­­­Prediction of NBA team success using team performance parameters

Zachary Katz and Marco Lagos  
STAT 410  
Professor Meng

­

**Table of Contents**

Background

Data Information

Variables

Basic Summary and Visualization

Lasso Selection Model

Summary and Discussion

References

Reflection

Appendix

3

3-4

4-5

10

12

10

14

16

18

**Background**

Basketball has been a beloved sport for decades, and the National Basketball Association (NBA) is one of the most prestigious and well-known professional basketball leagues in the world. The NBA has attracted fans from all over the globe, with millions tuning in to watch their favorite teams compete. While the primary objective of each team is to win games, the sport has become a lucrative business, with players' salaries often reaching astronomical heights. The question of whether paying players more in the NBA results in more team success has been a topic of debate for years. On one hand, paying players more may attract top talent and motivate players to perform at their best. On the other hand, teams may struggle to balance their budgets and invest in other areas, such as coaching, scouting, and team chemistry, that can contribute to team success.

To explore this question, many researchers have turned to statistical analysis of NBA data. They have looked at the relationship between a team's total salary and its number of wins. Some studies have found a positive correlation between salary and wins, suggesting that paying players more does lead to more team success. Others have found little to no correlation, suggesting that other factors, such as coaching and team chemistry, may be more important.

However, this question is not the only one that can be addressed through statistical analysis of NBA data. Researchers have also looked at the impact of various statistics on team success. For example, they have studied the rise of three-point shooting and the decline of two-point shooting, as well as the impact of defensive statistics such as steals and blocks. By identifying the statistics that are most strongly associated with team success, researchers hope to provide insights that can inform team strategies and player development.

In recent years, the availability of advanced NBA statistics has made it possible to conduct more sophisticated analyses of NBA data. These statistics provide a more detailed picture of player and team performance, allowing researchers to study nuanced relationships between variables. However, it's important to keep in mind that statistical analysis alone cannot fully capture the complex dynamics of a basketball game. Other factors, such as player injuries and team chemistry, can also have a significant impact on team success.

Statistical analysis of NBA data can provide valuable insights into the factors that contribute to team success. Whether studying the relationship between salary and wins or the impact of specific statistics, researchers can use statistical tools to uncover patterns and relationships that may be difficult to discern through observation alone. While statistical analysis is not a substitute for on-court performance, it can help teams and players make informed decisions and strategies to improve their chances of success.

**Data Information**

The data for this study was collected from the National Basketball Association (NBA) across the 2000-2001 season to the 2019-2020 season. The data was downloaded as CSV files from a publicly available source and imported into R for analysis.

The data includes a variety of NBA statistics, broken down into different categories. Shooting statistics include information on the accuracy and frequency of different types of shots, such as three-pointers, two-pointers, and free throws. Total team statistics include more general information on a team's performance, such as the number of points scored per game, the number of rebounds, assists, turnovers, and fouls committed. Total statistics opponent include data on how well opponents performed against a team, such as their shooting percentage, points scored, and turnovers forced. Advanced statistics include more complex metrics that go beyond basic box score statistics, such as player efficiency rating (PER), true shooting percentage (TS%), and win shares (WS).

The data set consists of 20 years of NBA data, with a total of 2,460 observations and 69 variables. The data was cleaned and processed in R to ensure that it was free of missing values and errors. The data provides a rich and diverse set of NBA statistics that can be used to explore the factors that contribute to team success and identify areas for improvement.

**Variables**

The data set used in this study includes a total of 69 variables that can be divided into three main categories: shooting statistics, total team statistics, and advanced statistics.

Shooting statistics include the following variables:

* MP: Minutes played
* FG%: Field goal percentage
* Dist.: Average distance of shot attempt
* 2P: Total number of two-point field goals made
* 0-3: Number of two-point field goals made from 0-3 feet from the basket
* 3-10: Number of two-point field goals made from 3-10 feet from the basket
* 10-16: Number of two-point field goals made from 10-16 feet from the basket
* 16-3P: Number of two-point field goals made from 16 feet to the three-point line
* 3P: Total number of three-point field goals made
* 2P, 0-3: Percentage of two-point field goal attempts made from 0-3 feet from the basket
* 3P, %FGA: Percentage of field goal attempts from three-point range
* 3P%: Three-point shooting percentage
* Att.: Total number of free throw attempts
* Md.: Total number of free throws made

Total team statistics include the following variables:

* FG: Total number of field goals made
* FGA: Total number of field goal attempts
* FG%: Field goal percentage
* 3P: Total number of three-point field goals made
* 3PA: Total number of three-point field goal attempts
* 3P%: Three-point shooting percentage
* 2P: Total number of two-point field goals made
* 2PA: Total number of two-point field goal attempts
* 2P%: Two-point field goal percentage
* FT: Total number of free throws made
* FTA: Total number of free throw attempts
* FT%: Free throw shooting percentage
* ORB: Total number of offensive rebounds
* DRB: Total number of defensive rebounds
* TRB: Total number of rebounds
* AST: Total number of assists
* STL: Total number of steals
* BLK: Total number of blocks
* TOV: Total number of turnovers
* PF: Total number of personal fouls
* PT: Total number of points scored

Advanced statistics include the following variables:

* Age: Average age of players on the team
* W: Total number of wins
* L: Total number of losses
* PW: Pythagorean wins, an estimate of a team's expected wins based on their points scored and allowed
* PL: Pythagorean losses, an estimate of a team's expected losses based on their points scored and allowed
* MOV: Margin of victory, the average point differential per game
* SOS: Strength of schedule
* SRS: Simple Rating System, an estimate of a team's strength relative to the average team
* ORtg: Offensive rating, an estimate of points scored per 100 possessions
* DRtg: Defensive rating, an estimate of points allowed per 100 possessions
* NRtg: Net rating, the difference between offensive and defensive rating
* Pace: Number of possessions per game
* FTr: Free throw rate, the ratio of free throw attempts to field goal attempts
* 3PAr: Three-point attempt rate, the ratio of three-point attempts to field goal attempts
* TS%: True shooting percentage, an estimate of shooting efficiency that takes into account two-pointers, three-pointers, and free throws

**Basic Summary and Visualization of Variables**

1. Copy in this section
2. Present our lasso model and explain how we got there
3. Talk about the model and different graphs
4. Put our references which is data site and text book
5. Write reflection
6. Do lab
7. Put code at the end