Nikhil Gopal

2-26-18

Schiel Sports Stats C

Is MLB Batter Salary Correlated with Runs Scored?

**Intro:**

In baseball, a winner is determined by the overall number of runs scored throughout the match. The batting team hits the baseball, and the fielding team tries to prevent them from scoring. If the batter hits the ball within the field, he gets the opportunity to run “around the bases” and can score 1 run if he gets back to home base. If the batter hits the ball outside the field, he scores a “home run” and any batter currently on base gets to go home. The fielding team tries to prevent the batting team from scoring runs by earning “outs” for catching the baseball before it hits the ground or touching the batters with the ball or for striking out a batter, and a few other things. By earning three outs, the fielding team gets the opportunity to bat, and each team gets nine opportunities to bat before the game ends.

Since runs are the way score is calculated in baseball, I decided to correlate them with salary in this analysis. It is true that some players are “benched” in that they don’t get as much time to play and therefore score runs, but the goal of this project was to see if salary correlated to runs scored for the MLB, including benched players. The goal of this analysis is to see if a player’s “ability” as demonstrated by their runs scored is correlated with their salary.  
**Hypothesis:**

There is a strong, positive correlation between the number of runs a batter scores, and the salary he earns in the MLB.

**Methods:**

The data were obtained from Sean Lahman’s baseball data archive. This data set includes many baseball statistics from the 1800s to 2016. The salary data was only available for 1985 to 2016, so this project focused on those years. To my knowledge, the data contains information for the whole MLB.

The number of runs scored, and the salary data were originally put in different files. To analyze them in Excel, I needed them to be in the same file. I also needed to avoid including players from years outside 1985-2016 where I didn’t have salary information, so I wrote a python script to read both files and copy and paste the data into one file that I could analyze in Excel. I also programmed the script to write a column of inflation adjusted salaries into the excel graph. All graphs and trendlines were made in Excel. A full copy of all my code and all the files I used in this project can be found here: <https://github.com/ng4567/sportsstats>. The CPI values used to calculate inflation were obtained from the US Department of Labor Bureau of Labor Statistic: <http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/>.

**Results:**

See graphs below:

Both the inflation-adjusted and raw salary graphs have low R scored values (0.0689 and 0.0541) respectively.

**Conclusion:**

Based upon the data, there is almost no correlation between number of runs a batter scores and their salary. The variance that can be explained by the regression line model is approximately 0.05% for the unadjusted data and approximately 0.07% for the inflation adjusted data. Since such a low proportion of the variance can be accounted for by the model, I can’t convincingly say that there is a strong correlation between MLB player salary and Runs Scored.

asdfa