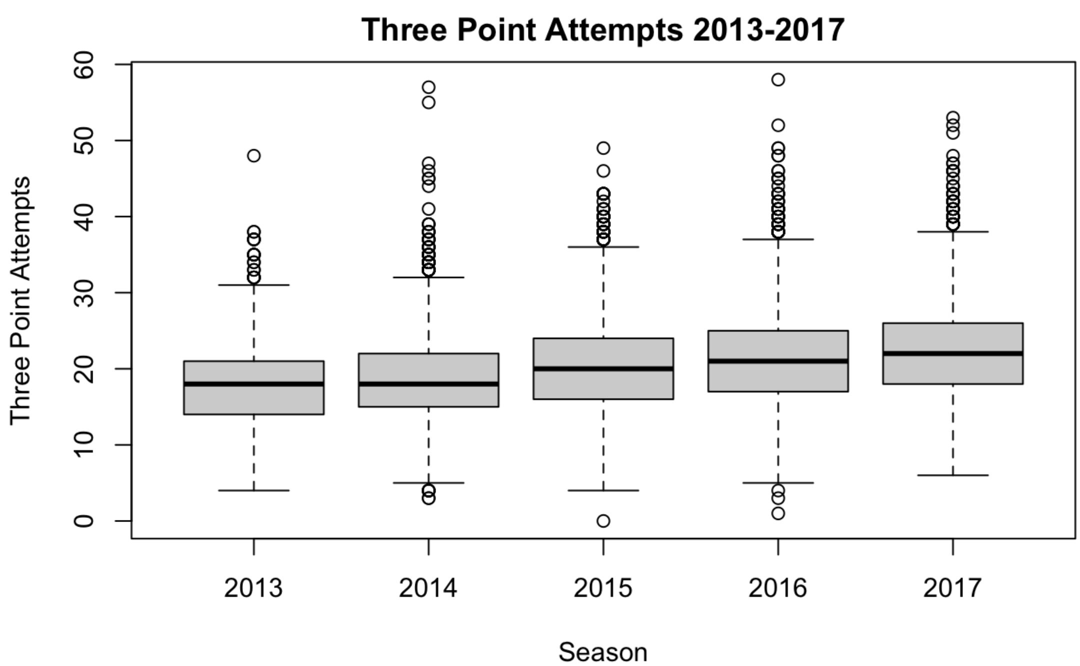

Figure 2 Three Point Attempts 2013-2017

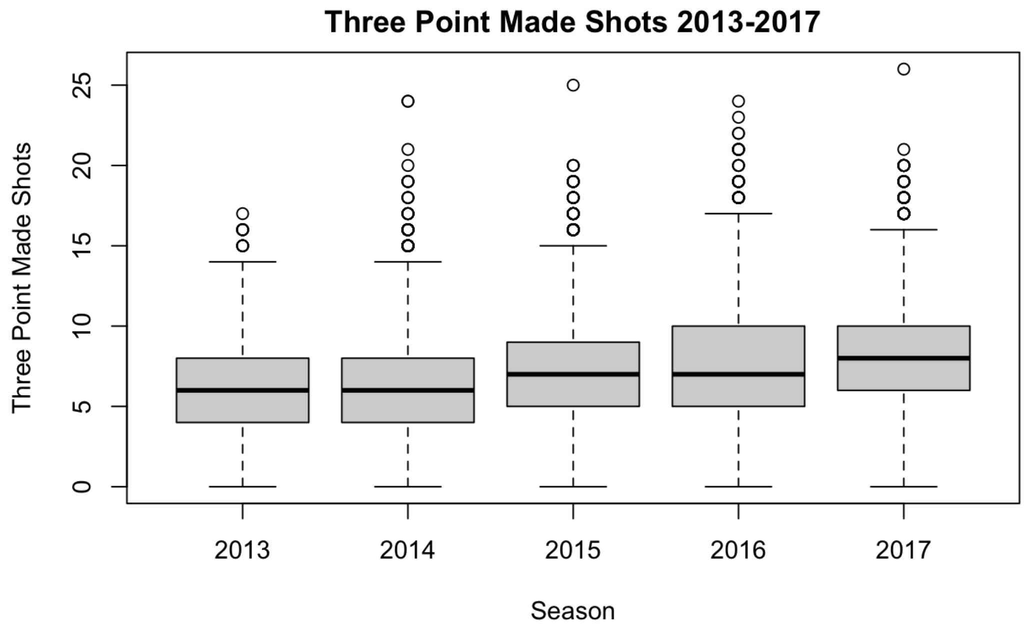

Figure 3 Three Point Made Shots 2013-2017

B. Three Point Shot Distance Trend Across Seasons

The shot charts display greater frequency of long-distance three point shots, so a time series plot was created to determine if the average distance of three point shots increased through the years. *Figure 4* below presents a general upward trend from 2013 to 2018. The lowest three point shot distance is around 277.7 inches at March 2014 and the highest average three point shot distance is around 288.7 inches at March 2018. The difference between the shot distance is roughly 11 inches, which is a 3.96% increase.

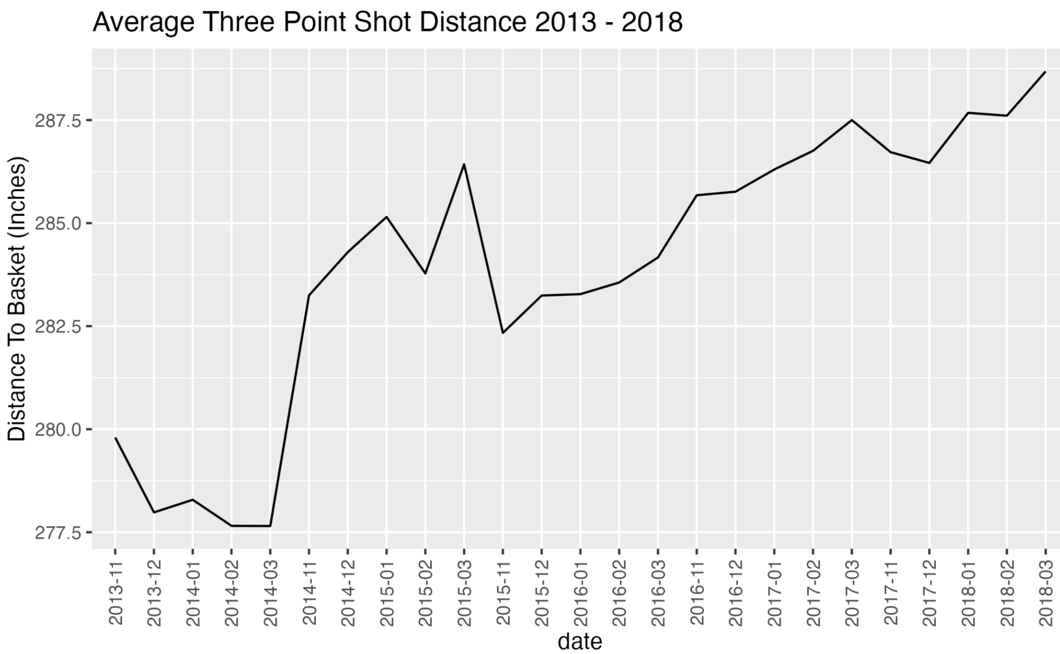

Figure 4. Average Three Point Shot Distance 2013 to 2018

Seasonal data of NCAA average shot distance from November Through March of each year

Seasonal Mann Kendall Test

The tau value is 0.8 with a *p*-value of 1.1771e-05. Therefore, there is a monotonic trend in the average distance of three point shots. Thus, the increase in average three point shot distance is statistically significant.

II. *Mixed Effects Logistic Regression Assumptions*

A. *Outliers*

In *Figure 5* below, there are points off turnovers greater than 60. Considering the maximum turnover value is 32, it means almost all turnovers have been converted to two point shots, which is unlikely. Therefore, observations with points off turnovers greater than 60 were filtered.

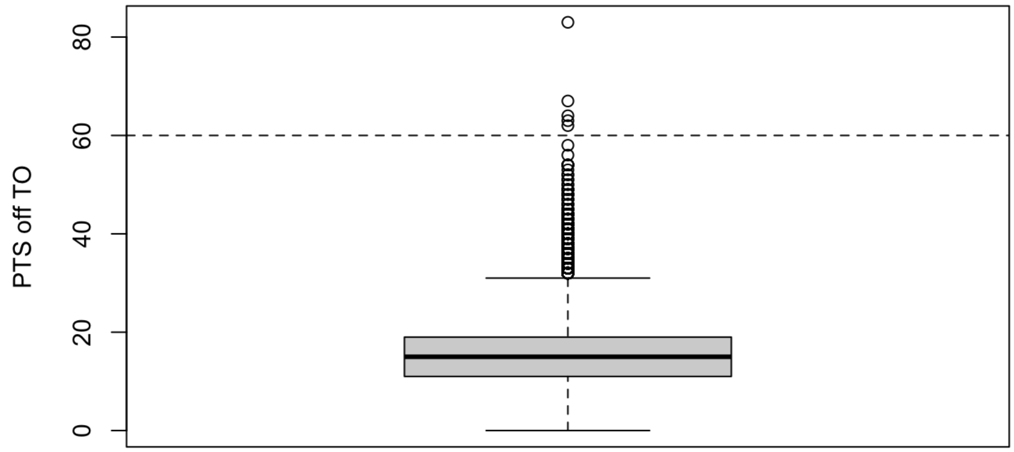

Figure 5. Boxplot of Points Off Turnover

Observations above dashed-lined are input errors

Observing *Figure 6* below, there are observations where second chance points are greater than 150. Considering even the total game points do not exceed 150 points, these are likely data input errors and were thus filtered from the dataset.

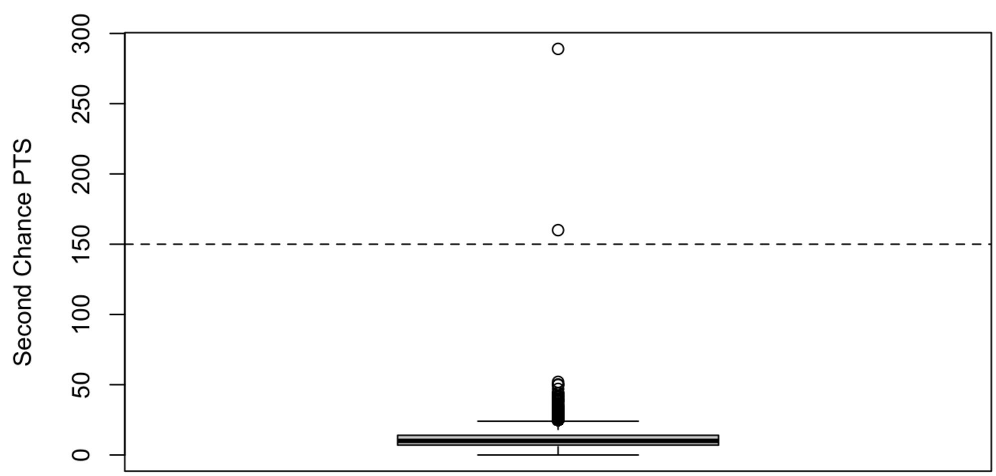

Figure 6. Boxplot of Second Chance Points

Observations above dashed lines are input errors

B. *Multicollinearity*

Table 1 below presents the VIF values of the 16 variables in the data.

Variable	VIF	Variable	VIF
Three Point Attempt	4.68	Assists	1.66
Three Point Made	3.76	Turnovers	1.81
Two Point Attempt	5.92	Steals	1.64
Two Point Made	3.42	Blocks	1.06
Free Throw Attempt	7.25	Personal Fouls	1.35
Free Throw Made	6.79	Fast Break Points	1.21
Offensive Rebounds	3.03	Second Chance Points	1.96
Defensive Rebounds	2.09	Points Off Turnovers	1.69

Table 1. Variance Inflation Factor (VIF) Table

Values greater than 5 display issues of multicollinearity

Table 1 above presents that two point attempts, free throw attempts, and free throws made have issues of multicollinearity. In order to resolve these issues, two points made were excluded from the model, which was highly correlated with two point attempts (0.72). Similarly, the free throw attempt variable was excluded.

C. *Linearity*

Observing *Figure 7* below, all numeric variables roughly display linearity as seen through the generally linear regression line at each plot.

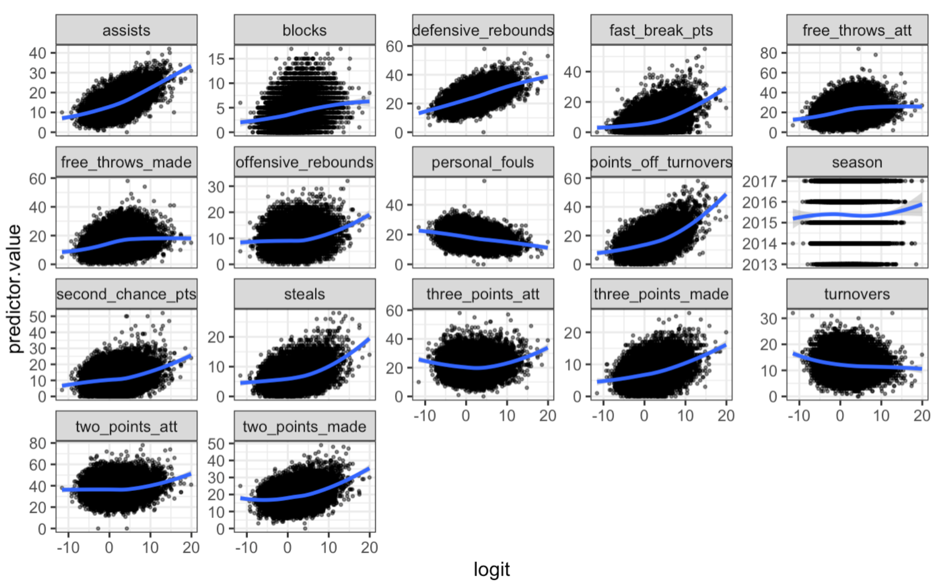

Figure 7.

III. Logistic Random Effects Mixed Model Results

Logistic Random Effects Mixed Model was used to determine the impact of the three point shot on the binary dependent variable, *win*. Results are shown through *table 1* below.

Predictors	Estimate	Standard Error	Z value	Pr(>|z|)
(Intercept)	-0.0279	0.2252	-0.124	0.901
Three Point Attempt	-0.3563	0.0072	-49.670	<2e-16***
Three Point Made	0.3978	0.0124	32.003	<2e-16***
Two Point Attempt	-0.2035	0.0051	-39.999	<2e-16***
Free Throws Made	0.0876	0.0041	21.124	<2e-16***
Offensive Rebounds	0.1689	0.0088	19.090	<2e-16***
Defensive Rebounds	0.3684	0.0065	56.561	<2e-16***
Assists	0.1914	0.0066	28.786	<2e-16***
Turnovers	-0.2863	0.0072	-39.817	<2e-16***
Steals	0.1997	0.0097	20.616	<2e-16***
Blocks	0.0781	0.0095	8.222	<2e-16***
Personal Fouls	-0.1531	0.0058	-26.298	<2e-16***
Fast Break Points	0.2482	0.0052	4.683	2.83e-06***
Second Chance Points	0.0812	0.0059	13.761	<2e-16***
Points Off Turnovers	0.1685	0.0050	33.985	<2e-16***

Table 1. Logistic Mixed Effects Model Summary Result

*Significance Level : '***' 0.001 '**' 0.01 '*' 0.05*

Marginal R^2 : 0.784

The results were interpreted using both the Marginal Effects at the Means (MEM) and the Average Marginal Effects below.

Marginal Effects at the Means

MEM was calculated using the following equation:

$$\widehat{\beta}(P)(1 - P) \text{ where}$$

$$P = \frac{1}{1+e^{-\widehat{Y}}}, \widehat{\beta} = coefficient, \widehat{Y} = \text{value after mean substitution}$$

The value of *P* above after substituting mean values of independent variables is 0.858. Thus, the probability that a team with average box-score statistics wins is 85.8%. Moreover, the MEM of *three point attempts* using this probability equals to -0.0434. Thus, an increase in one three point shot attempt at the average box-score stats *decreases* the probability that the team wins by 4.34%, holding all other variables constant at the average. In addition, the MEM of *three-point-made* is 0.0484. Thus, an increase in one three point shot made at the average box-score stats *increases* the probability that the team wins by 4.84%, holding all other variables constant at the average.

Average Marginal Effects

In order to calculate the overall marginal effect of the variables, Average Marginal Effects (AME) was calculated. Computationally, the AME is calculated by taking the average of all the values Marginal Effect values at different points. The results are presented in *Table 2* below.

Predictors	AME	Standard Error	Z value	Pr(>|z|)
Three Point Attempt	-0.0322	0.0005	-64.508	0.0000***
Three Point Made	0.0360	0.0010	35.095	0.0000***
Two Point Attempt	-0.0184	0.0004	-45.589	0.0000***
Free Throws Made	0.0079	0.0004	21.9730	0.0000***
Offensive Rebounds	0.0153	0.0008	19.6897	0.0000***
Defensive Rebounds	0.0333	0.0004	81.9382	0.0000***
Assists	0.0173	0.0006	30.9472	0.0000***
Turnovers	-0.0259	0.0006	-46.296	0.0000***
Steals	0.0181	0.0008	21.3688	0.0000***
Blocks	0.0071	0.0009	8.2683	0.0000***
Personal Fouls	-0.0138	0.0005	-27.9667	0.0000***
Fast Break Points	0.0022	0.0005	4.6904	0.0000***
Second Chance Points	0.0073	0.0005	13.9770	0.0000***
Points Off Turnovers	0.0152	0.0004	37.7312	0.0000***

Table 2. Logistic Mixed Effects Model Summary Result

*Significance Level : '***' 0.001 '**' 0.01 '*' 0.05*

The AME value of *three point attempts* is -0.0322. In other words, on average, an increase in one three point shot attempt decreases the probability of winning by 3.22%, holding other variables constant. Moreover, the AME value of *three points made* is 0.036. Thus, on average, an increase in one made three point shot increases the probability of winning by 3.6%, holding other variables constant.

Model Evaluation

I. *Confusion Matrix*

According to *table 3* below, the accuracy of the model is 86.78%. Moreover, the model specificity is 0.7736. In other words, among the games that were lost, the model predicted the loss 77.36% of the time. The model sensitivity value is 0.9126. Thus, among the games that were won, the model predicted the win 91.26% of the time.

	Actual Win	Actual Loss
Predicted Win	4603	542
Predicted Loss	441	1852

Table 3. Confusion Matrix of Win Prediction

II. *Receiver Operating Characteristic (ROC) Curve*

As seen in *Figure 8* below, the AUC value of the model is 0.9386, which indicates that the model performs much better than random chance. Thus, the model is capable of classifying wins and losses very accurately.

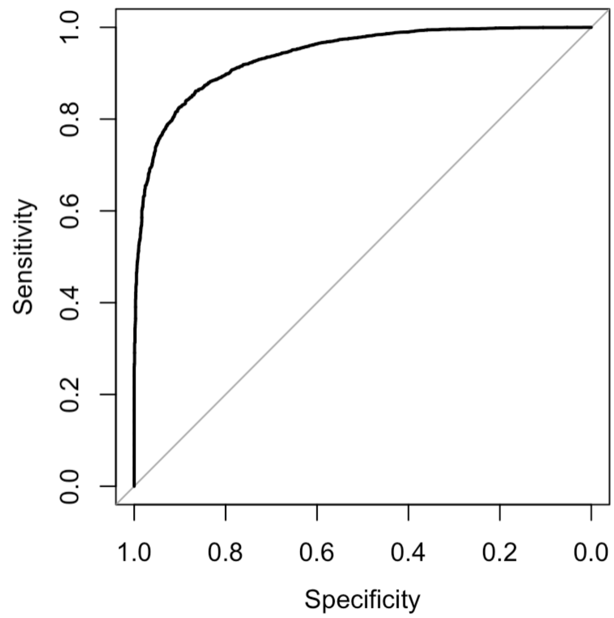

Figure 8. ROC Curve of Logistic Mixed Effects Model

AUC Value: 0.9386

III. Wald Test

The Wald Test reveals that the variables of interest, three point attempts and three point made shots have a *p*-value less than 2.2e-16. Therefore, the three point shot attempts and three point made shots are jointly significant in the model. Results are shown through *Table 4* below.

	Restricted Model	Unrestricted Model
Number of Parameters	14	16
AIC	17653	14421
BIC	17767	14551
Log-likelihood	-8812.6	-7194.4
Chi-Squared	-	3236.3
Degrees of Freedom	-	2
Pr(>Chisq)	-	< 2.2e-16 (***)

Table 4. Wald Test ANOVA Results
Three Point Attempts and Three Point Made Shots removed in restricted model

Discussion

College Basketball Strategy

The results of the regression presents that on average, increasing the number of three point shot attempts decreases the probability of winning in the NCAA. This does not align with the expectation that prioritizing three point shots increases the chances of winning, as shown through the recent shifts in strategies of NBA teams. Specifically, the average marginal effect of taking an additional three point shot is -0.0322, whereas that of a two point shot is -0.0184. This is a 54.55% percentage difference. However, it is worth noting that the three point made shots have the highest AME with a value of 0.036. In other words, although *taking* many three point shots is detrimental, *making* three pointers increases the chances of winning the most. Despite the potential harm of taking more three pointers, time series analysis revealed that there exists an upward trend in three point attempts from 2013 to 2018. This suggests that college teams may be striving to capitalize on the benefits of making three point shots. Ultimately, it can be expected that efficient three point shooters may be valued as teams increasingly attempt more three point shots. However, it may be helpful for college teams that do not possess three point shooters or specialists to prioritize two point shots near the basket.

The regression shows that defensive efforts increase the probability of winning the most after three point made shots. On average, an increase in one defensive rebound increases the probability of winning by 3.33%, holding all other variables constant. Moreover, the third variable that contributes to winning the most is, steals, in which an increase in one steal increases the probability of winning by 1.81%, ceteris paribus. In addition, It is worth noting that turnovers decrease the probability of winning the most, in which one turnover decreases the probability of winning by 2.59%, ceteris paribus. Proficient defenders can force turnovers for the opposing team and decrease the opponent's chance of winning.

Comparison with the NBA

The high impact of three point made shots and defense in college basketball align with the prominence of "three-and-d" role players in the NBA, who focus on hitting three point shots and being active on defense (Wimbish, 2020). Thus, the study reveals that "three-and-d" players in the NCAA may possess qualities that translate into the NBA. Ultimately, observing player performance in the NCAA may be valuable in identifying talent for the NBA drafts, since three point shots and defense are prominent in both leagues. Moreover, based on this finding, teams can design player development programs to heighten their chances of winning.

The NBA has also observed teams attempting an increasing proportion of three point shots in their shot attempts throughout the years. The high proportion of three point shots translate to greater win percentages. Teams who attempted more than 55% of their shot attempts with three pointers in a given game won over 78% of those games (Jaguszewski, 2020). Considering the overlap in the impact of three point shots in winning games in the NBA and NCAA, it can be observed that the three point strategy is evolving the game of basketball as a whole.

Limitations

Although the NCAA analysis displays similar trends as the NBA in terms of the impact of three point made shots, the study does not show causality between the NBA and NCAA. Therefore, there are opportunities for future research on whether the prevalence of three point shots in the NBA influenced the NCAA to adopt similar strategies. Moreover, factors such as player position were unavailable in the database. Therefore there are opportunities for future research on shot locations based on positions after acquiring position data from alternative sources. Furthermore, there are studies that investigate specific situations when teams prefer to attempt three point shots during NBA games (Zhuang, 2020). For instance, a study analyzed how three point shot selections differ during scenarios such as fourth-quarters, overtime, and playoff games (Zhuang, 2020). Therefore, analyzing NCAA shot pattern behaviors based on different scenarios may be a potential research opportunity.

The data available for investigating trends in three point shot distance was from 2013 to 2018, which leaves out the most recent years of 2019 to 2021. Thus, this research could not be generalized to reflect the most recent college basketball trend.

Acknowledgements

I thank Dr. Sarah Supp (Denison University) for her suggestion to apply a mixed effects model for the project, and Dr. Laura Boyd (Denison University) for her guidance through the interpretation of statistical model results.

References

Beshai, Peter. "Buckets: Visualizing NBA Shot Data."

Freitas, Lucas. "Shot distribution in the NBA: did we see when 3-point shots became popular?. *German Journal of Exercise and Sport Research* 51.2 (2021): 237-240.

Goldsberry, K. (2019, December 17). *How deep, audacious 3-pointers are taking over the NBA*. ESPN. Retrieved December 11, 2022, from https://www.espn.com/nba/story/_/id/28312678/how-deep-audacious-3-pointers-taking-ba

Google. (n.d.). *Exporting table data*. Google Cloud. Retrieved November 12, 2022, from https://cloud.google.com/bigquery/docs/exporting-data

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Jaguszewski, Maciej. "Increasing role of three-point field goals in National Basketball Association." (2020).

Medcalf, M. (2018, March 28). *How college basketball has fallen in love with the 3-pointer*. ESPN. Retrieved November 12, 2022, from https://www.espn.com/mens-college-basketball/story/_/id/22874000/college-basketball-lve-3-pointer-thank-golden-state-warriors

NCAA Basketball. NCAA, Sport Radar. Google Big Query. 15 Sept. 2022. https://console.cloud.google.com/marketplace/details/ncaa-bb-public/ncaa-basketball

Powell, S., & Archive. (n.d.). *Stephen Curry reached 'his own level' and left behind a new-look NBA*. NBA. Retrieved November 13, 2022, from https://www.nba.com/news/stephen-curry-changed-game-long-before-3-point-milestone

Reich, B. J., Hodges, J. S., Carlin, B. P., & Reich, A. M. (2006). A spatial analysis of basketball shot chart data. *The American Statistician*, *60*(1), 3-12.

Sarlis, Vangelis, and Christos Tjortjis. "Sports analytics—Evaluation of basketball players and team performance." *Information Systems* 93 (2020): 101562.

Tjortjoglou, S. (2015, July 28). *How to create NBA shot charts in Python*. Savvas Tjortjoglou. Retrieved October 24, 2022, from http://savvastjortjoglou.com/nba-shot-sharts.html

UCLA Advanced Research Computing. (n.d.). *Mixed Effects Logistic Regression*. OARC Stats. Retrieved October 25, 2022, from https://stats.oarc.ucla.edu/r/dae/mixed-effects-logistic-regression/

UCLA. (n.d.). *How Are The Likelihood Ratio, Wald, and Lagrange Multiplier Tests Different And Or Similar?* OARC Stats. Retrieved November 13, 2022, from https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faqhow-are-the-likelihood-ratio-wa d-and-lagrange-multiplier-score-tests-different-andor-similar/

Wimbish Jun 2, J. (2020, June 2). *Five all-time NBA 3-and-D specialists who defined role that is central to today's game*. CBSSports. Retrieved November 14, 2022, from https://www.cbssports.com/nba/news/five-all-time-nba-3-and-d-specialists-who-defined-ole-that-is-central-to-todays-game/

Zhuang, F. (2020). NBA three point shot attempts and accuracy distribution. *Scientific and Social Research*, *2*(2).

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, *115*(5), 654–657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

Appendices

Appendix I.

Che Hoon Jeong (11.13.2022). *NBA-NCAA-analytics* [electronic resource: github repository]. https://github.com/chejeong/NBA-NCAA-analytics