Stefano DelPiccolo

4-30-2019

CSCI 4831-001

Hoenigman

Contents

TBB is my proposed new statistic that is responsible for counting the total number of bases a player is responsible for across the whole season.From TBB you could easily calculate my other proposed statistic Bases Per At Bat (BPAB) Which as suggested would be TBB/AB. This statistic would be a true team player stat as getting and moving people on bases is the most important part of baseball. Currently if a batter is consistently moving people to let's say third bases but not batting them in there is no way to evaluate that. Only the player who bats the person in would get credit, but the player that batted them from first to third bases deserves just as much credit if not more than the person that batted them from just third to home. There is no true statistic that takes in only bases moved. So I created one.

Theses statistics would be most closely related to Runs Batted In (RBI), in which essinactly count the movement of other players on the bases. The main difference being that I count every bases and not just the bases the scores the run. A correlation test could be dividing TBB by 4 to see how closely these stats compare. Since I am also counting bases that the batter himself gets on it could also be highly related to On Base Percentage (OBP). If a batter has a high OBP then his BPAB would also be higher because they are directly correlated.

The way this statistic was calculated is by using the statcast 2018 season data. The first thing I did was drop all pitches in which no event happens there should be no bases moved so all of these were dropped. I then look at every at bat and create 4 new columns. The first 3 columns are the base state after the at bat. This is done by look at the base state of the next at bat and this will be the change that the previous at bat caused. The last column would be the change in the score. You have to look at the change of score to account for the runners who scored. These runners would no longer be on base and need to still be accounted for in the total bases moved. Each at bat will then go through various if else statements to determine the situation that happened. For example the first if statement is a check to see if anyone is on base. This is the case where it is easiest to count the bases from this at bat. Its either 1,2,3, or 4 for a home run. There is only more and more cases from here on out. The rest of the cases are dependant on how many runs were scored during the at bat. This is also where the statistic cannot be 100% accurate. Looking at a case where two runs or score. There is a variety of ways this can happen. The list is as follows. Third base and second base can score, Third base and first base can score,3 base and the batter can score, Second base and first base can score, second base and batter can score, and finally first base and the batter can score. I had to make certain assumptions because using the statcast data there is no for sure way to see who actually scored the run, you can only see that runs were scored. For example if bases are loaded and 2 people score, I have to assume that third and second scored. There might be very few cases where in this situation somehow third and first could score and second could get out, but due to to the limitations of statcast I would have no idea this happened. I made the best out of what I had. I store all the data in a dictionary with the key being the batter ID this way after each bat .The comments of every situation I could think of are in the code If I missed a situation I would love feedback on how I could improve the code.

The way I calculate bases could mean that even if 2 runs are scored in two different situations the amount of bases could be vastly different. For example if two people score the bases would be very different if third and second scored vs batter and first bases score. In the first situation is responsible 3 bases (also depends on were first base and the batter got to) vs the other situation is responsible for 7 bases. In opinion this is a fair way to evaluate bases. It is a much harder task to bat in people from hitting and 1st base you would almost certainly have to hit a homerun could cause this event. In the other situation all it would take is a long pop fly to achieve this result. That's why I think it is fair to be worth double the bases. It is more worth it to hit in runners farther from home. If done in one at bat there is less risk of getting out then having to move them over multiple at bats therefore the extra bases is justified.