SONOMA STATE UNIVERSITY | MATHEMATICS & STATISTICS

# Unlocking Player's Value: Predictive Modeling of Performance

Brandon Keck and Arthur Jones
(Math 465, Strawberry Shortcake), Sonoma State University, Rohnert Park, CA 94928

## Background

For our analysis, we are looking at what is called a player's ROTO value. Imagine that you have just become the general manager of the San Francisco Giants, and have been given $100 dollars to spend on any players you wanted for your team, and all the other managers were in the same situation. Based on these circumstances, the ROTO value given is the value a player earns based on their performance. However, you can't just choose all the best players, as you only have a $100 budget, which means that we are going to have to find some diamonds in the rough.

## Methods

Variables Used:

- ROTO: Known as Rotisserie Chicken for MLB, players are assigned a value based upon player stats.
- Slugging Percentage: Is a calculated value based upon extra base hits.
- At-bats: The number of at-bats a player has.
- Home Runs: The number of Home Runs a player has in a given season.
- Hits: The number of hits a player has in a given season.

## Results

We began by creating a simple linear model. The first of these being the predictor variable SLG vs ROTO. While we did receive a small p-value our adjusted R^2 of only 0.2552, still needed some attention. We began adding predictor variables to our model. Next we included at-bats. This greatly increased our adjust R^2 value to 0.7626. This was a huge improvement of our model. We continued to add terms to the model. We included the predictor variable home runs and again our adjusted R^2 jumped to 0.8227. At this point we knew that we could adjust a very useful model. With our final predictor variable of hits we achieved an adjusted R^2 of 0.9568. We had finally created a useful model in predicting ROTO!

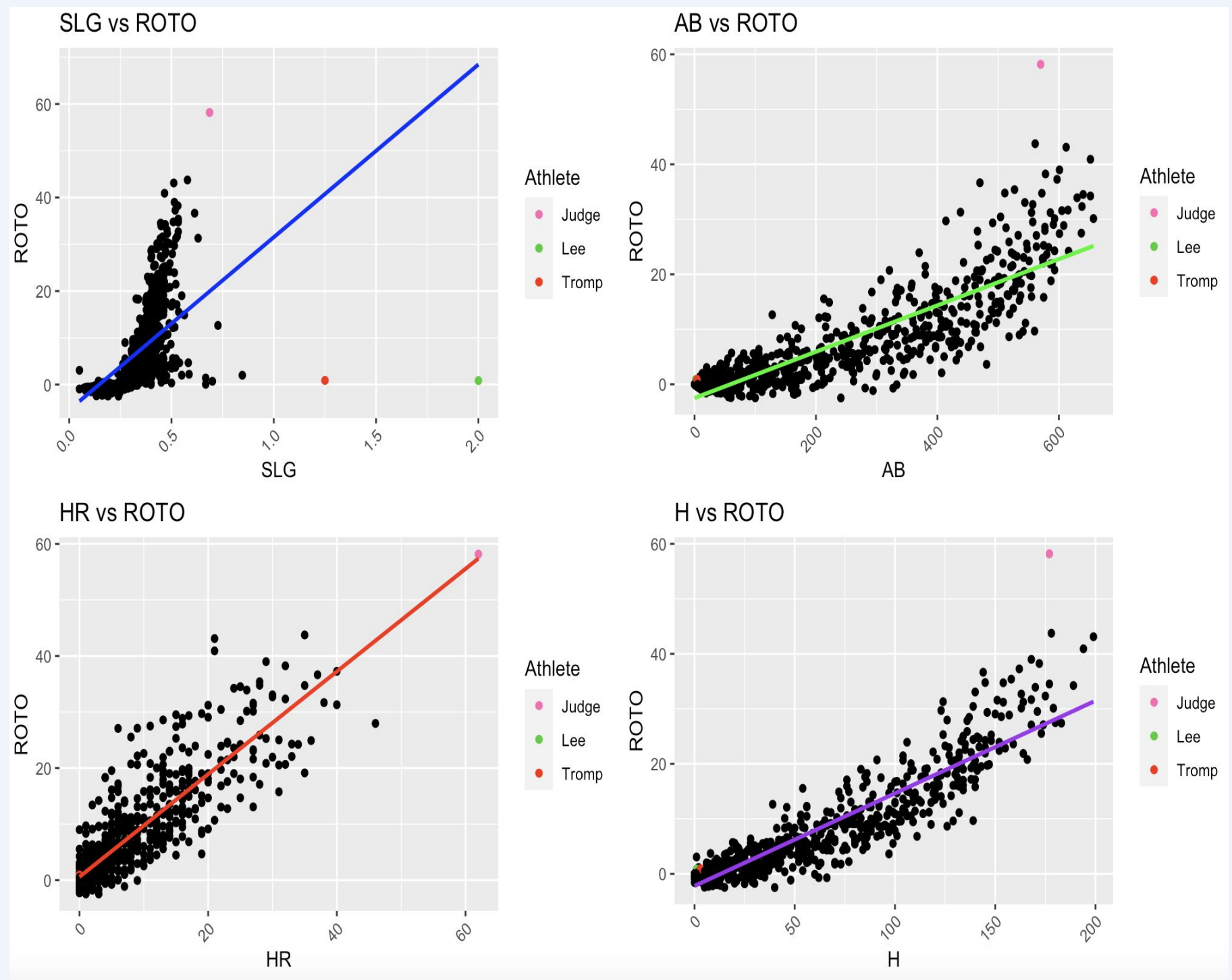

**Fig 1.** These graphs represent each predictor variable against the response variable ROTO.

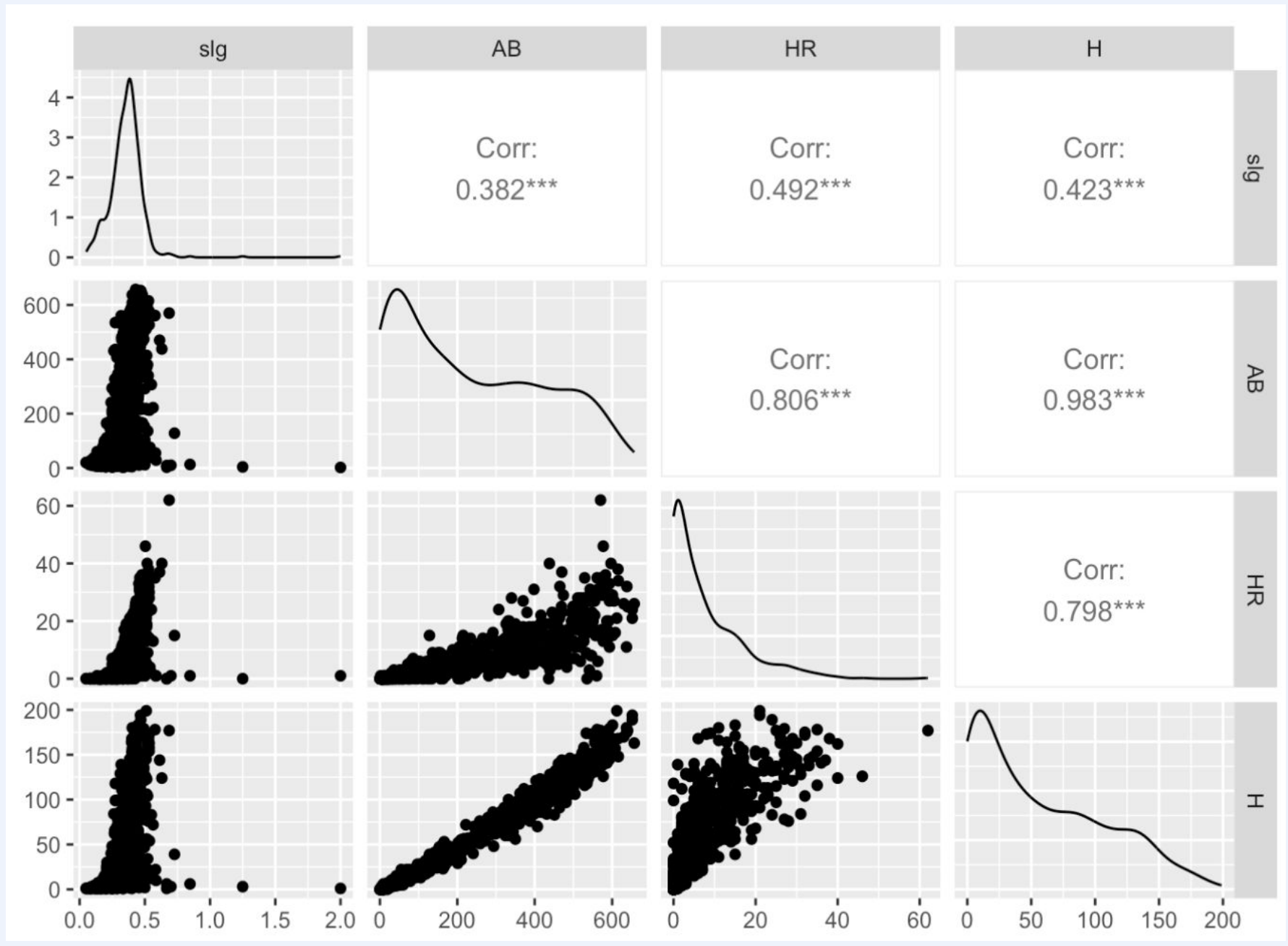

**Fig 2.** This graph represents the relationship between predictor variables.

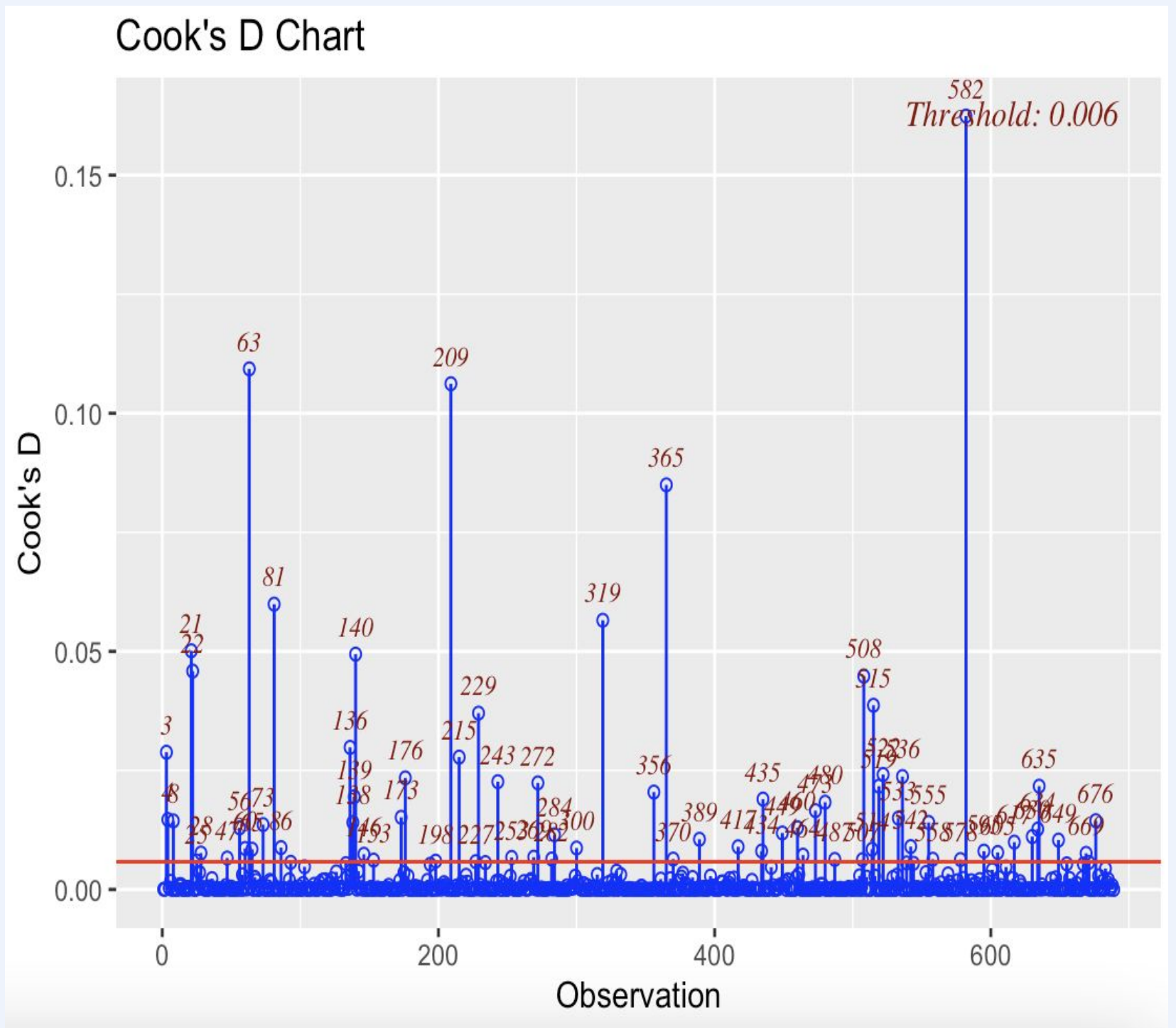

**Fig 3.** This graph plots outlier points of our model that have a significant impact on our model.

## Discussion

After looking at numerous models, our best model looks at the relationship between SB, BB, AB, HR, and H in terms of ROTO. We observed an adjusted R^2 of 0.994. What this means is that 99.4% of the variation in ROTO can be explained by the variables, so our model would be a very good predictor of the ROTO variable.Our residual standard error is low, so we can expect our results to be accurate. We decided to remove SLG from our model and added BB's and SB's. Using our new best model we decided to look at the outliers. Figure 3 shows us a Cook's D graph which helps us identify outliers in our new linear model. There are numerous players that are above our threshold of .006, but we observe that there are a select few players that are much more influential to our model. For example, we see that player 582 is the biggest outlier in our chart. Player 582 is Juan Soto, and he is regarded as one of the best players in the MLB. He is highly influential in our new model because it includes walks (BB), and Soto is always among the league leaders in walks drawn. Therefore, his presence has a huge impact on our model. We concluded that our new model is much better than our previous models because our outliers are now players that are regarded as outliers due to a high level of performance as opposed to players with inflated percentages based on limited sample sizes.

## Acknowledgments

**A special thanks to our professor Dr. Elaine Newman for teaching us new and fun regression techniques!**