**Introduction/Background:**

Over the years, data analytics has become more relevant in college and professional sports. There have also been new variables or statistics that have come out that measure a variety of different things including player ability, team ability, etc. One of these statistics is plus minus, which is used in basketball. Plus minus is a measure of how a player impacted a game when they were in (Plus/Minus - Nba.Com: Jr.. NBA.”). Ideally, a player with a larger positive plus minus indicates this player had a big impact and the team scored more points than the opponent. The plus minus statistic does have some drawbacks though. A player could have a positive plus minus, but this could also be dependent on who else was on the court at the time. If all the other teammates were starters, then the plus minus might be biased. There are other plus minus statistics that adjust this problem, but datasets that already exist have the original plus minus, so for research purposes the original plus minus will be used in this analysis.

In this analysis, the relationship between plus minus and points, by position type will be analyzed using NBA data from the 2023-2024 season, using R. The data was obtained through two separate datasets that were then joined together. The first dataset is one that is already in R through a package called “nbastatR”. The second data set was one that contained the players position, which was needed for the analysis and was obtained through github. This dataset was called “2023-2024 NBA Player Stats” and included both regular season and playoff data. For the analysis only regular season data will be used.

The relationship is being investigated, specifically for the 2023-2024 regular season because of a transition in the way teams have been playing offense. Many teams now have either been trying to build teams by starting with young players and developing them, some have used centers as the main point of the offense, etc. With this analysis, the importance of a player type can be analyzed. Obviously, each players skill is different and some players are outliers based on their position. But the importance of each player type could be something teams could maybe use. This could also be useful in seeing the best combinations of players to put on a floor.

As mentioned above the analysis involves examining the relationship between plus minus and points by position type. The analysis will answer the question of what is the association between average plus minus and average points based on position type in the NBA? Plus minus and points are numeric, while position type is a factor variable. Plus minus is the response variable while points and position type are the predictor variables. One thing to note is for each player in the dataset, plus minus and points were averaged. Without averaging for each player there would have been over 40,000 rows. When merging the datasets together, some players on different data sets were listed differently usually due to asterisks or symbols in their name and due to limited time, these issues were not fixed. So, the dataset contains 469 players or 469 data points altogether. Plus minus was defined above and the number of points is the average number of points per game for each player in the dataset for the 2023-2024 regular season. There are five different positions in the NBA: center, power forward, point guard, small forward, and shooting guard. Typical responsibilities and player type will not be described in detail or explained, but it should be known that for each different position type, the types of players usually have different weights, heights, skills, etc. Because of this, an interaction for points and position type will be included in the model to allow for different slopes based on the position type. Here are some summary statistics based on position type.

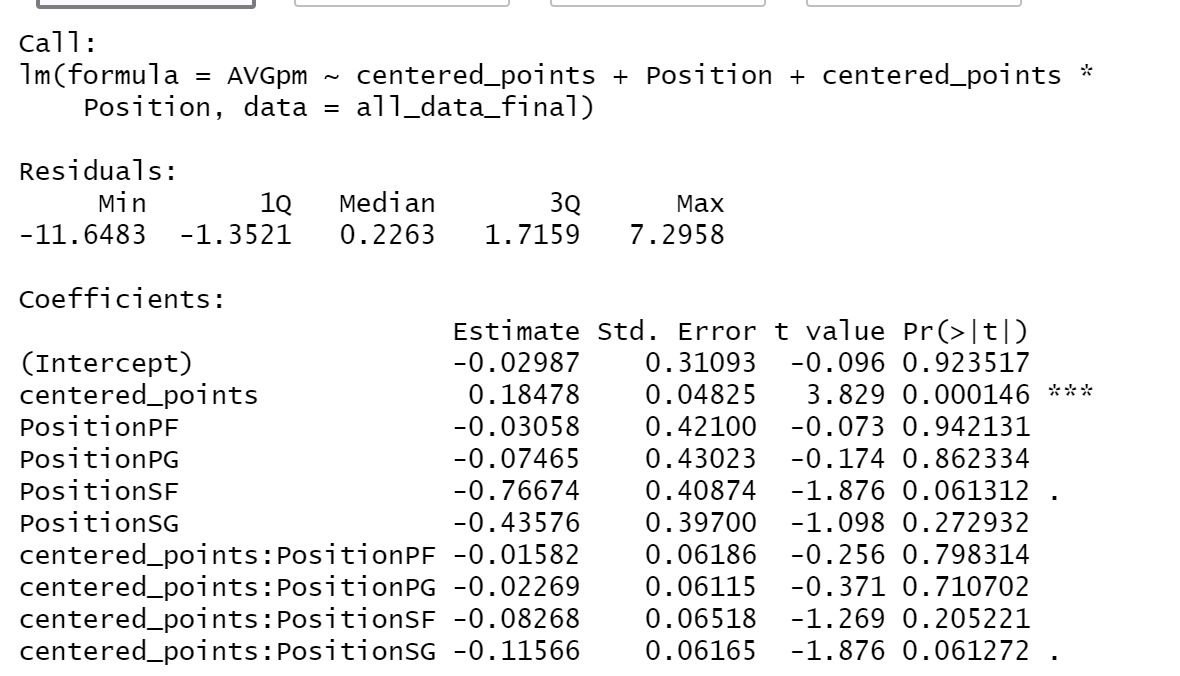
######Include summary statistic kable table

For the model, centering points was done for interpretation purposes. This was done by taking each player’s average points and subtracting that from the mean of the average points for the position the player is.

For the analysis, a model will be created using stan\_glm in r. stan\_glm will be used instead of lm because using stan\_glm probability statements could be made based on values. As mentioned above, average plus minus will be the response variable while average points scored and position type will be predictors. There will be an interaction between average points scored and position type to allow for different slopes based on position type.

Natural log transformations were considered for the response and numeric predictor, doing a transformation of only the response, another one for only the predictor, and then one for both. Visually, this did not help much with the relationship between average points and average plus minus, so no transformations were used for the model.

Here are the results from the model.



References:

“Plus/Minus - Nba.Com: Jr.. NBA.” *NBA.Com*, 13 Oct. 2015, jr.nba.com/plusminus/.