Factors Determining NBA Salaries

Info Crunchers

Introduction

In a Dario AS article, it was revealed that in the 2022-2023 NBA season, Ishmail Wainright of the Phoenix Suns is the lowest-paid player at $633,891, while Steph Curry, the highest-paid player, earns $48,070,014 annually (Rudder, 2023). Thus, this analysis investigates potential predictors of player salaries. This report addresses three key research questions focusing on these ongoing disparities. **Does a player with a higher true shooting percentage have a higher salary on average? Does a player that plays more games per season have a higher salary on average? Do older players have a high salary than younger players on average?**

In a prior study, it was found that a higher field goal attempt rate indicates a player's offensive importance, leading to higher pay (Yin, 2022). This suggests a potential link between salary and true shooting percentage where an increase in salary correlates with an improved true shooting percentage. Another study in The Sports Journal highlights the belief that longevity in the league, reflecting more playtime and enhanced skills, justifies higher salary demands (Sigler, 2018). The question arises: does playing more games per season and acquiring hands-on skills correlate with a higher salary? Lastly, an Atlanta Press study reveals evidence that for every 1-year increase in an NBA player's age, their salary increases by $0.414 million (Zhao, 2022). This implies older NBA players tend to earn more. The report will further explore the relationship between predictors like age and player salaries.

Methods and Analysis

A Multiple Linear Regression analysis was determined to be the most appropriate analysis. Prior to any modeling building, the response variable, Salary, is transformed using Box-Cox transformation sine the histogram of the response variable was heavily skewed right, making the response variable not suitable for regression. The test revealed a lambda (λ) of approximately 0.5, so the appropriate transformation was to take the square root of the response variable. Although the data still appears to lack perfect symmetry, the histogram of the response variable appears to be unimodal and only slightly right-skewed.

For the first step, a variable screening procedure was conducted with all 14 quantitative variables. First, stepwise regression was performed followed by backward elimination after fitting the first new model and discovering high individual and average VIF values. The backward elimination with a p-value removal level of 0.15 left FGA, TS., and GP as the only quantitative variables in the model.

For the second step, model building with quantitative variables was conducted. The initial model was: $E(sqrtSalary) = \beta_0 + \beta_1 GP + \beta_2 GS + \beta_3 MP + \beta_4 MPG + \beta_5 PER + \beta_6 TS. + \beta_7 FT. + \beta_8 X3P. + \beta_9 UR. + \beta_1 0OWS + \beta_1 1DWS + \beta_1 2WS + \beta_1 3VORP + \beta_1 4FGM + \beta_1 5FGA$. The revised model based on variable screening from the first step was: $E(sqrtSalary) = \beta_0 + \beta_1 FGA + \beta_2 TS. + \beta_3 GP$. To test significance, the Global F test of this model was performed at the 0.05 alpha level; with an F statistic of 24.05 and p-value of 1.078e-11, this quantitative model was deemed significant.

For the third step, model building with qualitative variables was conducted. Dummy variables were created for the variables Age and allStar. The initial model was: $E(sqrtSalary) = \beta_0 + \beta_1 FGA + \beta_2 TS. + \beta_3 GP + \beta_4 AgeU + \beta_5 allStarY$ where AgeU is 1 if age is under 30 and 0 otherwise while allStarY is 1 if noted as yes and 0 otherwise; in the model, the base level is those ages 30 and older and non-All-Star players. After conducting individual T-tests for the dummy variables, significant p-values were found at the 0.05 alpha level (1.54e-08 for Age and 0.01549 for allStar). Thus, the linear relationships were retained in our model without any revisions.

For the fourth step of the model-building process, higher-order terms and interactions were considered. The relationship between the explanatory and response variables exhibits no polynomial relationship that could be modeled. Furthermore, assessing interactions between qualitative vs. qualitative and qualitative vs. quantitative, it is shown that no lines intersect on the interaction plots, and the slopes of each level of the variables on the scatterplots did not change. Lastly, the quantitative vs. quantitative relationship, being difficult to visualize, did not present as an issue either as potential interactions were already addressed in the variable screening portion. Thus, the same model remains.

In the fifth step of the modeling process, we assessed the model's suitability for analysis. The lack of fit assumption is supported by residual plots, indicating no clear non-linear trend. While the constant variance assumption revealed some "fanning in" patterns, the residual vs fitted plot suggests constant variance. Although the normality assumption shows slight deviations in the histogram and QQplot, regression remains robust for large samples (n>50), and the independence assumption is reasonably met as we lack time series data. We transformed the response variable earlier (square root) to address potential issues, determining no further transformation was necessary. Proceeding, we should remain cautious of a potentially skewed histogram.

Prior to completing the model-building process, multicollinearity was again checked. With no individual VIFs above 10, and a mean VIF of 1.446, multicollinearity, while still present, was not concluded to be an issue.

The seventh step was to judge the influential observations using various techniques. The five outliers were Kevin Love, Jonathan Isaac, Cade Cunningham, RaiQuan Gray, and Kevin Durant. Love, Isaac and Gray are outliers in the y-direction (± 2 threshold for studentized residuals). Cunningham, Gray, and Durant are outliers in the x-direction (>0.12 leverage threshold). Isaac and Gray are also influential outliers (>0.04 threshold for Cook's Distance). No observations were removed as none were identified as known errors. The model remains robust showing only slight changes in the individual p-values of each beta and a minor increase in the adjusted R-squared value with or without influential points.

Results

Based on the analysis, the final prediction equation is: $\widehat{sqrtSalary} = 1.435 + 0.148FGA - 0.092TS. + 0.014GP - 1.229AgeU + 1.054allStarY$. The model is sufficient based on a p-value of less than 2.2e-16 for the global F test and an adjusted R-square value of 0.60.

Conclusions

When other variables are held constant, there is a 0.148 million dollar increase in square root of salary for each field goal attempt by a player. For instance, for Markelle Fultz (observation 14), using the prediction equation, his predicted square root salary is 1.988 million dollars. The actual square root of Fultz's salary is 4.062 million dollars. This makes the residual 2.074 million dollars. When changed from square root of salary to salary, Fultz's predicted salary equals 3.952 million dollars while his actual salary is 16.500 million dollars. This gives a residual of approzimately 12.548 million.

In terms of limitations, certain duplicated players, due to being traded and being on multiple teams, were removed which can slightly hinder having a representative sample. In order to improve upon the analysis, duplicate players could be incorporated into the analysis by combining the multiple observations into one since their salary stays consistent through the season. Since player salaries are generally a result of multiple years of play, past seasons could've been incorporated into the data. This would require looking at a player's averages through a specific interval of years rather than a single season. Age could have also been broken down into more levels, painting a clearer picture of how salaries would change with age groups. Lastly, a larger sample size could've been taken to have a better representation of the population.

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Player	Player	Individual players in the NBA 2022-23 season
Salary	sqrtSalary	Square root of the player's salary in millions of U.S. dollars
Position	Position	Player's NBA team position; C=center, PF=power forward, PG=point guard, SG=shooting gaurd, SF=small forward
All Star	allStar	Players indicated as a part of the All-Star team for the 2022-23 season
Age	Age	Player's age as of February 1st of the 2022-23 season categorized as under 30 and 30 and over
Games Played	GP	Number of games the player played in
Games Started	GS	Number of games which the player started in
Minutes Played	MP	Number of total minutes the player played
Minutes Per Game	MPG	Average minutes per game played
Player Efficiency Rating	PER	Measure of per-minute production standardized such that the league average is 1

Variable Name	Abbreviated Name	Description
True Shooting Percentage	TS%	Measure of shooting efficiency accounting 2-point field goals, 3-point field goals, and free throws.
Free Throw Percentage	FT%	Percent of free throws made to attempted
3 Pointer Percentage	3P%	Percent of field goal attempts from 3-point range
Usage Rate Percentage	UR%	Estimate of percentage of team plays used by a player while they were on the floor
Offensive Win Share	OWS	Estimate of the number of wins contributed by a player on offense
Defensive Win Share	DWS	Estimate of the number of wins contributed by a player on defense
Win Share	WS	Estimate of the number of wins contributed by a player
Value Over Replacement Player	VORP	Box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season
Field Goal Made	FGM	Number of average field goals made per game
Field Goal Attempted	FGA	Number of average field goals attempted per game

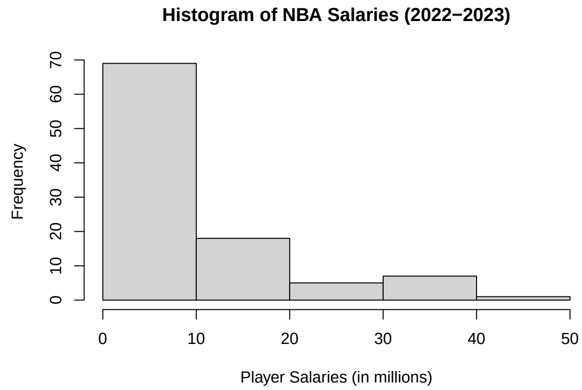
Appendix B: Data Rows

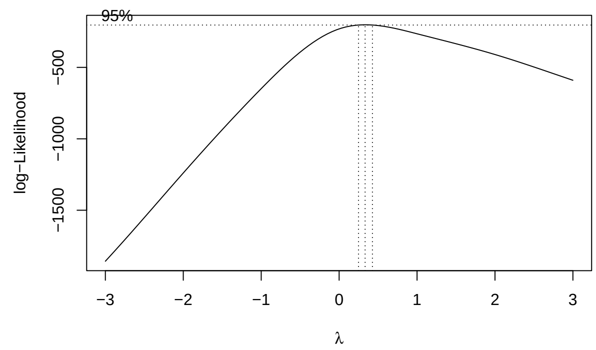
	Player	sqrtSalary	Position	allStar	Age	GP	GS	MP	MPG	PER
1	Cameron Johnson	2.426499	Other	No	Under 30	42	41	1199	28.5	17.1
2	Pat Connaughton	2.393406	Other	No	30 and Over	61	33	1443	23.7	10.0
3	Kevin Love	5.527836	Other	No	30 and Over	62	20	1240	20.0	13.1
4	Anthony Gill	1.355024	Other	No	30 and Over	59	8	624	10.6	11.1
5	Keita Bates-Diop	1.370664	Other	No	Under 30	67	42	1452	21.7	14.9
6	KZ Okpala	1.379178	Other	No	Under 30	35	3	248	7.1	6.2

	TS.	FT.	X3P.	UR.	OWS	DWS	WS	VORP	FGM	FGA
1	0.617	0.842	0.404	20.6	2.1	1.4	3.5	1.4	5.3	11.3
2	0.531	0.659	0.339	13.8	0.7	1.8	2.5	0.3	2.7	6.9
3	0.548	0.879	0.334	19.2	0.6	2.0	2.6	0.8	2.7	6.8
4	0.604	0.731	0.138	12.3	1.0	0.4	1.3	-0.2	1.2	2.2
5	0.609	0.793	0.394	16.7	2.1	0.6	2.7	0.5	3.5	6.9
6	0.554	0.875	0.333	8.4	0.1	0.2	0.3	-0.1	0.5	1.1

Appendix C: Final Model Output and Plots

Response Variable





Variable Screening

```
                          Elimination Summary
------------------------------------------------------------------------
        Variable                    Adj.
Step    Removed     R-Square    R-Square     C(p)        AIC        RMSE
------------------------------------------------------------------------
   1    FGA           0.4602      0.4375    4.0016    315.8932    1.1345
   2    TS.           0.4568      0.4398    2.6003    314.5281    1.1321
   3    GP            0.4527      0.4415    1.2975    313.2623    1.1304
------------------------------------------------------------------------
```

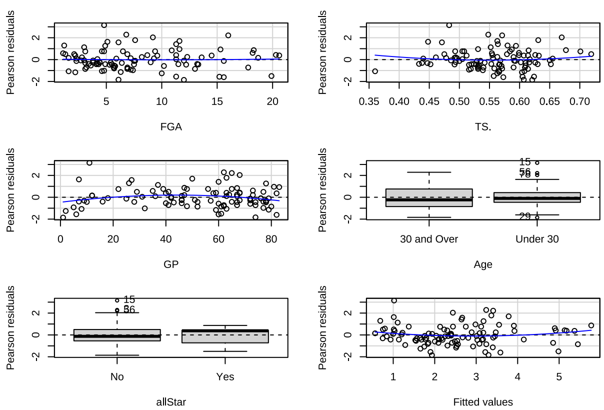
Regression Outputs

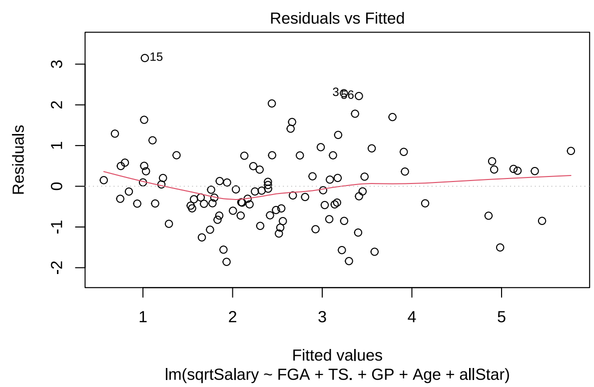
```
Call:
lm(formula = sqrtSalary ~ FGA + TS. + GP, data = nbaOg)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6185 -0.7208 -0.1664  0.6763  3.0483

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.671241   1.040986  -0.645   0.5206
FGA          0.157196   0.026146   6.012  3.3e-08 ***
TS.          2.362645   1.989577   1.188   0.2380
GP           0.012696   0.005734   2.214   0.0292 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.161 on 96 degrees of freedom
Multiple R-squared:  0.4291,    Adjusted R-squared:  0.4113
F-statistic: 24.05 on 3 and 96 DF,  p-value: 1.078e-11
```

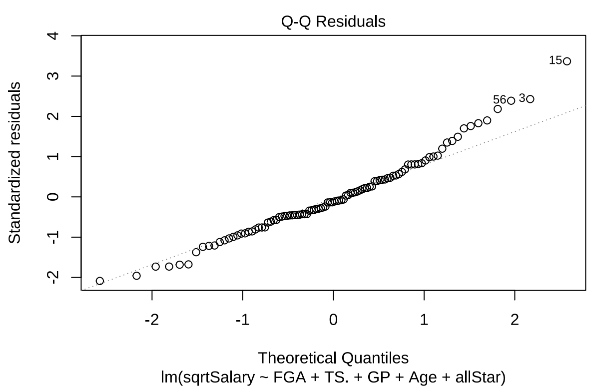
```
Call:
lm(formula = sqrtSalary ~ FGA + TS. + GP + Age + allStar, data = nbaOg)
```

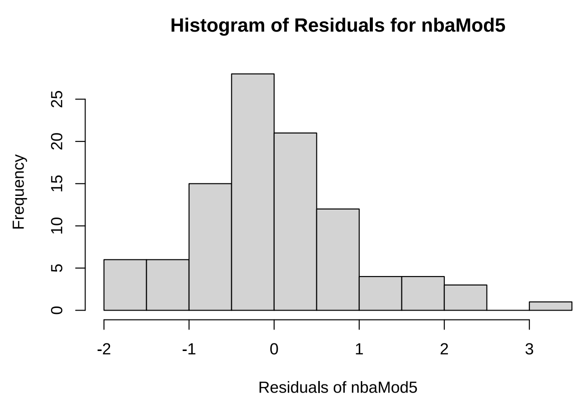
```
Residuals:
    Min      1Q  Median      3Q     Max
-1.8565 -0.5534 -0.1160  0.4953  3.1505

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.435126   0.915604   1.567  0.12038
FGA          0.147569   0.027358   5.394 5.12e-07 ***
TS.         -0.092240   1.685672  -0.055  0.95648
GP           0.013767   0.004761   2.892  0.00476 **
AgeUnder 30 -1.229494   0.215602  -5.703 1.36e-07 ***
allStarYes   1.053658   0.427340   2.466  0.01549 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9596 on 94 degrees of freedom
Multiple R-squared:  0.6178,    Adjusted R-squared:  0.5975
F-statistic: 30.39 on 5 and 94 DF,  p-value: < 2.2e-16
```

Assumptions





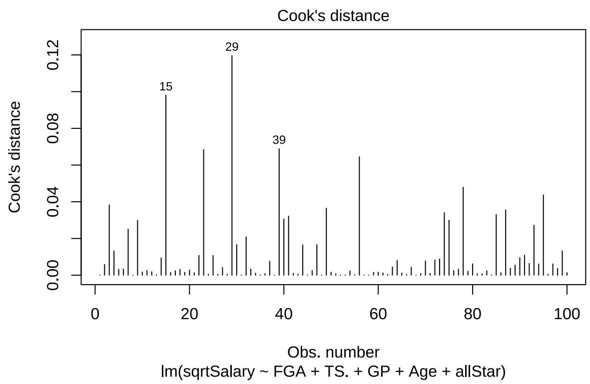


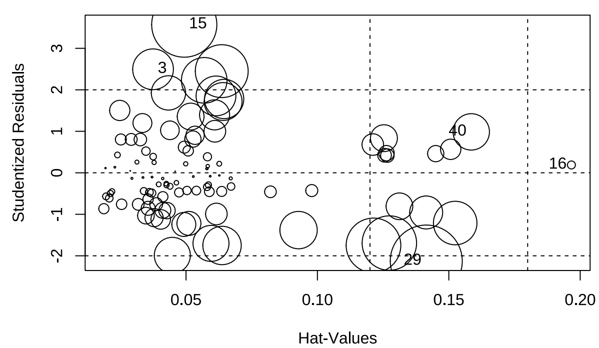


Outliers and Influential Observations

	StudRes	Hat	CookD
3	2.4949272	0.03744964	0.038238151

15	3.5714839	0.04932634	0.098043429
16	0.1891269	0.19664866	0.001474412
29	-2.1265972	0.14141828	0.119664837
40	0.9863533	0.15853529	0.030558294





Appendix D: References

HoopsHype. (2023). NBA Player Salaries. HoopsHype. Retrieved October 15, 2023, from https://hoopshype.com/salaries/players/

NBA. (2023). NBA Advanced Stats. NBA.com. Retrieved October 15, 2023, from https://www.nba.com/stats/players/traditional?SeasonType=Playoffs

Rudder, P. (2023, May 26). Who is the lowest paid NBA player in the 2022-2023 season?. Diario AS. https://en.as.com/nba/who-is-the-lowest-paid-nba-player-in-the-2022-2023-season-n/

Sigler, K. (2018, June 19). NBA players' pay and performance: What counts? The Sport Journal. https://thesportjournal.org/article/nba-players-pay-and-performance-what-counts/

Sports Reference. (2023, April 10). 2022-23 NBA Player Stats: Advanced. Basketball Reference. Retrieved October 15, 2023, from https://www.basketball-reference.com/leagues/NBA_2023_advanced.html

Yin, C., & Jamin, W. (2022). How do different factors affect the NBA players' salary. https://rstudio-pubs-static.s3.amazonaws.com/928945_ef3cd002eb2140e0a0875b74f805518f.html

Zhao, Y. (2022). Model prediction of factors influencing NBA players' salaries based on multiple linear regression. Proceedings of the 2022 2nd International Conference on Economic Development and Business Culture (ICEDBC 2022), 1439–1445. https://doi.org/10.2991/978-94-6463-036-7_213