Career Resume vs Salary: Does A Hitter's Performance and Accolades Impact Their Pay?

Malcolm Hsu[1]

December 16, 2023

Abstract

The goal of this paper is to assess the many components of a Major League baseball player's resume, awards and statistics included, in order to evaluate whether their salaries reflect the strength of their statistical season and accolades. Cross sectional data from the 2023 season from Baseball Reference, including batting average (AVG), on base percentage plus slugging percentage (OPS), all-star awards, most valuable player (MVP) awards, and years in the Major Leagues, in conjunction with agent-predicted/contract salary data from the USA Today database, allows for the assessment of player salary based on their prior year and career performance. Our results may help determine whether contracts are structured based on player performance or if there are other elements, in-game or non-in-game factors, that determine a player's salary. The number of awards won per player, while accounting for experience, has a significant impact on the player's salary, while batting average and OPS have certain caveats.

Keywords: salary, contract, AVG, OPS, all-star, MVP

[1] Hsu: mhsu01@wesleyan.edu. Mathematics & Economics, Wesleyan University, 2025

Career Resume vs Salary: Does A Hitter's Performance and Accolades Impact Their Pay?

1 Introduction

This paper investigates the different statistical effects of a Major League Baseball player's season and career in order to determine the correlation between statistics and salary. To measure a player's statistics, I use their batting average, which measures how many times a player successfully records a base hit, and on base plus slugging percentage, which is the sum of the player's on base percentage, how many times they reach base per at bat, and the player's slugging percentage, how many bases the player reaches per at bat. I also take into consideration the number of all-star selections and the number of most valuable player awards each player has won in their career, as well as the number of years each player has played professionally in the MLB. In order to collect the most accurate data, I only selected players that played in more than half of the games in the season, at least 81 out of the 162 regular season games. This cross-sectional data was collected from Baseball Reference.[2] To measure the player's salary, I took cross-sectional data from the USA Today database[3] which was collected during the 2023 season and included each player's predicted salary or contract salary.

Generally speaking, it would seem that a higher batting average and OPS, more awards, and more years played in the league would mean a higher salary.[4] However, this is not always the case as there are many other factors, some impossible to observe, that impact a player's salary. Salaries can be constructed and negotiated via a player's agent, or a player themself, working with the front office of the team. The MLB's Collective Bargaining Agreement sets a minimum

[2] https://www.baseball-reference.com/
[3] https://databases.usatoday.com/major-league-baseball-salaries-2023/
[4] See figures 4 and 5 for BA vs salary and OPS vs salary and figures 1 and 2 for salary by all-star and mvp

salary for every player of the team's 40-man roster, which can be increased via negotiation. Along with the assured base salary, a lot of players also have performance based incentives, milestone rewards, and provisions, though incentives are not included in a player's salary.

2 Literature Review

Most of the economic publishings done on the factors of an MLB players' salary were done in the 20th century, and analyzed statistics in general, such as batting average. Also, all but a couple publications dealt with the payroll of each team in comparison to their standings of that same year. These approaches are different from mine because I take into account three other variables, number of all-star awards, number of MVP awards, and years in the league. The biggest difference, however, is that none of the papers analyze the correlation between pay and performance at the individual level. In Depken's paper, he aims to describe the correlation between wage disparity and team performance, claiming that teams with greater wage disparity perform worse. This claim is similar to mine, as we both are trying to connect performance to pay, however, Depken writes at a team level, while I write at an individual level. Similar to Depken, Tao, Chuang, and Lin write about salary dispersion and team performance, arguing that a team's payroll, as a control variable, has a negative impact on team performance, while payroll relative to the league as a whole has an insignificant effect on team performance. Finally, Watnick attempts to set a base for regressions on statistical data from the MLB by salaries. He describes the process and interpretation of key variables that would affect a player's salary. This model is closest to the one I am estimating as Watnick's data was taken from a player-based database, having statistics for every player.

3 Empirical Approach

I aim to use multivariate ordinary least squares to estimate the regression model for salary based on batting average, OPS, the number of all-star awards, and the number of MVP awards per player. This model is estimated by the following equation:

$$salary_i = -5805 - 25.64ba_i + 14.24ops_i + 17.01allstar_i + 46.85mvp_i + 10.1years_i + \epsilon_i$$

$Salary_i$ measures the i^{th} predicted salary based on the model, ba_i is the i^{th} observation of batting average, ops_i is the i^{th} observation of OPS, $allstar_i$ is the i^{th} observation of the number of all-star awards, mvp_i is the i^{th} observation of the number of MVP awards, years is the i^{th} observation of the number of years the player has played, and $error_i$ is the error term of the i^{th} observation. In this model, ba_i and ops_i are quantitative variables, while $allstar_i$, mvp_i, and $years_i$ are categorical variables. The constant, -5805 indicates that if a player were to have a zero batting average and OPS, have won no awards, and never played in the professional league, their salary would be -5805 thousand dollars. This however, is not statistically or economically significant, as explained in the later section. The coefficient of batting average indicates for a 1 point increase in batting average, salary will decrease by 25 thousand dollars. The coefficient of OPS indicates for a 1 point increase in OPS, salary will increase by 14 thousand dollars. The coefficient of allstar indicates that for each all-star award won, assuming all players have played the same number of years in the MLB, salary will increase by 17 thousand dollars. The coefficient of mvp indicates that for each MVP award won, assuming all players have played the same number of years in the MLB, salary will increase by 47 thousand dollars. Finally, the coefficient of years indicates that for each additional year played in the MLB, salary will increase by 10 thousand dollars.

4 Data

4.1 Statistics and Awards Data

The data for batting average, OPS, number of all-star and MVP awards, and years played in the MLB was taken from Baseball Reference, the leading website in baseball statistic databases. Here, stats are tracked and stored per player, per team, of every player who has played professionally, including all levels of the MLB, some semi-professional leagues, and even collegiate players. The data is kept in line with the official statistics charted during and after every Major League game, to ensure that the statistics presented are accurate.

For each team, I selected the data of players who played in more than half of the regular season games, to get a more accurate representation of "full-seasoned" players. There are a handful of rookie players who have had midseason breakouts, and their statistics would be extremely inflated compared to those players who played full seasons, and thus would affect the data. I chose batting average and OPS as the two pure statistics as they most accurately represent the true performance of each player. Batting average cannot be affected by secondary factors such as walks and hit by pitches, that is, it is purely based on if the player can get a hit or not. OPS is a statistic that essentially measures the batting strength of each player, accounting for extra base hits and how many times the player reaches base. The key distinction between OPS and batting average is that for batting average, a home run and a single are counted the same, while for OPS, the hitter is awarded four bases as well as one time reaching base, truly capturing the type of hit. I next chose the number of all-star and MVP awards since these are the two most coveted awards for hitters. Only two MVP awards are given each year, one for each of the National and American League, and only forty hitters are selected as all-stars, twenty for each league. Since players have a greater chance of winning an award the longer they play, I ensured that the measurement of awards was independent of the number of years played by dividing each

number of awards by the number of years that player has played.[5] Lastly, I accounted for the number of years each player has played professionally in the MLB to see if longevity affects salary, though I am not counting it as a variable of interest.

4.2 Salary Data

Unfortunately, salary was not easily available from Baseball Reference, so I took salary data from a database from USA Today. In order to make the coefficients on my regression model more visually appealing, I divided each player's salary by one thousand, representing salary in thousands of dollars.

5 Analysis

5.1 Hypothesis Testing

The null hypothesis of this regression model is that the coefficients on each of the five variables, batting average, OPS, number of all-star awards, number of MVP awards, and years in the league are equal to zero. The alternate hypothesis is that these coefficients are not equal to zero. In Table 1, we see that the estimated coefficients on all five variables are nonzero, indicating that I can reject the null hypothesis in favor of the alternate hypothesis.

5.2 Methods

In order to maintain consistent results and clean data, I multiplied batting average and OPS by 1000 to get whole numbers instead of decimals. For example, a 0.200 batting average would become 200, consistent with the common verbal description "200 batting average." The same applies for OPS. In order to make the number of all-star and MVP awards consistent and not correlated with years, I divided each of allstar and mvp by years to get a proportion of awards won for each player compared to the player with the greatest number of awards (values

[5] See Results for an explanation of omitted variables

would take integer values between 0 and 1). I then multiplied both variables by 1000 to get consistent values with batting average and OPS. Next, I multiplied the years variable by 100 to get consistent values with batting average and OPS, similarly to what I did with allstar and mvp. Finally, I divided the salary variable by 1000 for simplicity of coefficients so that base 10 expansions would not be that large. In the new data, salary is valued at thousands of dollars.

First, the R-squared value of 0.637 indicates that our estimated regression model somewhat explains the trend in salary, fairly strongly. Essentially, about 64% of the variability in the data can be explained by the estimated model.

The first variable of interest, batting average, has a contradicting sign to what one would expect. The coefficient on batting average is negative, indicating that a higher batting average means a smaller salary. Though this variable is statistically significant at the 10 percent level, it's not economically significant in this case since it would not make sense for a team to offer a lower salary for a player who has a higher batting average, and vice versa.

The second variable of interest, OPS, has the right sign one would expect, that is, a higher OPS means a higher salary. This means OPS is economically significant. Also, we can see by Table 1 that OPS is statistically significant at the 1% level and has a small standard error compared to the value of the coefficient on OPS. Therefore, we can conclude that there is a positive correlation between OPS and salary.

The next variable of interest, allstar, has the right sign, being positive, and indicates that the more all-star awards won, holding years played constant, the more the player will be paid. Generally speaking, it's common knowledge that players who have won all-star awards are the "top" player in the league, and thus it would be assumed that they are on the higher-paid end. We see that allstar is statistically significant at the 1% level, and therefore there is a positive

correlation between the number of all-star awards won and salary, while keeping years played in the league constant.

The last variable of interest, mvp, also has the right sign, positive, and shows that the more MVP awards won, the higher the salary, while being economically significant. The coefficient of mvp is the largest, as expected, meaning that players with MVP awards have a way higher salary than other players. Similarly to allstar, MVP is accounted for while holding years played in the MLB constant. A conclusion that there is positive correlation between the number of MVP awards won and salary can be made since mvp is statistically significant at the 1% level.

I can also consider years, not a variable of interest. Years is both economically and statistically significant at the 1% level, and means that the longer that a player is in the league, the higher salary, which makes sense as players negotiate their salary from year to year. Also, players who do not perform well usually do not stay in the league for that long, so a player who has been in the MLB for longer usually means that they are statistically good.

5.3 Results

Using multiple performance based statistics and awards, I found that both award types had positive impacts on player salary, in ways that were similar to general assumptions that the more awards a player has won means they had mostly likely performed better than others, and therefore are more likely to have higher salaries. When it came to performance based statistics, a higher OPS was an indicator of a higher salary, while a higher batting average meant a lower salary. The batting average component not economically significant could have been due to a few reasons. First, it can be said that some players regress over their time in the MLB, due to age, injuries, among other factors. Next, players could have had a bad season in the 2023 year, which was analyzed. There are endless reasons as to why batting average was not as strong of a

variable as others since a player's skill of only being able to get a hit is not the only factor of their salary.

Focusing on the four variables of interest, it is certain that there is omitted variable bias included. By selecting only batting average and OPS, I am already leaving out other statistics such as slugging percentage, number of home runs, wins above replacement, etc. A key variable in our model, not considered a variable of interest for this paper, has a tremendous impact on a player's salary, and thus is considered an omitted variable if just considering the variables of interest.

6 Conclusion

Analyzing the model at the 1% confidence level, I can conclude that a player's OPS, the number of all-star awards, and the number of MVP awards are positively correlated with their salary. I found that for every one point increase in a player's OPS, their salary increases by about fourteen thousand dollars. For every all-star award a player receives, holding the number of years played constant across all players, their salary increases by about seventeen thousand dollars. And for every MVP award a player receives, again holding the number of years played constant across all players, their salary increases by about forty seven thousand dollars.

These findings are mostly consistent, with the exception of batting average, with the general consensus of factors of salary. Connecting to the literature surrounding the baseball salary realm, our findings can be used to dive into salary dispersion/disparity at a team or player level. Since there are not many economic pieces attempting to analyze baseball performance versus salary at the individual level, this analysis provides insight into the economic implications of professional baseball players' careers. Our analysis can be extended to the determination of salaries, for agents and players alike, in order to find the most elements of a player's salary.

8. Appendix

Figures

VARIABLES	(1) salary1
ba1	-25.64*
	(14.43)
ops1	14.24***
	(4.818)
allstar1	17.01***
	(2.069)
mvp1	46.85***
	(10.57)
years1	10.10***
	(0.889)
Constant	-5,805*
	(3,003)
Observations	311
R-squared	0.637

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Regression Output
Note: standard errors in parentheses
Note: ops1, allstar1, mvp1, years1 are statistically significant at the one percent level, while ba1 and years1 are statistically significant at the ten percent level.

VARIABLES	MEAN	SD	MIN	MAX
salary1	6,453	8,494	154.8	40,000
ba1	251.3	28.55	173.0	354.0
ops1	743.8	90.87	530.0	1066.0
allstar1	110.0	182.0	0	1000.0
mvp1	6.53	32.1	0	333.33
years1	588.1	362.9	100.0	2000.0

Table 2: Summary Statistics

Note: salary is measured in thousands of dollars, ba is multiplied by 1000, ops is multiplied by 1000, allstar is represented as a proportion by years multiplied by 1000, mvp is represented as a proportion by years multiplied by 1000, years is multiplied by 100

Figure 1: Average Salary by Number of All-Star Awards

Note: salary is in thousands of dollars

Figure 2: Average Salary by Number of MVP Awards

Note: salary is in thousands of dollars

Figure 3: Average Salary by Number of Years
Note: salary is in thousands of dollars

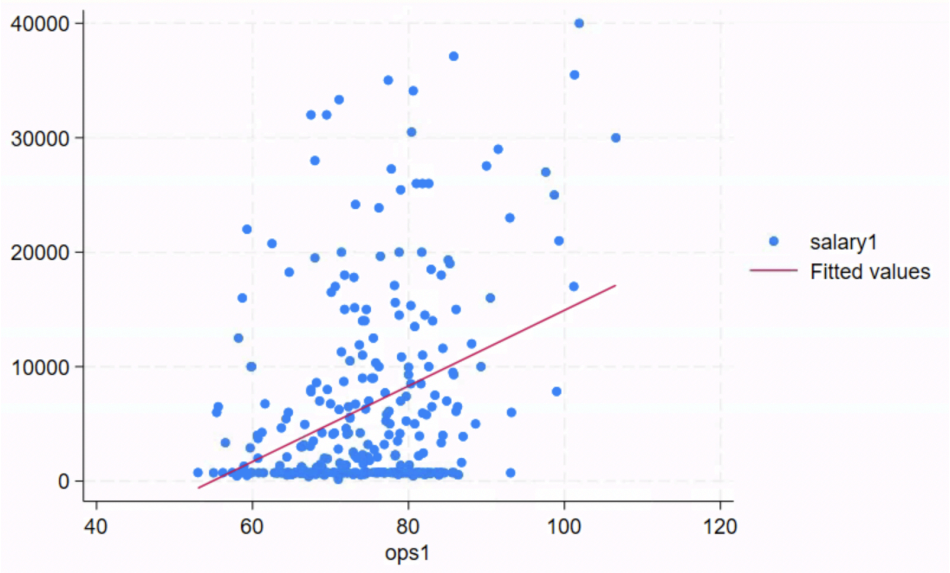

Figure 4: Scatterplot of Salary by OPS with Fitted Line
Note: salary is in thousands of dollars
Note: visual positive correlation between OPS and salary

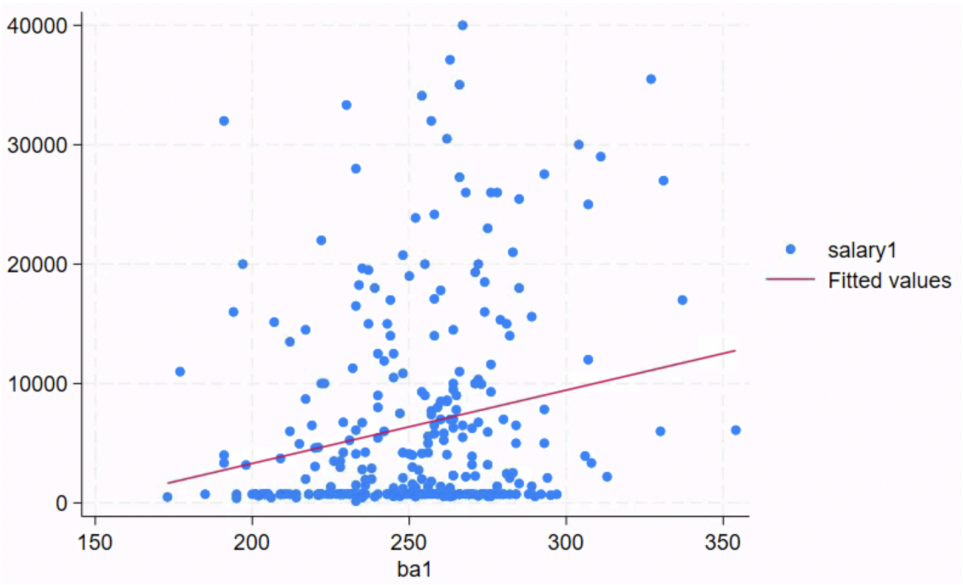

Figure 5: Scatterplot of Salary by Batting Average with Fitted Line
Note: salary is in thousands of dollars
Note: visual positive correlation between BA and salary

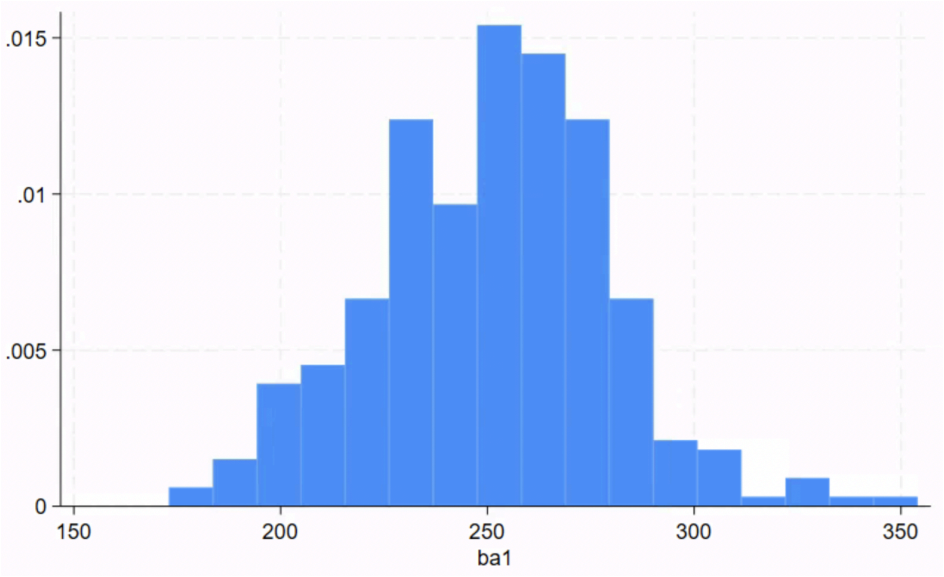

Figure 6: Distribution of Batting Average
Note: batting average = ba*1000

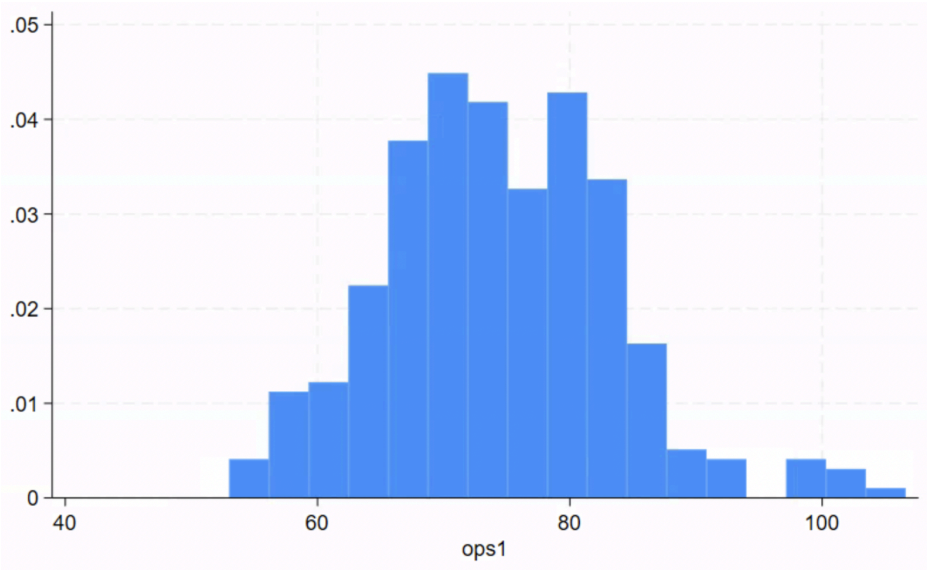

Figure 7: Distribution of OPS
Note: OPS = ops*1000

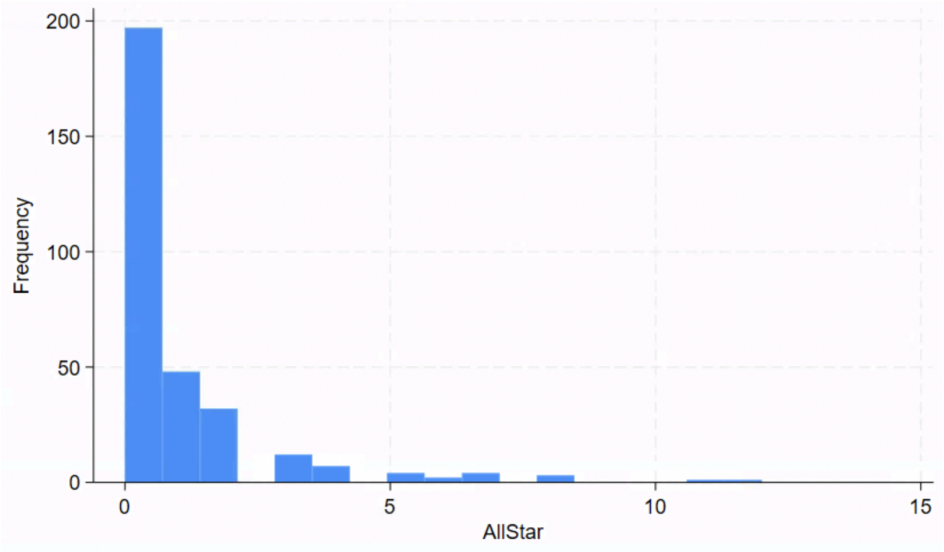

Figure 8: Frequency of All Star Awards

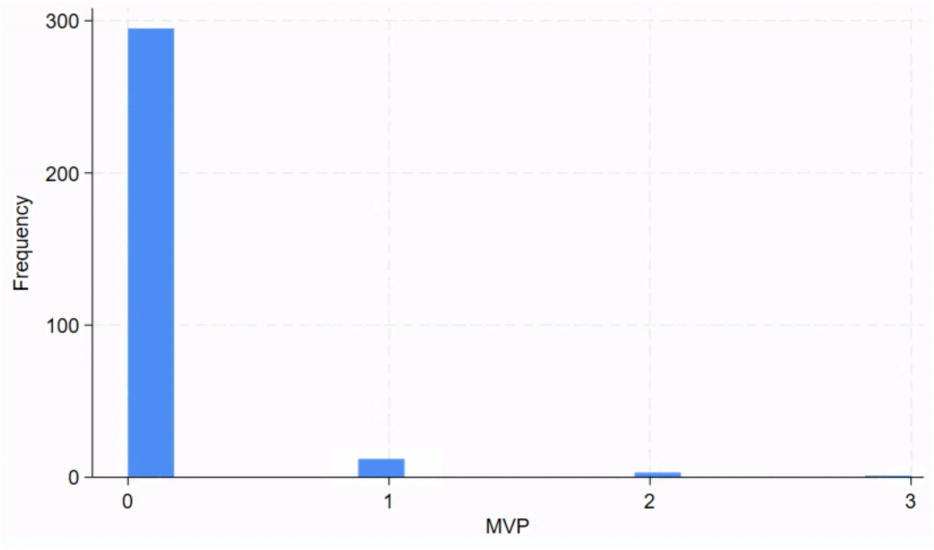

Figure 9: Frequency of MVP Awards

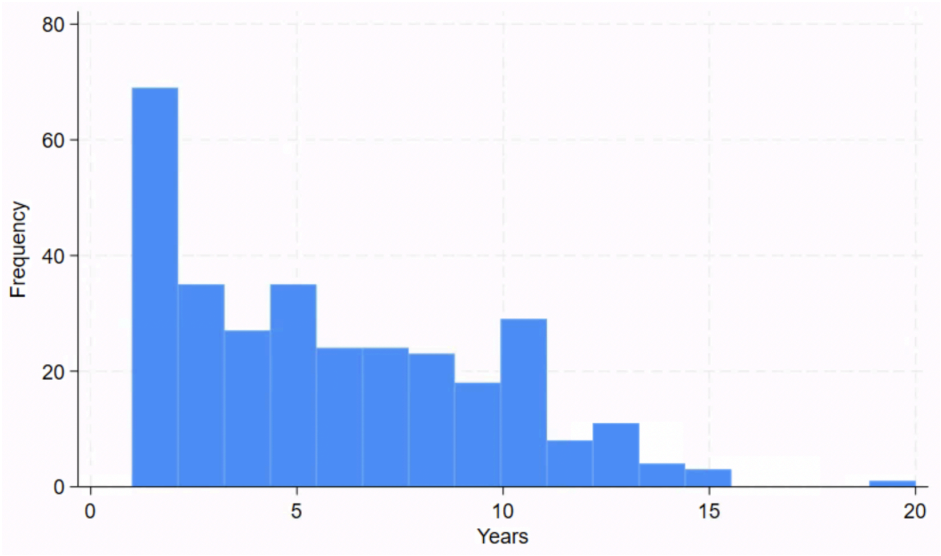

Figure 10: Frequency of Years of Experience in the MLB

Stata code:

```
import excel "C:\Users\mhsu01\Desktop\allstatsonly (2).xls", sheet("Worksheet") firstrow
case(lower)
generate ba1 = ba*1000
generate ops1 = ops*1000
generate salary1 = salary/1000
generate allstar1 = (allstar/years)*1000
generate mvp1 = (mvp/years)*1000
generate years1 = years*100
ssc install outreg2
```

```
reg salary1 ba1 ops1 allstar1 mvp1 years1
outreg2 using "rr.tex", replace
summarize salary1 ba1 ops1 allstar1 mvp1 years1
outreg2 using "ss.tex", stats(N mean sd min max) replace
histogram ba1
histogram ops1
hist allstar, freq
hist mvp, freq
hist years, freq
graph bar (mean) salary1, by(allstar)
graph bar (mean) salary1, by(mvp)
graph bar (mean) salary1, by(years)
twoway (scatter salary1 ba1) (lfit salary1 ba1)
twoway (scatter salary1 ops1) (lfit salary1 ops1)
```

References

Depken, Craig A. 2000. Wage disparity and team productivity: Evidence from major league baseball. *Economics Letters*, *67*(1), 87-92. https://doi.org/10.1016/S0165-1765(99)00249-9.

Tao, Y., Chuang, H., & Lin, E. S. 2016. Compensation and performance in Major League Baseball: Evidence from salary dispersion and team performance. *International Review of Economics & Finance*, *43*, 151-159. https://doi.org/10.1016/j.iref.2015.10.037.

Watnick, Mitchell R. 1998. Pay for Play: Are Baseball Salaries Based on Performance? *Journal of Statistics Education.* **6**(2), 1. https://jse.amstat.org/v6n2/datasets.watnik.html.