Modern Basketball: Which Player Statistics Affect NBA Salaries

STAT 3220 - Fall 2021

Research Questions

Can NBA player salaries be predicted using statistics?

- Do players with higher point scoring or higher usage rates (a newer, more advanced basketball statistic) earn higher salaries?
- Do multi-position players earn higher salaries?
- Does a player's age affect their salary?

Introduction

- With data analytics becoming an emerging field in sports and player data becoming increasingly available, methods of management and player development are changing
- Sports analytics first gained popularity in the moneyball era of baseball; now spreading to other sports and we want to see the effect on basketball (Sigler)
- Players want to make as much money as possible because they never know when their career will end; knowing which statistics count could put them in a better position to do this (Lopez)

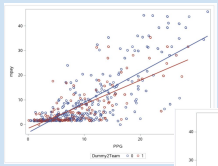
Goal: Evaluate if certain statistics can predict NBA player salaries

Data Summary

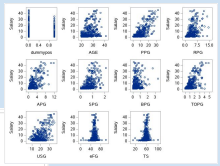
Salary: Player's 2021-22 season pay, Quantitative
Team: Player's 2021-22 team, Qualitative
Position: Basketball Position for each player, Qualitative
Age: Player's age at start of 2021-22 season, Quantitative
PPG: Average points scored per game, Quantitative
APG: Average assists per game, Quantitative
RPG: Average rebounds per game, Quantitative
SPG: Average steals per game, Quantitative
BPG: Average blocks per game, Quantitative
TOPG: Average turnovers per game, Quantitative
USG: Percentage of team involvement while player is on the floor, Quantitative
eFG: Field goal percentage with 3-pointers worth more than 2-pointers, Quantitative
TS: Shooting efficiency percentage on all shots (including free throws), Quantitative

*Original datasets collected from reliable sports analytics websites and modified by removing similar variables to avoid multicollinearity

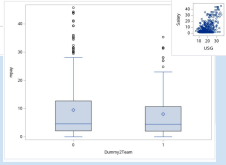
EDA



Number of positions boxplots - no true difference



Test for qualitative-quantitative interaction between ppg and number of positions shows no true relationship -> no interaction (similar for all)



No true patterns in scatterplots of explanatory variables vs. salary - no higher order terms

Three-stage Model Analysis

1. Quantitative predictors (test quantitative interactions):

$Salary = \beta_1 Age + \beta_2 ppg + \beta_3 rpg + \beta_4 apg + \beta_5 spg + \beta_6 bpg + \beta_7 topg + \beta_8 usg + \beta_9 efg + \beta_{10} ts + \beta_{11} ppgapg + \beta_{12} ppgts + \beta_{13} ppgefg + \beta_{14} ageusg$

New: $Salary = \beta_1 Age + \beta_2 ppg + \beta_3 apg + \beta_4 spg + \beta_5 bpg + \beta_6 usg + \beta_7 efg + \beta_8 ppgapg$

2. Qualitative predictors (no qualitative interactions):

$Salary = \beta_1 Age + \beta_2 ppg + \beta_3 apg + \beta_4 spg + \beta_5 bpg + \beta_6 usg + \beta_7 efg + \beta_8 ppgapg + \beta_9 Dummy2Team$

Where Dummy2Team = {1 if player plays multiple positions, 0 otherwise}

New: $Salary = \beta_1 Age + \beta_2 ppg + \beta_3 bpg + \beta_4 ppgapg$

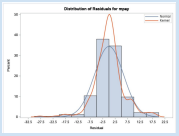
3. Final Model Suitability:

Final: $Salary = \beta_1 Age + \beta_2 ppg + \beta_3 bpg + \beta_4 ppgapg$

F-Statistic: 185.1 Root MSE: 5.91667
P-value: <0.0001 Adj. R-Squared: 0.6924

Assumptions

*Response variable transformations were tested to fix the constant variance. While the log(salary) model was the closest to fixing the assumption, there was still a violation so we kept the model with the highest adjusted R-squared value (other assumptions met)



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	25991	6497.69698	185.61	<.0001
Error	324	11342	35.00704		
Corrected Total	328	37333			

Coding Quantitative Variables:

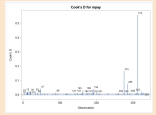
- Coding of quantitative variables occurs in two forms:
  - Transformation of salary
  - Restriction of age and removal of variables with empty values

Variable Screening:

- Stepwise regression yields Age, ppg, apg, spg, bpg, and eFG as significant predictors of NBA player salary with SLEntry and SLStay values of 0.15
- No multicollinearity because all VIF values <10 and average VIF <3

Influential Observations:

- There are two observations noted, observations 275 and 312, that are both influential and outliers.



Conclusion

Prediction Equation:

Salary (in millions of dollars) = -11.879 + 0.394*age + 0.562*ppg + 3.244*bpg + 0.105*ppgapg

- Bpg (blocks per game) coefficient is the largest
  - Increase in one block per game would be estimated to increase the player's salary by over three million dollars
- With our background knowledge, we know it appears to have more weight because even the top shot-blockers in the NBA average at most around three blocks per game
  - For ppg (points per game), the league leaders average over 30 points per game, which is why bpg is weighted more

Example:

Aaron Gordon
Actual Salary = 16.409 million dollars
Predicted Salary = -11.879 + 0.394*25.67 + 0.562*14.6 + 3.244*0.8 + 0.105*(14.6*4.2) = 15.474 million dollars
Residual = 0.935 million dollars

Effectiveness:

- Model is useful (with caution), but can be improved in the future
- Violation of constant variance limits the model to application only within the experimental region
- Constant variance violation hurts the validity of the model, but the adj. R-squared of 0.6924 is relatively strong and actual R-squared value over 0.7 indicates that over 70% of the variation is explained by this model
- Large Root MSE is concerning, big spread in model

Areas of future improvement:

- Better predictive dataset would only include players who received new contracts following year (measure yearly difference over 5-year span)
  - Rookie contracts based on predicted performance (Age > 22 in model)
  - Model is relatively reliable for non-rookies
- NBA minimum salary could be the reason for the lack of constant variance
- Less playing time means a player has lower values of each statistic, which makes the data right skewed and heavily concentrated at those low values

Works Cited

Lopez, Aaron. "Life after NBA Comes Sooner than Many Players Think." Denver Nuggets, NBA.com/Nuggets, 21 July 2015, https://www.nba.com/nuggets/features/junior_bridgeman_20100610.html.

Sigler, Kevin, and William Compton. "NBA Players' Pay and Performance: What Counts?" The Sport Journal, 19 June 2018, https://thesportjournal.org/article/nba-players-pay-and-performance-what-counts/.

"2020-2021 NBA STATS: Player Box Score & Advanced Metrics." *NBAstuffer*, 24 May 2021, https://www.nbastuffer.com/2020-2021-nba-player-stats/.

"2021-22 NBA Player Contracts." *Basketball Reference*, 2021, https://www.basketball-reference.com/contracts/players.html.

Tyler Gorecki, Stone Larson, Jiahao Xie, Zain Zahir