Math 261A Project 1 Draft Report: NBA Player Wingspan vs Total Rebounds

Jan Blackburn

Title: Exploration of NBA Player Wingspan on Rebound Performance
Author: Jan Blackburn, Math 261A
Date: September 24, 2025

Abstract

Analysis of professional basketball players physical attributes and their potential impact performance is critical to helping professional basketball teams improve their odds of winning in order to draw in fans and increase team revenues. This paper seeks to examine one key physical attribute, wingspan as a potential predictor on total rebounds which is an important measure of player's defensive contribution. This paper evaluates the relationship between these two variables by applying a simple linear regression using player level data for the 2024-25 National Basketball Association (NBA) statistics. In the course of conducting this analysis, it is found that there is a relationship between a player's wingspan and the total number of rebounds. This is followed with a discussion on how a more robust model can be developed by building a multiple regression model as path for further exploration.

Introduction

Professional basket ball in the United States is a big business as the market value of NBA teams increasing substantially in recent years (1). This trend has increased the importance of finding and recruiting players who can improve team performance and helped spawn a robust literature in evaluating player performance that is viewed as a key area of investment2. One area of analysis and research is evaluating the physical attributes as a potential predictor of performance on a number of offensive and defensive performance metrics.

When thinking of physical attributes for basketball players, height is what comes to mind. In addition to height, wingspan is frequently used as a key physical attribute that can provide

players with advantages in defensive performance. Common metrics used to evaluate defensive performance include rebounds, steals and blocks which are defined in more detail in the Data section. After evaluation of these metrics in the data set, Total Blocks was selected since the amount of data available is more robust and amenable to regression modeling.

A player's wingspan is often correlated to a player's height, prompting use of an alternate measure,Wingspan Advantage, which measures the difference between a player's wingspan and height. Wingspan Advantage most often has a positive value (wingspan > height) with a higher value suggesting greater potential for defensive capability. After evaluation between Wingspan and Wingspan Advantage, for which measure would be better as a the predictor variable, Wingspan was selected due to potential confounding issues with Wingspan Advantage.

The resulting regression model showed that there is a linear relationship between Wingspan and Total Rebounds but there is the opportunity to explore a more robust regression model by incorporating additional variables into a multiple regression model.

The rest of this paper provides an overview of the data sources, methods and results from the analysis as well as a discussion of future model changes that can be explored to improve model fit and predictive power.

Data

The data for this study was sourced from NBA Data. This data set contains player level statistics collected from the 2024-25 NBA season. Many of the metrics are normalized to averages over 100 ball possessions in order to account for differences in the number of games and minutes played. This data set was compiled from a number of online data sources by the original author of the data set.

The data set contains 419 rows and 26 columns that includes information and statistics for each player including player physical attributes and a number of offensive and defensive performance metrics. A summary of the variables and definitions is included in the Appendix.

Since Wingspan and Wingspan Advantage are key physical characteristics thought to contribute to enhanced defensive performance, each of these variables was evaluated as the predictor variable for the regression model. The definition of each measure is specified below:

- **Wingspan (inches):** Measure of the length of a player's arms from the fingertips when the arms are held out parallel to the ground.
- **Wingspan Advantage (inches):** Difference between a player's Wingspan and Height. This statistic is often positive (i.e., Wingspan > Height) with a larger value being preferred.

In the data set, the following measures of defensive performance were available. Each is measured as the average over 100 ball possessions:

- **Rebounds:** Player takes possession of the ball after a field goal, changing ball possession between the teams.
- **Steals:** Player takes the ball from an opposing player, eliminating the opposing team's opportunity to score and creating new opportunity for the player's team to score
- **Blocks:** Player blocks a field goal attempt from the opposing team, preventing a potential scoring opportunity and initiating a potential for change in ball possession

Rebounds are considered more important than Steals or Blocks since they directly create new possessions and additional scoring opportunities and occur more often as shown in in Exhibit 1. Based on these factors, Total Rebounds was selected as the response variable for the regression model.

Exhibit 1: Key Measures and Statistics

Metric	Min	Q1	Median	Mean	Q3	Max
Wingspan	73.0	80.0	82.3	82.4	85.0	96.0
Wingspan Advantage	-1.5	3.0	4.5	4.5	6.1	10.8
Rebounds	3.0	6.0	8.0	9.1	11.3	23.0
Blocks	0.0	0.4	0.8	1.0	1.4	5.6
Steals	0.0	1.2	1.6	1.7	2.0	5.1

The data set received was very complete with just one row needing to be removed due to NA values so the data set ended up at 418 rows. An inspection of the data did not identify any extreme data suggesting that further data cleansing or transformation was warranted.

Methods

The first step of the analysis was an evaluation of both Wingspan and Wingspan Advantage as the predictor variable. The concern with Wingspan Advantage as the predictor variable is that it reflects the difference between two potentially correlated variables: player height and wingspan. This complication in the metric could result in some confounding effects when modeling the relationship with Total Rebounds. This possibility was investigated by analyzaing a scatter plot of wingspan advantage and total reboundsd as seen in Exhibit 2 below. As can be seen, there is fair degree of dispersion and there appears to be little basis for thinking a linear relationship exists between these two variables.

Exhibit 2: Scatter Plot of Wingspan Advantage vs. Total Rebounds

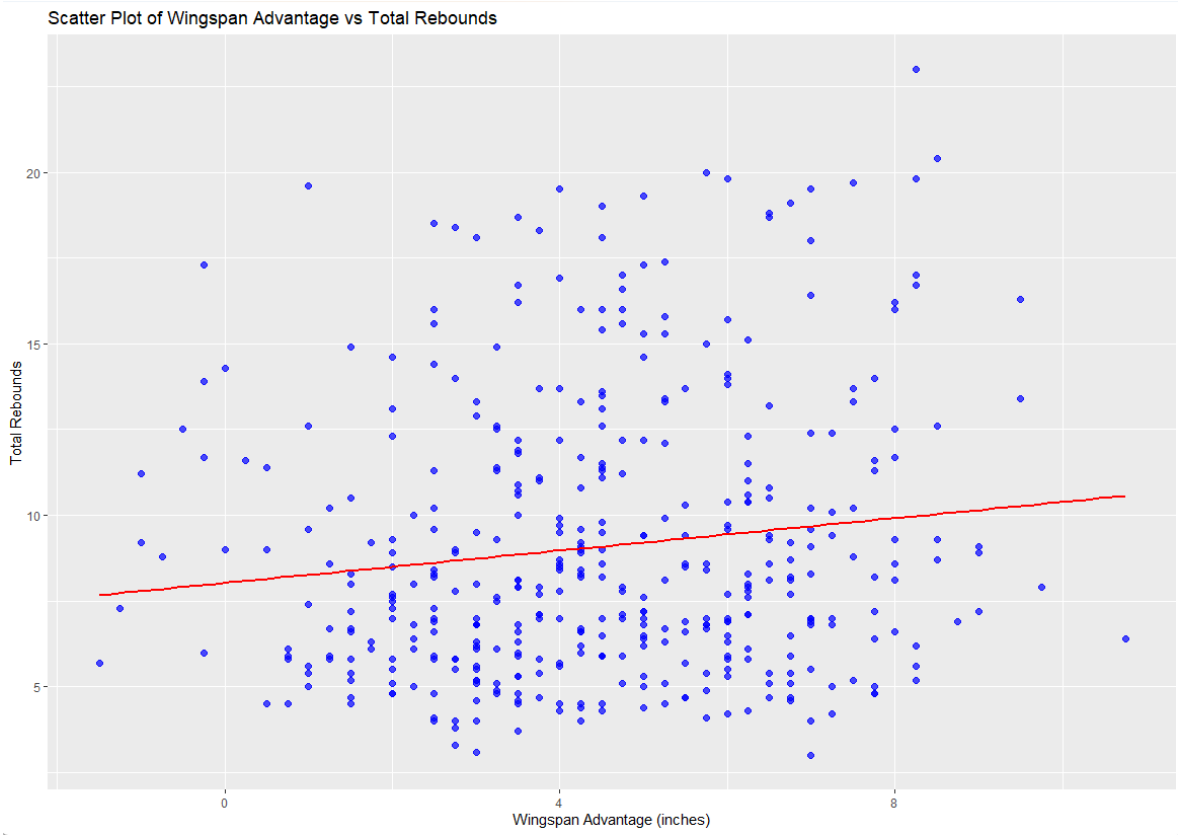
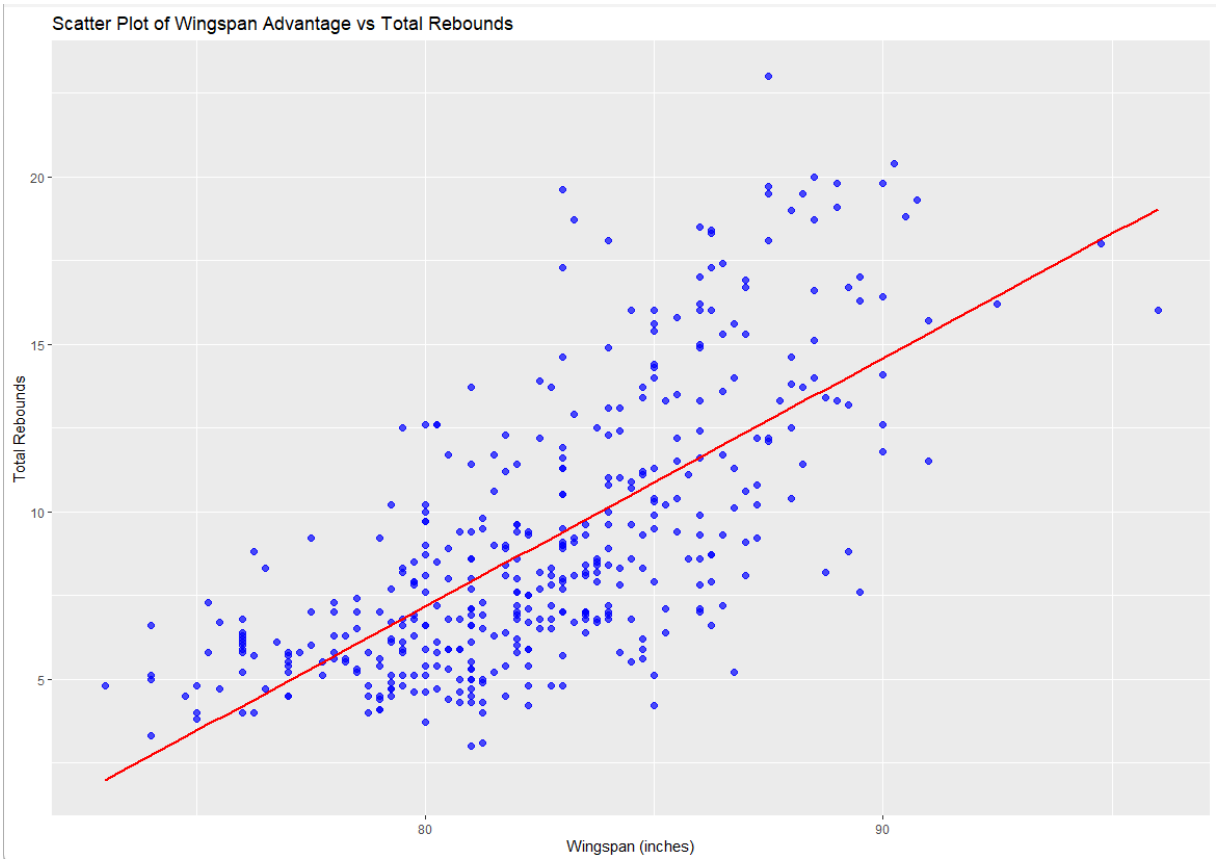


Exhibit 3 displays the same analysis using Wingspan versus Total Rebounds. This scatter plot illustrates that there is stronger case to be made for their to be a linear relationship with Total Rebounds and supports the decision to select Wingspan as the predictor variable.

Exhibit 3: Scatter Plot of Wingspan vs. Total Rebounds



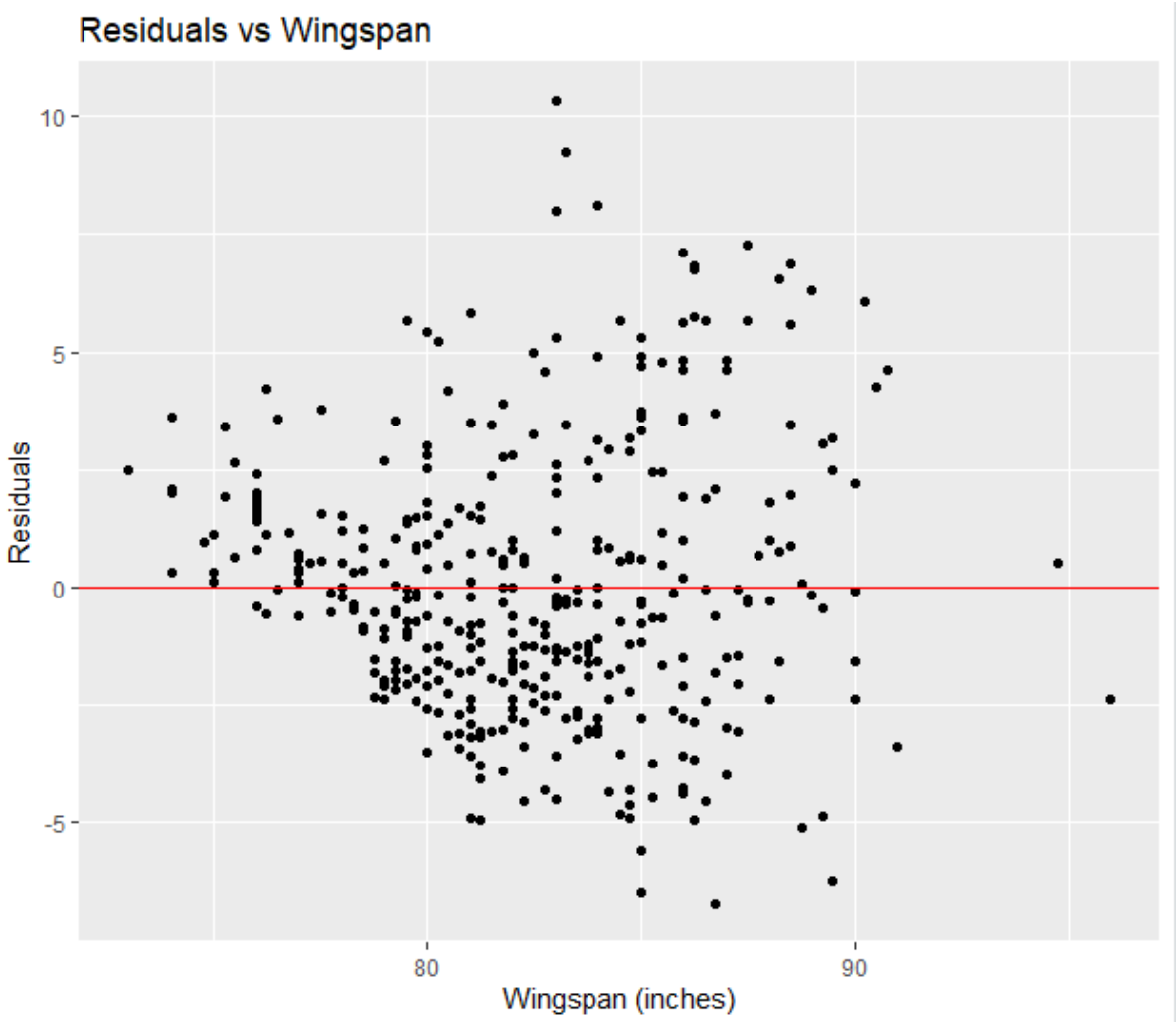
With the response and and predictor variables selected, the relationship between Wingspan and Total Rebounds was evaluated using a simple linear regression based on the least squares methodology using the R statistical analysis package. The specific attributes of the model are specified below:

$$R_i = B_0 + B_1 W_i + \epsilon_i$$

- R_i**:** Total rebounds achieved by player i in the last 100 ball possessions
- B_0: Intercept for the regression model
- B_1: Slope of the regression model that measures the average number of additional rebounds per additional inch of player wingspan
- W_i: Wingspan of player i measured in inches
- e_i: Error term of the regression model

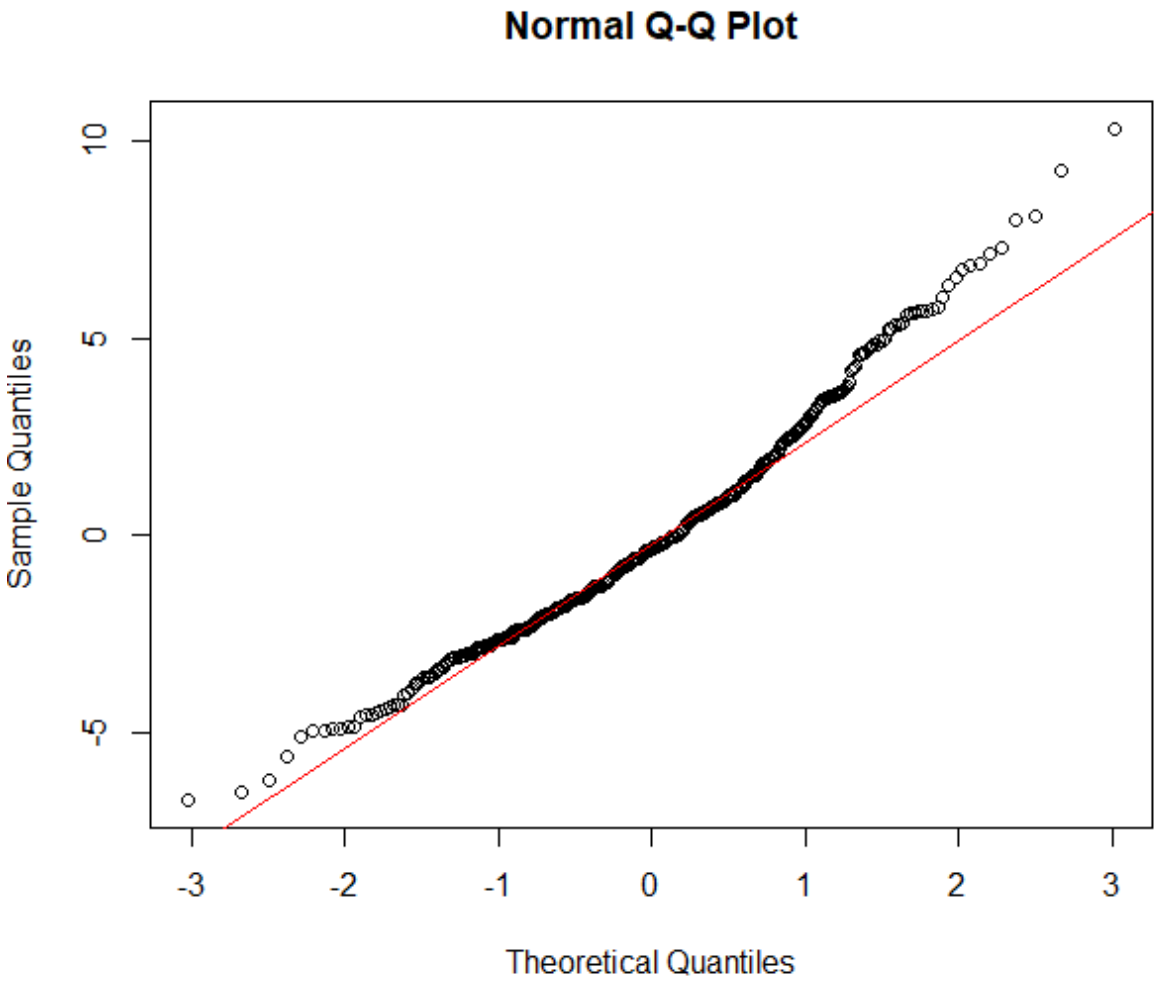
For the regression model to be valid, key assumptions for the model's error terms need to be valid including independence, zero expected value and constant variance. A residuals plot was generated to evaluate this visually as shown in Exhibit 4. Visually there didn't appear to be any relationship between the residuals and a player's Wingspan. The variance does appear to increase slightly as Wingspan increases, suggesting that variance is not constant. This could impact the quality of the parameter estimates and generate less precise predictions.

Exhibit 4: Residual Plots



Normality of the residuals was also analyzed using a Q-Q plot shown in Exhibit 5. The dispersion of points appears to be near normal around the center but diverges at either end suggesting that the tails are heavier than a normal distribution. Considering the large sample size (n=418) this is not a major concern nor would it have major implications for the robustness of the regression model.

Exhibit 5: Q-Q Plot



Results

The results of the simple linear regression are shown in Exhibit 5. Overall, the model is statistically significant with an high t-statistic and low p-value. Based on the slope estimate, it appears that a one inch increase in Wingspan results in additional 0.74 Total Rebounds over 100 ball possessions. This finding is consistent with expected results. The R^2 value of nearly 0.5 suggests some positive relationship with nearly half of the model variance explained.

Exhibit 5: Linear Regression Results

- **The F-statistic is 379.9 (1, 416df) with p-value <2e-16 and R^2 = 0.4773.**

Parameter	Estimate	Std. Error	t-statistic	p-value
Intercept	-52.18849	3.14692	-16.58	<2e-16
Wingspan	0.74191	0.03807	19.49	<2e-16

The following takeaways are inferred from the analysis:

- The initial approach to use Wingspan Advantage appears to be the right approach for the regression model. This likely due the potential confounding impacts of having wingspan and height in the calculation.
- There does appear to be a linear relationship between Wingspan and Total Rebounds as demonstrated through visual inspection and the results of the linear regression
- The resulting linear model suggests that each additional inch of wingspan increases the total number of rebounds by nearly 0.5 and nearly half of the total variance was explained by the model

This data set includes additional variables that could be used to further explore and refine the regression model and improve the understanding of the impact of player's Wingspan impact on rebound performance. In pursuit of this goal, it is recommended that a multiple regression model be fit to the data considering the following additional variables:

- **Player Position:** Include categorical variables to account for differences in a player's position since this could impact rebound level due to differences in defensive roles.
- **Player Height:** Incorporate player's height as this could help further refine the relationship between wingspan and and rebounds since this could be a confounding variable and could help isolate the true impact on rebound performance.
- **Refine Rebounds Definition:** Apply the model to offensive vs. defensive rebounds to determine if there is meaningful difference and refine as appropriate.

Further analysis of this relationship could help yield useful insights that could be used by those interested improved predictions for player performance that can help NBA teams improve their team performance.

References

Data Source:

https://data.scorenetwork.org/basketball/nba_wingspans_and_performance.html

Github Page:

https://github.com/janblackbu/Math-261-Projects

Foot Notes Sources:

1: https://en.wikipedia.org/wiki/Forbes_list_of_the_most_valuable_NBA_teams

2: https://news.mit.edu/2025/basketball-analytics-investment-nba-wins-and-other-successes-0325

Appendix: Data Definitions

Field Name	Definition	Data Type
name	Player's full name	Text
team	Player's team abbreviaation	Text
position	Player's primary position	Categorical
height_inches	Player's height in inches	Real
wingspan_inches	Player's wingspan in inches	Real
wingspan_advantage	wingspan_inches - height_inches	Real
g	Games played during season	Integer
mp	Minutes played during season	Integer
pts	Points scored per 100 possessions	Real
orb	Offensive rebounds per 100 possessions	Real
drb	Defensive rebounds per 100 possessions	Real
trb	Total Rebounds per 100 possessions	Real
ast	Assists per 100 possessions	Real
stl	Steals per 100 possessions	Real
blk	Blocks per 100 possessions	Real
o_trg	Offensive Rating: (Points Scored Possessions) / 100 possessions	Integer

Field Name	Definition	Data Type
d_rtg	Defensive Rating: (Opponent Points / Opponent Possessions) / 100 possessions	Integer
e_fg_percent	Effective Field Goal Percentages (FGM Made + 0.5 3 Point FGM Made) / FG Attempts	Real
ft_percent	% of Free Throws Made	Real
fg_percent	% of Field Goals Made	Real
avg_dist_of_fg	Average distance of field goal attempts (in feet)	Real
2pt_rate	Proportion of shots that are 2pt attempts	Real
3pt_rate	Proportion of shots that are 3pt attempts	Real
2pt_percent	Field Goal % on 2pt attempts	Real
3pt_percent	Field Goal % on 3pt attempts	Real
dunk_rate	% of Field Goal attempts that are dunks	Real