Statistic Analysis

Austin Pearman
CSCI 4831-001

When I began evaluating this statistic I was expecting most players to be returming a very high success rate in creating productive outs. I believe this assumption came from my own personal frustrations as a baseball fan. Watching players pop out with a man on third and one out and failing to create that crucial run is one of the greatest frustrations of any fan. It was much to my surprise that most players were creating productive outs at a relatively low rate. When looking at players with a minimum of 500 PrOOs (Productive Out Opportunities), Elvis Andrus topped the PrO% (Productive Out Percentage) ranking with 37.8% of his PrOOs resulting in productive at-bats. I was expecting good hitters to be able to create a productive out in about half of their opportunities. After all, all they really need to do is make contact right? Evidently not.

Looking at the statistics, I believe that, at least for most players, the PrO% is being weighed down by the inclusion of the "man on second" scenario. Advancing a man from second is simply not that easy. It most often either requires a decent smack to right field (most often by a left handed batter) or a well hit but slow rolling ground ball to the right side. Given how common this base state is, it is easy to see what may be dragging down the statistics.

In reality, good PrO% numbers did not look much different than a good batting average and only a little bit lower than a good OBP. The average PrO% for players with at least 500 PrOO is 0.214. This is approximately 40 points below the average league batting average in 2017. Given a batting average Mendoza Line of 0.200 and assuming a proportional relationship to PrO%, this would put a hypothetical PrO% Mendoza Line at approximately 0.160. I would like to note that while I believe in the validity of a productive out metric, I do not believe that the scope of this statistic is large enough to make a strong argument to replace any player by this metric alone. By the logic of the above numbersm I've created the following rough player assessment based on PrO%.

	Minimum	Maximum

Good	0.260	N/A
Average	0.160	0.260
Replacement	0	0.160

It is worth noting that the PrO% tended to unsurprisingly favor contact, "slap", hitters over power hitters. Famed slap hitter Ichiro Suzuki, turned in a phenomenal PrO% of 0.322. A full 79 points above the league average. Meanwhile, the home run hitting machine that is Giancarlo Stanton, turned in a replacement level PrO% of 0.152. I also found below average numbers for Carlos Gonzalez and Edwin Encarnacion while finding great numbers for Joe Mauer and Marco Scutaro. Does this mean that I believe that the Yankees should be shipping Giancarlo Stanton to the minors immediately? Absolutely not, he is one of the best players in baseball. That being said, if you need a sacrifice fly to win the game and have Ichiro on the bench. Giancarlo may not be your guy.

I created a correlation matrix (available in the jupyter notebook) to compare the PrO% statistic to other stats that may be useful in predicting a player's ability to create a productive out. One thing that I was certain I would see would be a strong negative correlation between PrO% and strikeouts. It seemed intuitive that a player's ability to create a productive out would be highly influenced by his ability to put the bat on the ball. Surprisingly, while there was a notable negative correlation coefficient when comparing PrO% to Strike Outs, the R-squared value came out to a mere 0.035. In fact, PrO% seemed to be poorly correlated with the vast majority of other metrics. It's strongest correlation was with RBI% (the percentage of PrO that resulted in RBIs) and even that was only 0.13.

This could draw me to one of two conclusions. First, I could conclude that the loose correlation between PrO% and other commonly accepted metrics of performance is indicative of the poor quality of a productive out metric. Second, I may conclude that this metric is simply along a different line of thinking and player evaluation metrics than those already in place. I do firmly believe that measuring a players ability to be productive, even in an out, is a valuable metric, but I acknowledge that there may be shortcomings in my calculation. I believe that future work on this subject should focus on assigning weights to different types of productive outs. If sabermetrics is about measuring a players ability to create runs and therefore wins for his team, a metric that

evaluates the ability for a player to do that, even while creating an out, must hold some value.