

## ASSIGNMENT: PROJECTIONS

We are going to try to improve upon the Marcel projections using batted ball tracking data.

### EXERCISES

1. (1pt) Download the past four years of MLB pitch-by-pitch data using the `download_baseballsavant()` function from the `sabRmetrics` R package. To verify that you have done this correctly, report the number of pitches you downloaded for each year.
2. (1pt) Using the linear weights we calculated in class (and which are available on Canvas), calculate the linear weight per 500 plate appearances for each batter-season. To verify that you have done this correctly, report the top 10 batter-seasons by linear weight, minimum 500 plate appearances. In addition to reporting names, report the number of plate appearance and the linear weight per plate appearance for each batter-season.
3. (1pt) Using the naive 5/4/3/2 weighting scheme (and ignoring aging), calculate the Marcel projection of linear weight for each batter in the latest year using the data from the three years prior. To verify that you have done this correctly, report the projection for (a) the player with the most plate appearances across all four years and (b) the player with the fewest plate appearances across all four years.
4. (2pts) Fit a batted ball outcome modeling using the `ranger()` function from the `ranger` R package. Use the numeric linear weight of each batted ball as the outcome, and include two features: launch speed and launch angle. To verify that you have done this correctly, report the training RMSE.
5. (2pts) Using grid search to choose the optimal weights, create Marcel-like projections for (a)  $xLW2$  per batted ball; (b) the residual  $LW - xLW2$  per batted ball; (c) strikeouts per plate appearance and (d) walks per plate appearance. These projections should predict the latest year using data from the three years prior to the latest year. To verify that you have done this correctly, report the top five and bottom five players in each of these four metrics (name and projection).
6. (2pts) Combine your projections from the previous exercise into a projection of linear weight per plate appearance. Be careful with how you handle the projections for metrics whose denominator is batted balls. To verify that you have done this correctly, report the top five and bottom five players according to this combined projection (name and projection).
7. (1pt) Evaluate the RMSE of your projection from Exercise #6 versus the RMSE of your projection from Exercise #3. To verify that you have done this correctly, report both RMSEs.
8. (extra credit, 1pt) Is this difference in performance between your Exercise #6 projection and your Exercise #3 projection statistically significant? Report 95% confidence intervals for the two RMSEs.

### SUBMISSION

Please submit your solutions to the exercises (the results you are asked to report) in a PDF file. Please submit your R code as well. You may do so by providing the code in line with the solutions in the PDF file (such as a knitted RMarkdown document) or uploading your R script separately.

### REMINDER

Please **anonymize** your submission by removing any personally identifiable information (including file paths in your R script that contain things like a username!).