

Measuring and Modeling Batted Ball Quality

WSU Mathematics and Statistics department

Discussion

After 5 different modeling types we concluded that the pitching variables within the Statcast data are not very predictive of hit speed. With that being said, we can still look at the lasso coefficient plots for each player and gain an understanding into what types of pitches lead to a lower or higher predicted hit speed, albeit with an R^2 of less than 0.35.

This problem was a tough prediction problem. Hit speed may vary drastically by the bat being centimeter's lower or higher than normally expected. With these factors into consideration, it lead us to suggest possible extensions of this work.

The Statcast data source is very rich and can answer many questions. With that in mind, one potential extension of this work would be to attempt to answer a different questions. If we recall Figure 3, one could use the first loading score as a response variable. This would be nice for modeling what types of pitches turn into home runs. A potential research question for this problem would be, what pitching characteristics make up a home run?

Another extension of this work may be to fit more complex models. This research was only done using linear model selection and regularization. If one were to pick up this data set and continue to predict hit speed, fitting more complex models may lead to a higher prediction accuracy. This may lose some interpretability, but if the model predicts hit speed more accurately then it will be an improvement.

A final extension of the work that we've done would be to add model more players and compare prediction accuracies. We arbitrarily chose to model the top ten position players according to the WAR statistic. This choice may have lead us down a path of less predictive accuracy. For example, if we were to model below average hitters, maybe there is more variability in the hit speed. Mike Trout may just be a player that can hit nearly every type of pitch hard. Whereas an average player may struggle to hit a certain type of pitch hard.

With all of this discussion, leads me to one last final point. That last final point is that prediction is hard. Take it from the New York Yankee, baseball legend, Yogi Berra, "It's tough to make predictions, especially about the future."