Batting Adaptability

Jake Liebow & Haris Khan
Sabermetrics

Batting Adaptability is the statistically calculated variance of the hits per pitches of each pitching type against a certain batter. This statistic can be used to pinpoint certain weaknesses in a specific batter! At the highest levels of play understanding weaknesses is essential to self-improvement and more than that, understanding your opponents weaknesses can be quintessential to defeating them. This is especially important given how difficult it can be to get a leg up on the competition in a professional sport. Everyone in the MLB is more or less an expert on Baseball, Baseball Strategy, as well as the nuances of hitting and pitching. It is because of this expert status that it can be incredibly difficult to even know what specifically to work on as a professional player, how do you improve an expert? As I stated earlier they must pinpoint their weaknesses or maybe even learn the weaknesses of their opponents and this statistic provides a simple analysis to point them in the right direction for either eventuality.

This statistic is fairly easy to calculate and provides great insight into the weaknesses of a Batter. This statistic is easily calculated with the statcast dataset, specifically in building this project and statistic we used all of the 2018 season data. We then pick a specific batter to calculate Batting Adaptability for and analyze the description column for specific events. The events we are looking for are any fouls, strikes, or hits, these are the only relevant events in the pitching description column. For example, we do not include balls because the batter didn't try to hit it. For this part of the calculations we are trying to build a dictionary of hit percentages per pitching type. So we can iterate through all pitches and for any fouls/strikes add one to the denominator and for any hits add one to the numerator and denominator. After we have analyzed every pitch a batter received in the 2018 season we can divide the numerator and denominator of each pitching type to get hit percentages for each pitching type. After this all we have to do is find the variance of each group and we get Batting Adaptability. This one simple number will give you amazing insight into the Adaptability and consistency of a batter. For example the top 5 batters of 2018 as listed by espn are Mookie Betts, J.D. Martinez, Christian Yelich, Jose Altuve, and Mike Trout. Their Batting Adaptability scores fluctuate between .003 and .015 which is a very low variance given the percentage nature of these scores. These batters are incredibly consistent and know how to deal with every kind of pitch! Now if we look back at the opposite end of the spectrum we can see batter Marwin Gonzales, 102nd as ranked by espn in 2018. Marwin Gonzales has a Batting Adaptability of .059 which is almost 4 times more varied than the Jose Altuve's Batting Adaptability of 0.015. So, we have a great way to pick out batters with inconsistent or not very adaptable batting style but how can we improve based on this data? How and where can we find the pitching type we are weak against?

If we were to examine a specific batter, their Batting Adaptability, and the data used to calculate it we could learn a lot. Unfortunately there isn't much data between every pitching type some pitches are rarely thrown and others are thrown very frequently. So it is difficult to pick out the best instrument for discovering insight on our batter. However, we recently discussed a technique in class called bootstrapping which was built to compare two populations with sample sizes too small for other tests to make sense. This test also fit like a glove, we can iterate through each pitching type and test the null hypothesis that within N trials of randomly selecting X(the number of pitches pitched with a specific pitching type) simulated pitches we will see that a specific pitching type is statistically greater in terms of hit percentage than the rest of the population of hit percentages of pitching types. We can then find a p-value as the probability that our Null hypothesis is true.

For example, when analyzing Marwin Gonzales we found a Batting Adaptability of 0.059 suggesting that his ability to hit the ball is widely varied by pitch type, it also means that the population of pitching types has a standard deviation of about 25% of its total possible values as the only possible values range from 0-1. After Bootstrapping each pitching type 3 pitching types were found to have rejected the Null Hypothesis with a significance level of 0.05. Curve balls with a p-value of 0, Slider balls with a p-value of 0, and Knuckle Curve balls with a p-value of 0.021. For Curve Balls this means that in 1000 simulations there were 0 trials where Curve Balls surpassed the hit percentage of the general population. So to improve Marwin must practice batting against Curve Balls, Knuckle Curve Balls, and Slider Balls. On the opposite end if you want to absolutely destroy him while pitching, throw only Curve Balls, his hit percentage is 19 out of 156 Curve Ball pitches. This statistic even has merits for the best batter in the world (*in 2018) Mookie Betts after running through the bootstrap simulation was found with a weakness to Slider Balls and Curve Balls with a p-value of 0.025 and 0.008 monitoring p values with a collective significance level of 0.05.

There is one more use case for this statistic, if we iterate through every pitcher that has pitched a sufficient amount (30 pitches in our case) then we can pull out Batting Adaptability for our batter against each pitcher. With this information a professional could optimize their training for their next opponent. For example if the Boston Red Sox are about to go up against the Yankees then I would recommend that BOS Star Batter Mookie Betts trains against Sinkers because he has hit zero out of six sinker pitches from CC Sabathia the Yankees pitcher as such the p-value was 0. Another problematic pitch type was a Fastball Cutter which Mookie hit three out of fourteen times, the p-value was 0.137 which is still noteworthy given the extreme Batting Adaptability score. We found this anomaly after analyzing all pitchers who pitched sufficiently against Mookie Betts and found that against CC Sabathia Mookie Betts Batting Adaptability rose from a consistent 0.006 to a terrible 0.19. We discovered this with the help of Batting Adaptability as it clearly explained who Mookie could bat consistently against and who he could not. This allowed us to pinpoint these weaknesses so either Mookie can fix them, we can predict their exploitation, or so that other pitcher's and teams can also exploit this weakness.

Overall this statistic most closely relates to Batting Average because it uses calculations similar to Batting Average in order to accurately compare and find the variance of all of the groups, groups being pitching types. However, it is still difficult to compare this statistic to Batting Average because a pitcher can throw multiple pitch types in a single plate appearance. This Statistic also relates to On Base Percentage, as in an adaptable player with a consistent handle on most pitching styles will be likely to get on base. However, comparing On Base Percentage with Batting Adaptability is still tricky and not quite reasonable given that a player can have good batting adaptability but low skill in the game. Someone who consistently hits 1/10 pitches regardless of type will have amazing Batting Adaptability and yet will still have low Batting Average and On Base Percentage. I find this odd given how much data is available, how easy it is to access, and how useful this statistic seems.