Are MLB Players Worth Their Salary?

Augustus Madsen

DSC 530

11/15/19

Hypothesis

For this project, my hypothesis was a two part question: “If Major League Baseball players are paid more for performing better, then players with a higher salary should have a higher WAR.” The statement boils down to the idea that higher paid athletes would generally have a higher wins above replacement, but is this always the case? Secondly, “If a player’s salary is contingent upon performance, then he should only be paid for the value he contributes.” This is where it gets a little controversial. There are players making $30 million a year, but they do not have the highest WAR value. Additionally, for my investigation, the wins they *are* contributing to their team might not be worth that high of a pay scale.

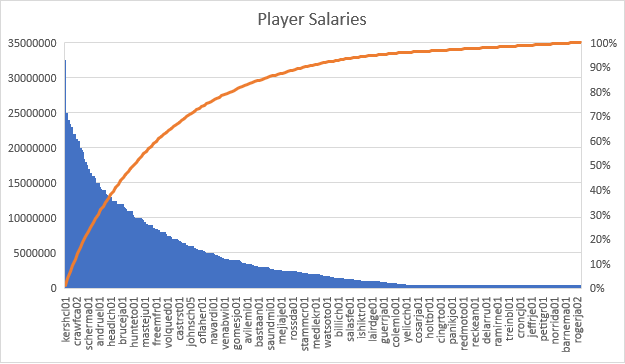
EDA

To begin my EDA, I used the Salary dataset from Kaggle and decided to remove all players from 1935-2014 (This seems like a lot, because it is), but I just wanted to focus on one year. I kept the 2015 data including the player names, team, league, and salary for the season. I had over 800 players to use for the 2015 season, which was more than enough for my investigation. First, I identified the types of variables. They were all strings, except the salary and WAR which were numeric. For WAR I used baseballreference.com to isolate individual performers and find their overall (offensive and defensive) outcome for later analysis. Then I looked at some basic metrics through Excel:

-Mean 2015 salary was roughly $4,000,000

-Median salary was $1.8 million

-Average WAR value around 2.0 (or 2 wins above a replacement level player at the position), later found to be 1.38



I made a quick graph mapping the 2015 players per team, and the left hand side shows the highest paid player, Clayton Kershaw, making over $30 million pear year, while CJ Cron was most likely making the league minimum around $400K.

Missed During Analysis

I think I missed out on doing some more advanced regression techniques. I follow a site called FanGraphs where the community contributors are constantly using advanced regression methods to back up their work. Typically this is applied in players’ worth for the team specifically, and how their efforts are helping within the nine innings. I could have also taken advantage of some more historical data to support my question – how the dollar has changed value over time, the measurements before WAR was established officially, and team net worth.

Variables for Analysis

There are definitely some other variables that would have helped in this analysis. For one, I could have expanded by using data from team earnings: ticket sales, vendors (food/beverage and apparel), etc. This would help me find a more detailed comparison between the amount of value a player is contributing and the team’s overall revenue. This way, I could more specifically see if what a team is paying a player is actually worth it. It’s one thing to estimate the value of one MLB victory with only salary/WAR, but I think it would become more accurate with some financial data to support the findings.

Incorrect Assumptions

At first, I assumed that a lot of the rookie and lower level contracts were overvalued in the sense that most players just coming up do not see a lot of playing time, and therefore their WAR suffers because they are not contributing regardless to have any effect. This assumption was incorrect because I cannot penalize players for not having the chance to contribute. They have no control over when they are put into the game. Additionally, several outliers were these rookie contracts where players were exceeding expectations. This is fairly rare in the Major Leagues, but it is important enough to recognize that some players are making *less* than what they deserve.

Challenges

One of my main challenges was gathering the WAR value itself. Baseball Reference does an impressive job cataloging these records, but I was not able to find an exact dataset which had the value with the salary included. I could have taken the time to fill in a new column in my spreadsheet to assign the WAR, but this would have taken a long time. This is why I focused on a few key players on more of the advanced analyses so that it would not be too much of a burden. Also, it was challenging (but fun and interesting) to figure out the calculations behind the wins above replacement stat. A lot of these modern stats were created my brilliant minds, and that shows in how the math is applied. I essentially had to break down the formula to see what pieces were assigning the value accurately to compare against the dollars per win.