Measuring and Modeling Batted Ball Quality

WSU Mathematics and Statistics department

Abstract

Baseball is a competitive game of offense, defense, athleticism, strategy, and numbers. The recent advent of the *StatCast* technology has yielded data on the exit velocity, hit angle, coordinates of the ball when it passes the strike zone, the spin rate of the baseball, the break angle, and much more. This data is available at *baseballsavant.mlb.com*. In this project, we investigate the following questions: What is a well hit ball? What are similarities between pitches? What pitching characteristics explain the quality of a hit ball? We used principal component analysis to analyze what a well hit ball is, what pitching variables affect the qualities of well-hit balls. We fit various models to explain exit velocity. We used five modeling techniques: full multiple linear regression, forward variable selection, backward variable selection, ridge regression, and the lasso. Using data for Mike Trout, each model had a low predictive performance. These R-squared value were calculated on the test data using 2/3, 1/3 cross-validation. Once the models were fit, we looked at the lasso coefficients. The top two largest coefficient were locations that corresponded to the upper-in and upper-middle strike zones. A possible implication is a pitcher may not want to pitch Trout on the inside.