

Sabermetrics Final Project Proposal

Austin L. Pearman

Often in this class we have spent time talking about measuring a player's contribution in terms of runs that the player in question is contributing to his team. We have done this in many ways. Some were inefficient such as counting RBIs, while others were better such as calculating how many runs an at bat was worth by run expectancy. One thing that I don't think we have touched on enough is the ability of a player to create a productive out, meaning an out that resulted in a run or advanced base runner. I will be calling this statistic $PrO\%$, or Productive Out Percentage. I believe this statistic would be an accurate metric to measure a player's contribution to the "small ball" part of the game.

Given that the data I need will be in the form of play by play information, I will be using the same Retrosheet database that we have been using in the class. I will be pulling all plays in which the batter was up to bat with runners on second or third and less than two outs. I will consider this to be the plays in which a player has an opportunity to create a productive out and will call the total number of these plays per player $PrOO$ or Productive Out Opportunities. I will then count the number of $PrOO$ s in which the batter was ruled out but a run scored or a runner advanced. This count will be called Productive Outs or $PrOs$. Thus, $PrO\%$ will be calculated as indicated below:

$$PrO\% = PrO/PrOO$$

To expand upon this idea, I am considering evolving the statistic to weight productive outs that result in runs heavier than those that merely advanced the base runner. I would also put emphasis on outs in which runners advanced and runs scored. I do not know what weights I would assign and it would most likely take a fair amount of tinkering to get right. For that reason I consider this to be a stretch goal for the project. I am considering using the ratio of run value for the two different kinds of at bats to weight them but that is just my preliminary idea.

To demonstrate this I am choosing to go with a simple approach. I have not touched web development for three years and will not have the time to refresh myself alongside finishing senior projects. For that reason, I am planning on creating a python application that will exist in the same directory as a csv file that will contain all of the necessary data. The application will prompt the user for the name of a player and will then fetch the PrO , $PrOO$, and $PrO\%$ for that player. The application will also display the percent of $PrOs$ that resulted in runs, the number of that resulted in an advanced runner, and the number that resulted in both. This will be comprehensive for all players with data in the retrosheet regardless of position. To run this, it will be required that the user downloads the python script as well as the csv file and that they are stored within the same directory and the csv file's name is not changed. This will be detailed in the README.

