

**GROUP B**

# **Integrative Experience Step 5: Beyond the Bubble**



**By: Aditya Narayanan, Eilat Tabak, Nikhil Rajkumar, Vindhya Rachur**



# Question of Interest

Our group will be analyzing whether NBA players salaries for the 2019-2020 NBA season are correlated to their performance statistics in the 2018-2019 NBA season.

It is important to note that this data is comprised of 307 players who played for a single team during the 2018-2019 season and signed with one team for the 2019-2020 season.





# Data

The overall sample contains statistics on 307 players with 31 variables. All the covariates are from the 2018-19 season. We filtered out the necessary covariates and used only those, and they are as follows:

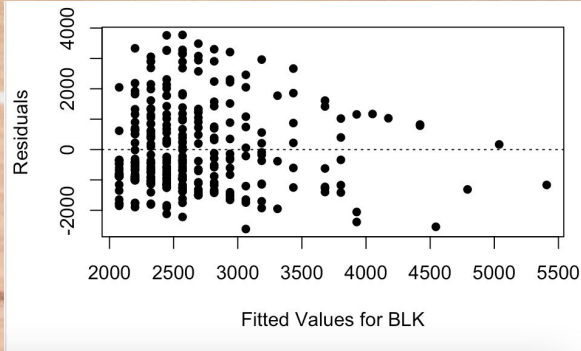
1. Position
2. Games Started (GS)
3. Average Points per Game (PTS)
4. Effective Field Goal Percentage (eFGP)
5. Offensive Rebounds (ORB)
6. Blocks (BLK)
7. Average Steals per Game (STL)
8. Defensive Rebounds (DRB)



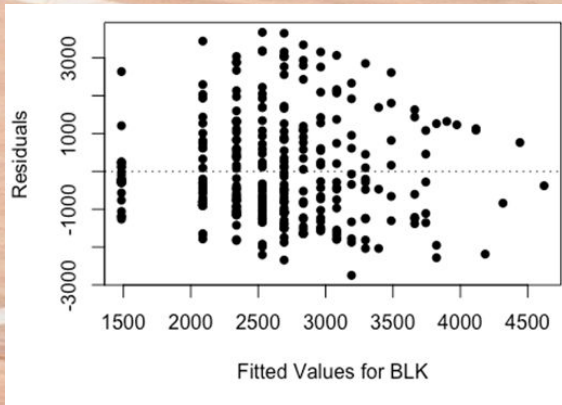
|            | Salary1920 | GS        | PTS       | eFGP      | ORB       | BLK       | STL       | DRB       |
|------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Salary1920 | 1.0000000  | 0.5708623 | 0.7285313 | 0.1867345 | 0.2860845 | 0.3175579 | 0.5668971 | 0.5691693 |
| GS         | 0.5708623  | 1.0000000 | 0.7073658 | 0.2655435 | 0.3931086 | 0.4089545 | 0.6083450 | 0.6040704 |
| PTS        | 0.7285313  | 0.7073658 | 1.0000000 | 0.2772421 | 0.3354827 | 0.3477263 | 0.6120097 | 0.6745208 |
| eFGP       | 0.1867345  | 0.2655435 | 0.2772421 | 1.0000000 | 0.3557261 | 0.3063724 | 0.1046881 | 0.3241580 |
| ORB        | 0.2860845  | 0.3931086 | 0.3354827 | 0.3557261 | 1.0000000 | 0.6646145 | 0.2689106 | 0.7143393 |
| BLK        | 0.3175579  | 0.4089545 | 0.3477263 | 0.3063724 | 0.6646145 | 1.0000000 | 0.2689024 | 0.6397746 |
| STL        | 0.5668971  | 0.6083450 | 0.6120097 | 0.1046881 | 0.2689106 | 0.2689024 | 1.0000000 | 0.5159893 |
| DRB        | 0.5691693  | 0.6040704 | 0.6745208 | 0.3241580 | 0.7143393 | 0.6397746 | 0.5159893 | 1.0000000 |

# Initial Simple Linear Regression Model

Before Transformation:



After Transformation:

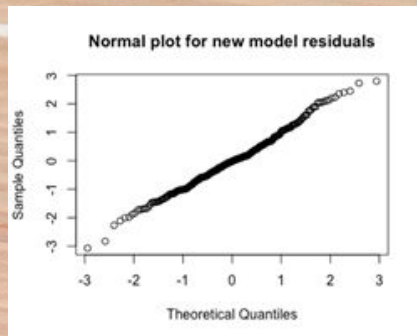
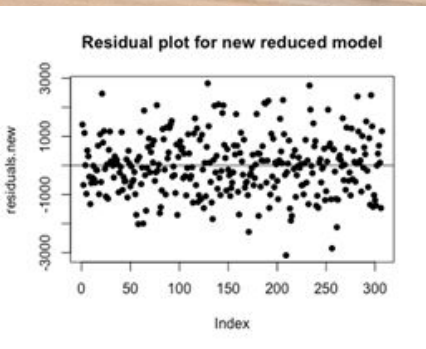


- From our initial analysis, we found that our theory of having a linearly positive correlation between our covariates and salary said that the better performance an NBA player had in the 2018-2019 NBA season, the higher their salary would be in the 2019-2020 NBA season.
- From our 7 covariates, we were able to use transformations to create a simple linear regression model that followed normal linear regression assumptions
- Assumptions being made: Linearity, Normality and Independence

| Observations | Residual Std. Error | $R^2$      | Adjusted $R^2$ |           |
|--------------|---------------------|------------|----------------|-----------|
| 307          | 0.5209              | 0.1258     | 0.1229         |           |
|              | Estimate            | Std. Error | t value        | Pr(> t )  |
| (Intercept)  | 1485                | 178        | 8.342          | 2.558e-15 |
| sqrt(BLK)    | 1909                | 262.2      | 7.281          | 2.839e-12 |



# Multiple Linear Regression Model



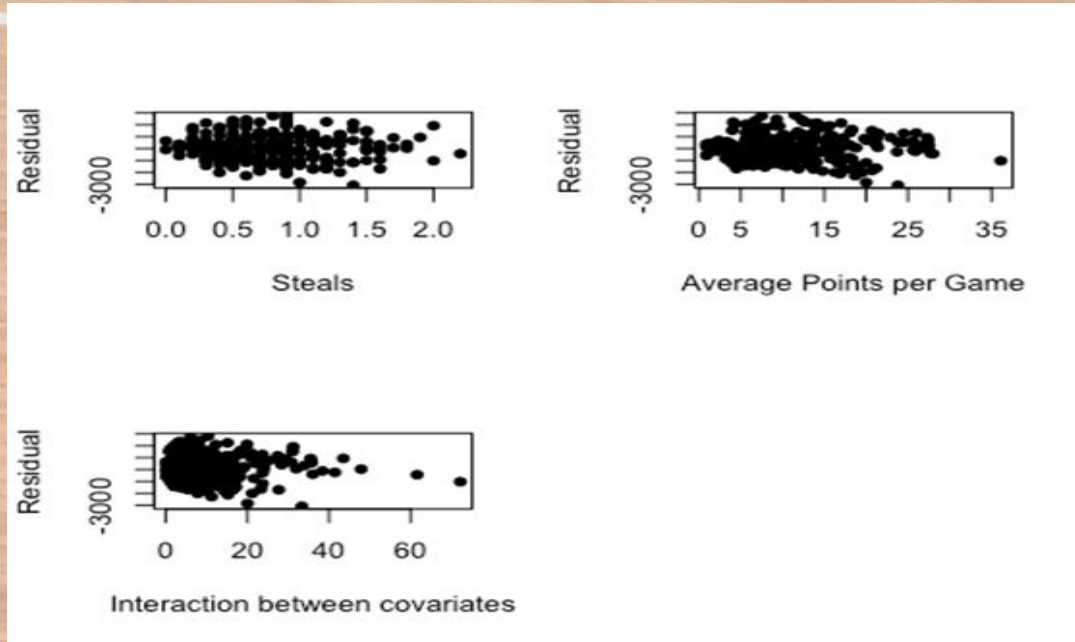
- First, we fit a multiple linear regression model with all of our transformed covariates and response.
- Hypothesis tests resulted in average points per game(PTS) and average steals per game (STL) to be the most significant covariates for the model.  
 $H_0 : B_k = 0$ ;  $H_a: B_k \neq 0$ .
- Our final model was **Salary = 613.8 + 738.2(PTS) + 140.8(STL)**



| Intercept | STL   | PTS   |
|-----------|-------|-------|
| 613.8     | 738.2 | 140.8 |

# Interactions Between Covariates

There was not a strong enough interaction between number of steals and average points per game to include the interaction term in the final linear regression model as shown in the graph. When plotting the respective covariates against the residuals as well as the interaction term against the residuals, the plot shows that the variance is not normally distributed.





# Limitations

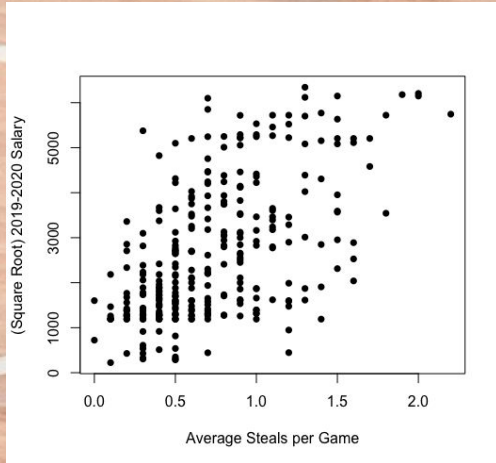
- After the transformations, we had to dismiss 5 possible covariates indicating that the transformations could have been more accurate.
- Any of those 5 possible covariates could have been associated with the 2019-2020 salary.



# Results

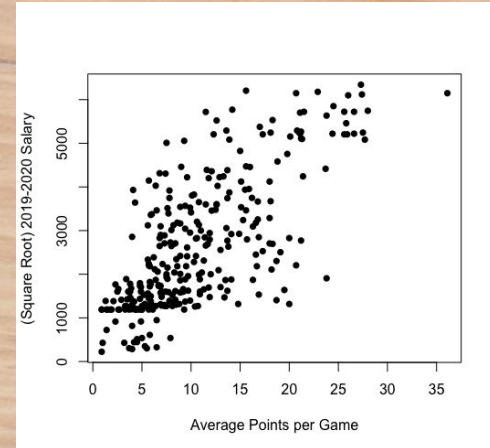
After completing Hypothesis testings for all covariates, it appeared that the average points per game (PTS) and average steals per game (STL) were the only two covariates that had significant evidence of association for the 2019-2020 salary.

$$Y = 613.8 + 738.2PTS + 140.8STL$$



# Conclusion

We are 95% confident that the average points per game and average steals per game in the 2018-2019 season are associated with the players 2019-2020 salary.







Questions?

