The Impact of NBA Player's Performance on Their Salary

Project Update 1

<u>Group 1:</u>

Jishen Wang

David Wiley

Scott McCauley

Texas Tech University

ISQS-6350-D01 Multivariate Analysis

Alireza Sheikh-Zadeh, Ph.D.

November 7, 2020

Introduction

Ever since professional baseball has gravitated towards data analytics for better decision making, the National Basketball Association (NBA) soon followed. The use of cameras now record every movement of both the ball and all 10 players 25 times per-second (Merrimack College, 2020). NBA analytics has made quite a contribution to the sport, and we will provide a few examples. NBA analytics plays a key role in selecting new players during the NBA draft, where teams sort through data on player description and performance metrics in hope of finding great talent. If a professional team makes the wrong player selection, it could set progress back by several years.

Scoring is another area affected by NBA analytics, where each individual player's free throw percentage and field goal locations are all closely analyzed in order to improve shooting form. Finally, NBA analytics plays a large role in assessing player matchups. By analyzing the strengths and weaknesses of their players and opponents, teams can strategize to place their players in favorable matchups. Our group will delve into this new and exciting field of analytics by performing various multivariate procedures on a dataset of NBA player performance metrics.

Dataset

The dataset was downloaded from the Kaggle[1] website. The data consists of various player statistics extracted from the 2018-19 NBA season. The dataset includes twenty-seven variables, which we have broken down into three main categories: player description (name, height, weight), player value (according to their salary), and player performance (points, rebounds, assists). For a complete list of variables and descriptions see Appendix A.

As an initial exploratory analysis on our dataset and to evaluate its potential for multivariate analysis, our team constructed a correlation matrix on seven of the most relevant variables in the dataset. These variables are perceived to be most relevant according to our team's analysis plan. The correlation matrix is shown below in Figure 1.

	Salary	Points	Blocks	Steals	Assists	Rebounds	Turnovers
Salary	1.00	0.62	0.30	0.48	0.49	0.49	0.55
Points	0.62	1.00	0.39	0.62	0.64	0.64	0.82
Blocks	0.30	0.39	1.00	0.32	0.09	0.69	0.32
Steals	0.48	0.62	0.32	1.00	0.65	0.43	0.63
Assists	0.49	0.64	0.09	0.65	1.00	0.29	0.81
Rebounds	0.49	0.64	0.69	0.43	0.29	1.00	0.56
Turnovers	0.55	0.82	0.32	0.63	0.81	0.56	1.00

Figure 1: Correlation Matrix from the NBA Performance Dataset

The matrix above has been rounded to two decimal places and highlights correlations over 0.5. We observe that blocks and rebounds are positively correlated (0.69) which supports our intuition since they both favor individuals of taller stature such as the Center and Power Forward positions. One surprising pattern observed is that turnovers and points have a high positive correlation (.82) indicating that players who score more points also have more turnovers (lose the ball more often). However, throughout this project we will conduct various analysis techniques to confirm these hypotheses.

The next graphic that our team chose to plot is a histogram of all NBA player salaries for the season.

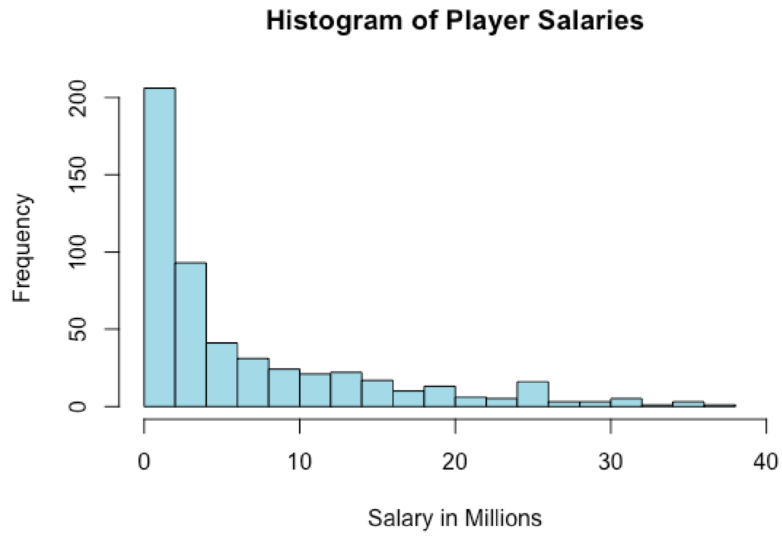

Figure 2: Histogram of NBA Salaries

After plotting this histogram we can observe that the salaries are distinctly right skewed. While the majority of NBA players make less than 5 million dollars, we observe that the upper bound of salary is much higher. Some players are even valued at almost 40 million dollars for this season. One of our interests for this analysis will be to identify what performance factors influence salary and how players can distinctify themselves in order to obtain higher salaries. Our hypothesis is that the path to higher salaries is segmented and that there is no singular performance metric that determines salary, but rather multiple clusters of performance combinations which can each indicate higher salaries. In order to assess this, we will graphically observe relationships between variables using the scatter plots shown below.

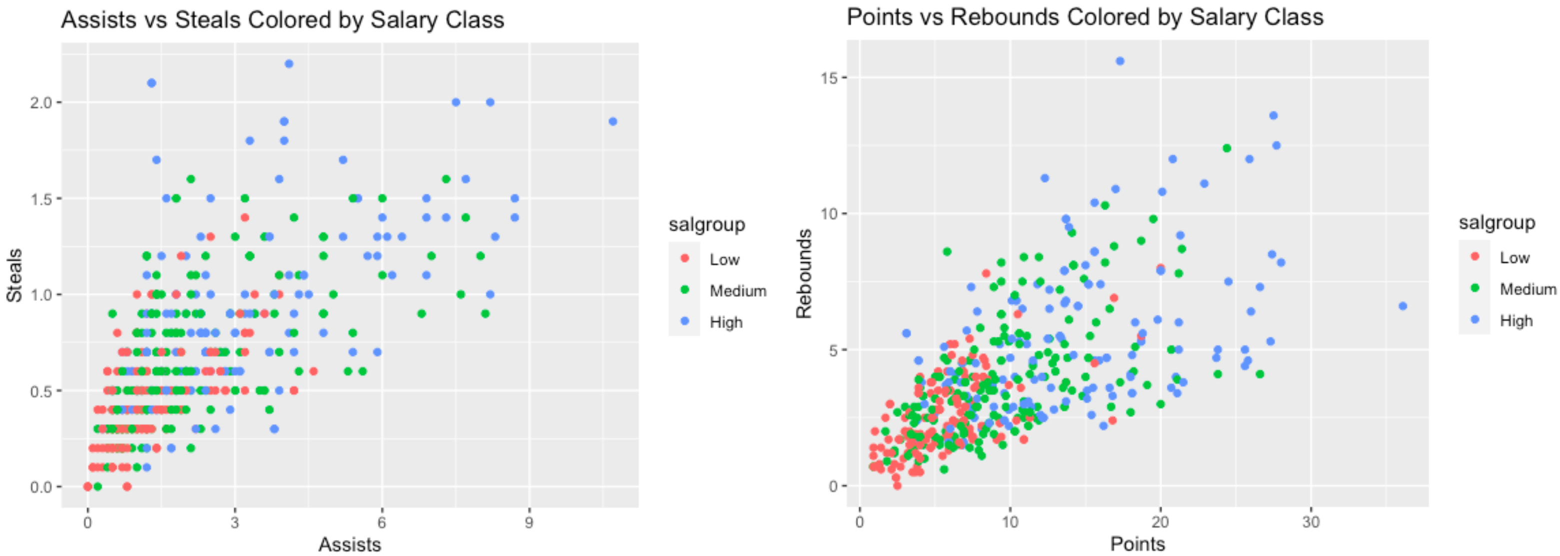

Figure 3: Color Coded Scatter Plots by Salary

The above plots display scatter plots of two sets of variables, color coded by salary group. The salaries have been grouped as ‘Low’, ‘Medium’ or ‘High’ using the below specifications:

- ‘Low’ salary group (shown in red): obtained salaries less than $2,000,000 this current season
- ‘Medium’ salary group (shown in green): obtained salaries from $2,000,000 - $8,000,000 this season
- ‘High’ salary group (shown in blue): obtained salaries greater than $8,000,000 for the current season

From the above plots we can observe that there is a general progression in salaries, from red to green to blue in both scatter plots. This indicates that individuals can obtain higher salaries if they obtain more steals and assists. However, players can also obtain higher salaries if they obtain more rebounds and steals. This supports our hypothesis that there is more than one method (combination of performance metrics) to obtain higher salaries.

Multivariate Analysis

We will perform three multivariate procedures (dimension reduction, clustering, and Confirmatory Factor Analysis (CFA)) on the dataset to better understand which statistical measurements included in the data are highly correlated. Specifically, dimension reduction techniques such as Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) will allow us to reduce the dimension of the data and create new variables that explain a majority of the variability within the dataset.

Next, we will categorize the players by salary into three classes: high, medium, and low. Our team will then conduct the same multivariate techniques on each of these salary groups and create new variables which are combinations of the original variables in order to better understand how player salary is impacted by player performance.

Motivation

Our team's motivation towards pursuing this project is centered around a strong interest in sports and a curiosity towards how analytics has influenced basketball and how it could continue to transform the sport. We believe that our project can potentially create value for NBA players, coaches, staff, and owners in many ways, which are specified below.

Performing multivariate analysis on these statistical measurements provides insight into the overall value of NBA players for the coaches, staff, and owners. In the NBA there is a time period called Free Agency where teams can sign any eligible players that are not under contract during that time. Analyzing a player's performance using principal component analysis can be utilized to identify the play styles (principle components) that account for the most variation within each salary category. Teams can use the results to assist in assigning value estimates to players and then deciding how much they are willing to invest in each player.

A similar measure of player value can be extended to selecting players from the NBA Draft. This is an event where the teams each take turns selecting new college players to join their teams. Similar to Free Agency, NBA teams can use principal component analysis to identify common play styles in each salary group. Teams can then match college player performance metrics to these playstyle metrics to assign value estimations. Using this, teams can select players of optimal value in the draft.

The final application to NBA analytics that our project can be used for are growth projections for younger NBA players. Using clustering techniques, NBA teams will be able to group players together and identify players with similar play styles and variable measurements. Teams can then identify younger players of interest and correspond them to older more experienced players in their cluster. Finally, teams can identify the growth and performance of the older players to estimate the potential growth and improvements for the younger players.

References

[1] Schmadamco (2019). *NBA Regular Season Stats 2018-2019*. Retrieved from https://www.kaggle.com/schmadam97/nba-regular-season-stats-20182019/.

[2] Merrimack College (2020). *How NBA Analytics is Changing Basketball.* Retrieved from https://onlinedsa.merrimack.edu/nba-analytics-changing-basketball/

Appendix A: List of Variables

Variable	Description	Type
Player Description		
Name	Name of the player	String
Height	Height of the player in inches	Inches
Weight	Weight of the player in pounds	Pounds
Team	Team the player played on	String
Age	Age of the player	Integer
Player Value		
Salary	Yearly salary of the player	Integer
Player Performance		
Points	Average number of points per game for the season	Decimal
Blocks	Average number of blocks per game for the season	Decimal
Steals	Average number of steals per game for the season	Decimal
Assists	Average number of assists per game for the season	Decimal
Rebounds	Average number of rebounds per game for the season	Decimal
FT%	% of freethrows made	Decimal
FTA	Average number of freethrows attempted	Decimal
FG3%	% of three point field goals made	Decimal
FG3A	Average number of three point field goals attempted	Decimal
FG%	% of field goals (excluding freethrows) made	Decimal
FGA	Average number of field goals (excluding freethrows) attempted	Decimal
MP	Average minutes played per game	Decimal
G	Games played this season	Integer
PER	Player Efficiency Rating = (Points + Rebounds + Assists + Steals + Blocks – Missed FG – Missed FT - Turnovers) / G	Decimal
OWS	Offensive Win Shares	Decimal
DWS	Defensive Win Shares	Decimal
WS	Win Shares	Decimal
WS48	Win Shares Per 48 Minutes	Decimal
USG	Usage Percentage	Decimal
BPM	Box Plus-Minus	Decimal
VORP	Value Over Replacement Player	Decimal