SIGNIFICANT PREDICTORS OF MLB BATTER PERFORMANCE USING STATCAST METRICS

Kyle Vander Meulen

California Baptist University

Email: [Kyle.VanderMeulen@calbaptist.edu](mailto:Kyle.VanderMeulen@calbaptist.edu)

**1. Introduction**

In ways like never before, Major League Baseball (MLB) teams are embracing analytical tools to further enhance team performance and decision making to maximize odds of winning. The available information that is usable for teams and fans continues to build upon itself over the course of Major League Baseball’s history. Analytics in baseball can be grouped into three different time periods: The Traditional Stats Era (up unto the 2000’s), Moneyball/Sabermetrics Era (2000’s – 2014), and now the emergence of the Statcast Era [11]. Major League Baseball defines Statcast as a, “state-of-the-art tracking system that uses high-resolution cameras and radar equipment to measure the precise location and movement of baseballs and baseball players” [10]. The tracking system allows front offices, broadcasters, and fans to quantify raw skills of players that weren’t available previously. This revolution has provided new outlooks for hitters attempting to gain advantages over their opponents. Hitters have focused specifically on the launch angle of the ball off the bat and the exit velocity of the ball after initial contact [9]. The number one issue is what combination of launch angle and exit velocity should hitters aim for?

The first objective of this paper is the comparison of five predictive models in relation to which is the most accurate representation of a hitter’s performance based upon his exit velocity and launch angle. With the data collected from the 2019 season our model estimates the probability (0 to 1) of hitting a hit for our respective players using the constants and model coefficients and forecasts expected hit percentage that a hitter will achieve based on their statistics over the course of the season. The result shows the intersection of the best launch angle and exit velocity that produces the highest perceived hit probability for our individual hitter will be helpful in further enhancing hit probability and improving poorer hit probability areas. The best fitting model has a mean absolute error of 5.13 percentage points from a hitter’s actual hit outcome.

The second objective of this paper is to examine significant factors in a batter’s exit velocity (BEV). This regression was able to sift out the statistically significant factors for each of our four hitters exit velocity at the 95% confidence level.

Gallo’s significant predictors are as such: Month [-1.57, 0.013], Launch\_Angle [0.05, 0.33], Distance [0.01, 0.07], DirectionPull [6.44, 19.48], DirectionStraightaway [2.98, 17.29], Event\_Field\_Out [-15.23, -0.28], Event\_Force\_Out [-53.2, -2.26].

Judge’s significant predictors are Month [-1.82, 0.025], Distance [0.025. 0.07], Event\_Field\_Out [-14.56, -1.66], DirectionPull [2.15, 11.78], DirectionStraightaway [1.59, 10.13], PitchTypeChangeup [-14.45, -1.21].

Yelich’s significant predictors are Distance [-0.021, 0.04], DirectionPull [4.63, 12.81], DirectionStraightaway [1.70, 9.34], Event\_Field\_Out [-15.00, -3.50], Event\_Force\_Out [-22.00, -2.99], Event\_Grounded\_Into\_Dp [-24.2, -1.05], Event\_Single [-13.00, -0.72].

Ozuna’s significant predictors are Distance [0.023, 0.05], DirectionPull [4.63, 12.12], DirectionStraightaway [-0.18, 7.46], Event\_Field\_Out [-11.45, -0.42], PitchType4-SeamFastball [3.09, 10.90], PitchTypeCutter [-0.01, 12.74], PitchTypeKnuckle-Curve [2.19, 22.50], PitchTypeSinker [3.12, 14.13], PitchTypeSlider [-0.74, 10.74], Pitch(MPH) [0.07, 0.95].

**2.** **Methods and Materials**

**2.1 Data Description**

The data we utilized included full season 2019 MLB Statcast Game Logs for four players: Aaron Judge, Joey Gallo, Marcell Ozuna, and Christian Yelich. The datasets for each of our respective players was obtained from Baseball Savant [4] and were scraped into individual datasets for each player. The analysis was completed in Rstudio 3.6.0 by using these specific packages:

readxl (1.3.1), devtools (2.3.0), lattice (0.20-38), gam (1.16.1), corrplot (0.84), mgcv (1.8-31), tidyverse (1.3.0), leaps (3.1), ggvis (0.4.5), ggpubr (0.3.0) and PerformanceAnalytics (2.0.4)

[17-27]. We intently focused on obtaining an accurate depiction of exit velocity and its potential predictors by including one right handed hitter from the American League, one right handed hitter from the National League, one left handed hitter from the American League, and one left handed hitter from the National League. This was done to ensure an accurate representation of exit velocity league-wide while taking into account batter handedness. Data preparedness was completed through filtering out the non-Batted Ball Events (K’s and BB’s) and subsetting the individual pitches. Additionally, a new variable named outcome was created to estimate and fit the probability of a hit (0 to 1) through the use of our gam and logit models.

**2.2 Variable Definitions**

Predictor variables of month, launch angle, distance, event, direction, pitch speed, and pitch type were used with the following definitions.

**Month** The timeframe by which our batter performed his at bats (Ordinal)

**Launch Angle (degrees)** The vertical angle at which the ball leaves a player’s bat after being struck (Interval)

**Distance** Represents the distance away from home plate that a batted ball lands (Ratio)

**Event** The end result of a player’s at bat (Nominal)

* Single
* Double
* Triple
* Homerun
* Field\_Error
* Field\_Out
* Fielder\_Choice\_Out
* Force\_Out
* Grounded\_Into\_Double\_Play
* Sac\_Fly

**Direction** The particular location of the field where a batted ball is hit (Nominal)

* Straightaway
* Pull
* Opposite

**Pitch (MPH)** The maximum speed of a given pitch at any point from its release to the time it crosses home plate (Ratio)

**Pitch Type** The different variety of pitches available that differ in velocity, break, and movement (Nominal)

* 4-Seam Fastball
* 2-Seam Fastball
* Changeup
* Slider
* Curveball
* Sinker
* Cutter
* Split-Finger
* Knuckle-Curve
* Knuckleball

Reference: <https://github.com/kylevandermeulen97/statcast> for further definitions, r package versions, datasets, and code

**3. Fitted Probability of a Hit**

We attempted to calculate hit probability and identify ideal parameters of exit velocity and launch angle for our four respective hitters. We propose that each hitter is unique and is in need of personalized exit velocity and launch angle dashboards. Some have found the optimal launch angle between 20 and 40 degrees, and when you add in exit velocity over 100 miles per hour you are looking at a guaranteed hit [6, 8]. Petti and Nathan are right in their statement about the optimal exit velocity and launch angle being within those ranges. However, if our scope for our player’s exit velocity and launch angle is too wide, then all we are going to get is an average hitter. But if we personalize exit velocity and launch angle for hitters, then the result will be close to perfection.

**3.1 Background**

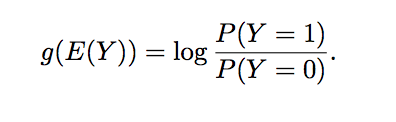
With the emergence of Statcast previous attempts to calculate hit prediction have been overwritten. The motivation for this study came from Jim Albert [1] who similarly combined exit velocity and launch angle to explain the Ball in Play results for individual players. Albert used a logistic model, but did not consider other predictive models. We built upon his analysis through smoothing the results, finding significant predictors for exit velocity, and gaining further insight into batter tendencies. On the contrary others [5,7] have pointed out the flaw in using raw exit velocity for analysis. First, if you use exit velocity as an indicator of player ability, then you must also accept that one player’s exit velocity is a function of his opponents. Essentially, they are saying that a player’s exit velocity is biased by the teams they face. We have to disagree as the batter is in control of the bat and the pitcher can’t change that. Additionally, looking at an average takes all of the greatness away from a player’s performance against certain pitchers and teams. If a player is exposing a pitcher, then that shows pure talented domination over an opponent. Lastly, to reiterate our point of an individualized exit velocity and launch angle, is an article by [2,3] who in one of his graphs highlights a key point. The relationship he visualized was between the expected runs added and exit velocity for contact between 22 and 28 degrees of launch angle in 2015. The valley in the chart, with bloop singles on one side and doubles on the other, has been labeled the “doughnut hole” because it’s surrounded by exit velocity and launch angles that are more productive. These visual dips focus on the fact that there is a large variance over the broad spectrum of players in being able to produce runs with the intersection of exit velocity and launch angle.

**3.2 Methodology**

**3.2.1 Models Assessed**

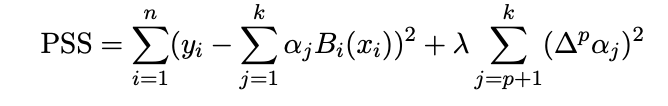
Five varieties of models are evaluated using the dependent variable outcome (0 to 1) to estimate the fit and probability of a hit: generalized additive models (GAM), p-spline smoothing, restricted maximum likelihood (REML), local regression (loess), and generalized cross validation (GCV). As the dependent variable is binary (0,1), the predicted hit probability is capped to ensure predictions fell within this range.

GAM is an additive modeling technique that captures the overall effect of the predictors through estimating the smooth relationships with the dependent variable. These relationships can either be linear or nonlinear. The method proves to be useful in uncovering nonlinear covariate effects [12]. We can then effectively predict outcome g(E(Y )) using the logit link:

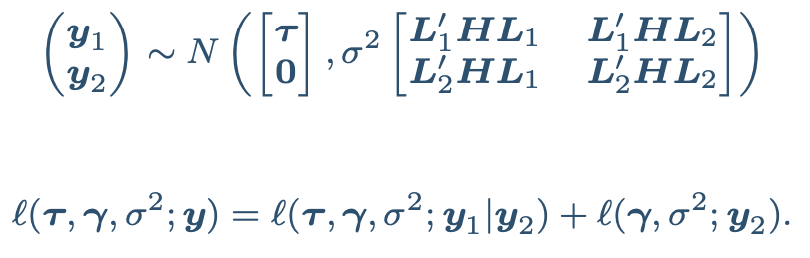


Within GAM is a method called smoothing splines, that can estimate the smooth function by minimizing the penalized sum of squares. Moreover, we utilized p-splines which penalize the differences between the coefficients to invoke smoothness. P-splines allow the building of a mixing and matching of a variety of additive smooth structure components. Essentially weights are put on each of our splines to avoid overfitting the data, while allowing for the fitting of the data through the splines.

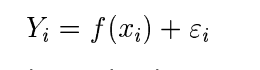
We chose a set of 10 *k* splines and Bj (x) denotes the j-th B-spline evaluated at x and αj denotes its coefficient. We can then effectively penalize changes at the knots for the outcome (*yi*) and fit the splines to our dataset [13].



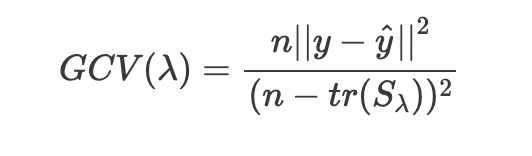
Thirdly is a method that focuses on estimating variance components. The Restricted Maximum Likelihood Estimation (REML) model approach is a form of maximum likelihood estimation that uses a likelihood function calculated from a transformed dataset. In distinction to likelihood estimation REML yields unbiased estimates for variance and covariance parameters. “The REML model’s likelihood is partitioned into two independent parts: one (y1 = L′1y) relating to the fixed effects and one part (y2 = L′2y) relating to the residual contrasts (zero expectation)” This REML estimate of variance is the usual MSE with a n – 1 denominator that includes an adjustment for degrees of freedom [15].



On the other hand, we decided to implore a model that was nearest neighborhood-focused rather than the penalized methods that were mentioned previously. Local regression (Loess) belongs to the nearest neighbor class and produces a smooth curve by fitting a weighted regression. This type of regression was utilized because of its well known use for a fitting a smooth surface between an outcome and its predictor variables. The properties are that the curve be smooth and that the curve minimize the variance of the residuals or prediction error. The standard model form takes f(x) which is an unknown function, e*i* is an error term representing variability in the other sources not found in f(x) [14].



Lastly, we included a cross-validation approach named Generalized Cross Validation (GCV). GCV is centralized around removing one data point at a time, smoothing the data with each validation and mitigating overfitting. For every linear smoother../../../var/folders/x7/f81w76vs05q3b8_qqr53qxk80000gn/T/TemporaryItems/(A%20Document%20Being%20Saved%20By%20screencaptureui%207)/Screen%20Shot%202020 the cross-validation criterion consists in minimizing the following quantity: where λ is the penalty coefficient, n the number of observations and tr(Sλ) is the trace of the matrix Sλ [15].

****

**3.2.2 Results and Discussion**

It is well accepted that exit velocity and launch angle affect a batter’s performance. We can see that the model’s mean absolute error (MAE) ranged from 0.0328 to 0.0583 across the board. The MAE is the difference between the true value and the predicted values for each individual prediction [16]. Which proves to be significant in knowing the models are making accurate predictions. The REML model performed best in Gallo and Judge’s case. While Yelich’s model that was best was local regression. Lastly, Ozuna performed well with the generalized cross validation. We can see the necessity to tailor the best model to each player in terms of the lowest MAE value.

**Table 1. Mean Absolute Error for all Models-Gallo**

Model MAE

REML 0.0513

Local Regression 0.0514

Generalized Cross Validation 0.0516

P-spline Smoothing 0.0524

Generalized Additive Model 0.0547

**Table 2. Mean Absolute Error for all Models-Judge**

Model MAE

REML 0.0447

Generalized Cross Validation 0.0449

P-spline Smoothing 0.0453

Local Regression 0.0496

Generalized Additive Model 0.0513

**Table 3. Mean Absolute Error for all Models-Yelich**

Model MAE

Local Regression 0.0442

REML 0.0464

Generalized Cross Validation 0.0465

P-spline Smoothing 0.0473

Generalized Additive Model 0.0545

**Table 4. Mean Absolute Error for all Models-Ozuna**

Model MAE

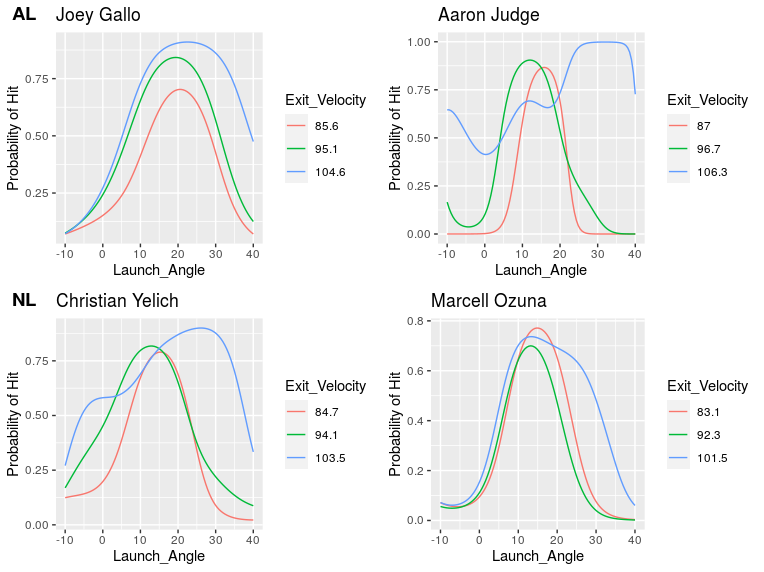
Generalized Cross Validation 0.0328

REML 0.0329

P-spline Smoothing 0.0335

Local Regression 0.0406

Generalized Additive Model 0.0583



*Figure 1*. Fitted probability of a hit as a function of the launch angle for each exit velocity.

To better understand the best model fits for each of our players, visualizations were made to better understand our batter’s tendencies in his exit velocity. Three different lines are graphed on each plot that highlight three different exit velocity values. The green line represents the mean exit velocity for our hitter over the whole 2019 season. While the red line represents an exit velocity 10% below the mean and the blue line signifies 10% above the mean. The 2x2 matrix in the first row has our two American league hitters and the second has the two National league hitters. Additionally, the first column has our two left handed hitters in Gallo and Yelich. Whereas the second column is our two right-handed hitters. Setting up the matrix in this way for helpful to compare and contrast the two leagues and their handedness. If we take a look at each individual plot we can begin to make some key observations for each of the players.

First, we can see that Gallo’s lines have the largest distributions of any of the hitters. In 2019 the numbers he put up affirm that he indeed had a large distribution and was highly inconsistent with his exit velocity. His numbers are as such: 2.53 AVG, 22 HR, and 61 hits in 241 at bats. Gallo ended up finishing sixth in the MLB in regards to average exit velocity. He is going for power which highly decreases his potential average and hit amount, but if he does make solid contact it most likely will be a homerun. Based on the three line graphs his optimal intersection of launch angle and exit velocity to maximize hit probability would be 20 degrees for each respective exit velocity. We can conclude that it is going to be harder to predict Gallo’s probability of a hit based on his exit velocity and his launch angle.

Secondly, unlike Gallo’s graph Judge’s lines are more tightly knit. He is more consistent and has the highest probability of a hit of any hitter as shown on the y-axis. Optimally, for the three respective exit velocities he wants to aim for 12-14 degrees. Something to note though is that 10% above the mean we see a dip around 15-20 degrees in his launch angle off of the bat. This is a significant note that can be brought to coaches to be able to work on to be able to smooth out that line to go up rather than cave in. Contrarily, competition could use this information to target certain areas of the strike zone that most likely produce 15-20 degrees of launch angle.

Thirdly, in similarity to Gallo tendencies are Christian Yelich’s. His line graphs although not as widely distributed as Gallo’s are distributed more than Judge’s and Ozuna’s. The optimal point to maximize hit probability over the three lines would be at an angle of 15 degrees. Once again we can conclude that it is going to be harder to predict Yelich’s probability of a hit based on his exit velocity and his launch angle.

Lastly, Marcell Ozuna had a graph distribution that was the tightest of them all. It is relatively easy to predict his probability of a hit based on his exit velocity and his launch angle. To maximize hit probability I would recommend once again, 15 degrees. So after taking an in depth look at each of our batter’s tendencies we can focus on weak spots, as well as amplify strong points.

**4. Predictors of a Hitter’s Exit Velocity**

We now shift our attention to the second objective addressed in this paper: What are significant factors in a batter’s exit velocity (BEV) and are there any universal factors in BEV calculation? Most perceptions have launch angle and pitch speed as significant in BEV calculation [5]. But other than that there is not an overwhelming amount of research. We wanted to gain a greater sense of what underlying factors go into a BEV and how can we leverage that to maximize performance; also, if some of those factors are significant, do they apply to the broad spectrum of players? This section details the significant variables in each of our individual regression results and then visualizations are given to gain even a greater sense of our batter’s strengths and weaknesses. We once again used MLB Statcast Game Logs from Baseball Savant [4] for the same four players for the 2019 season for the analysis of this section.

**4.1 Methodology**

Before we began there was data manipulation that had to be done for our independent variables. First, we subsetted each of the individual pitch types, directions, and events to be able to gain the full understanding and relationship with BEV. Secondly, we had to format our game date variable to a date and separate it into its own variables year, month, and date. This was completed to ensure that we could compare the significance of BEV over a time series (monthly). An OLS approach was used because of our correlation results, linearity in the coefficients and error term, normal distribution, and the variance met model assumptions. After that we went on to choose the best models for subset size in terms of BIC. It was determined after this test that we were going to include the best subset size for each player. We then proceeded to perform the OLS regressions and calculate the results. They are outlined below.

**4.2 Results and Discussion**

Here are the significant predictors of exit velocity for each hitter at the 95% confidence interval.

Significance levels:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘+’

Table 5: Gallo OLS Regression Summary

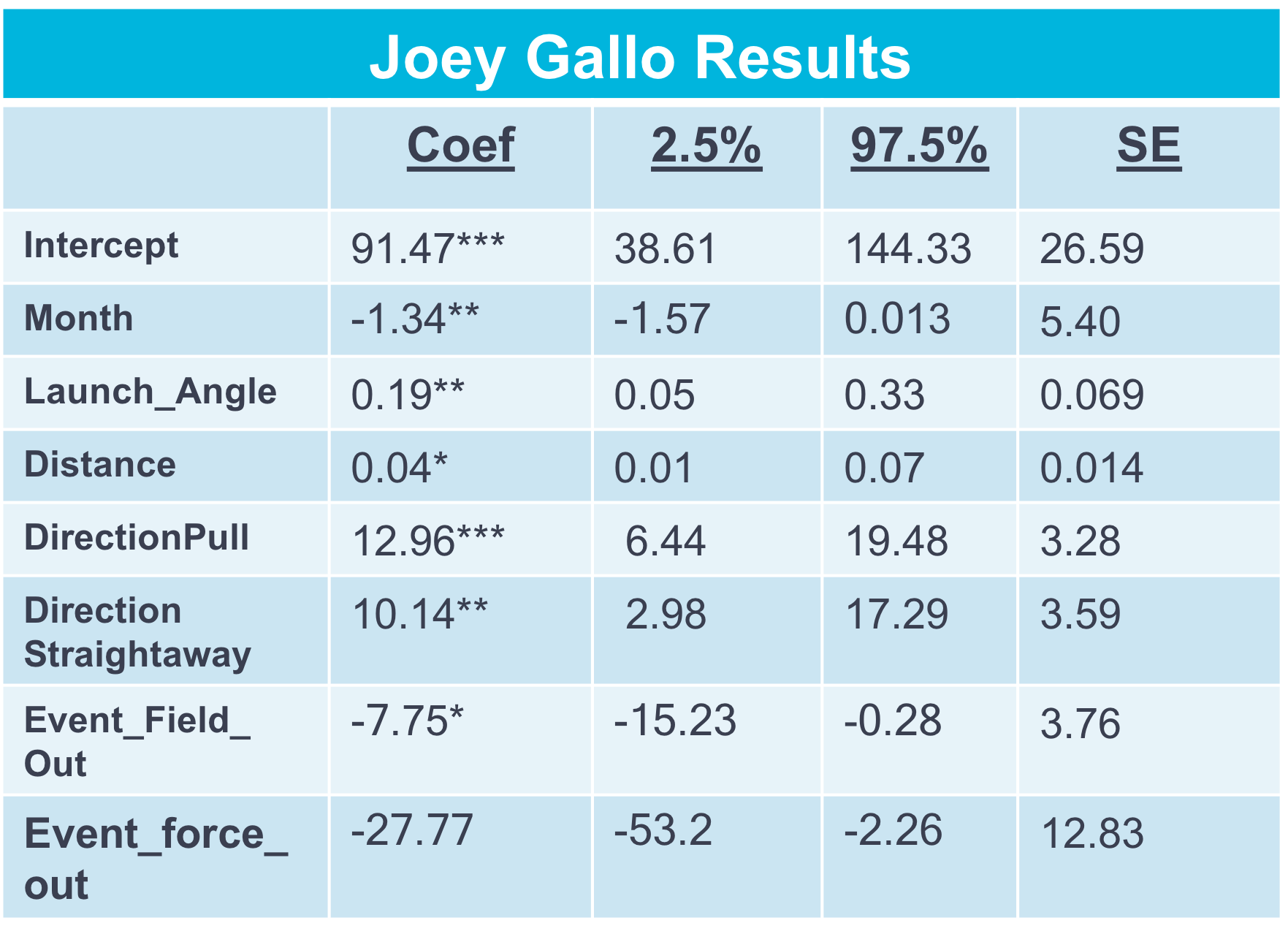


Table 6: Judge OLS Regression Summary

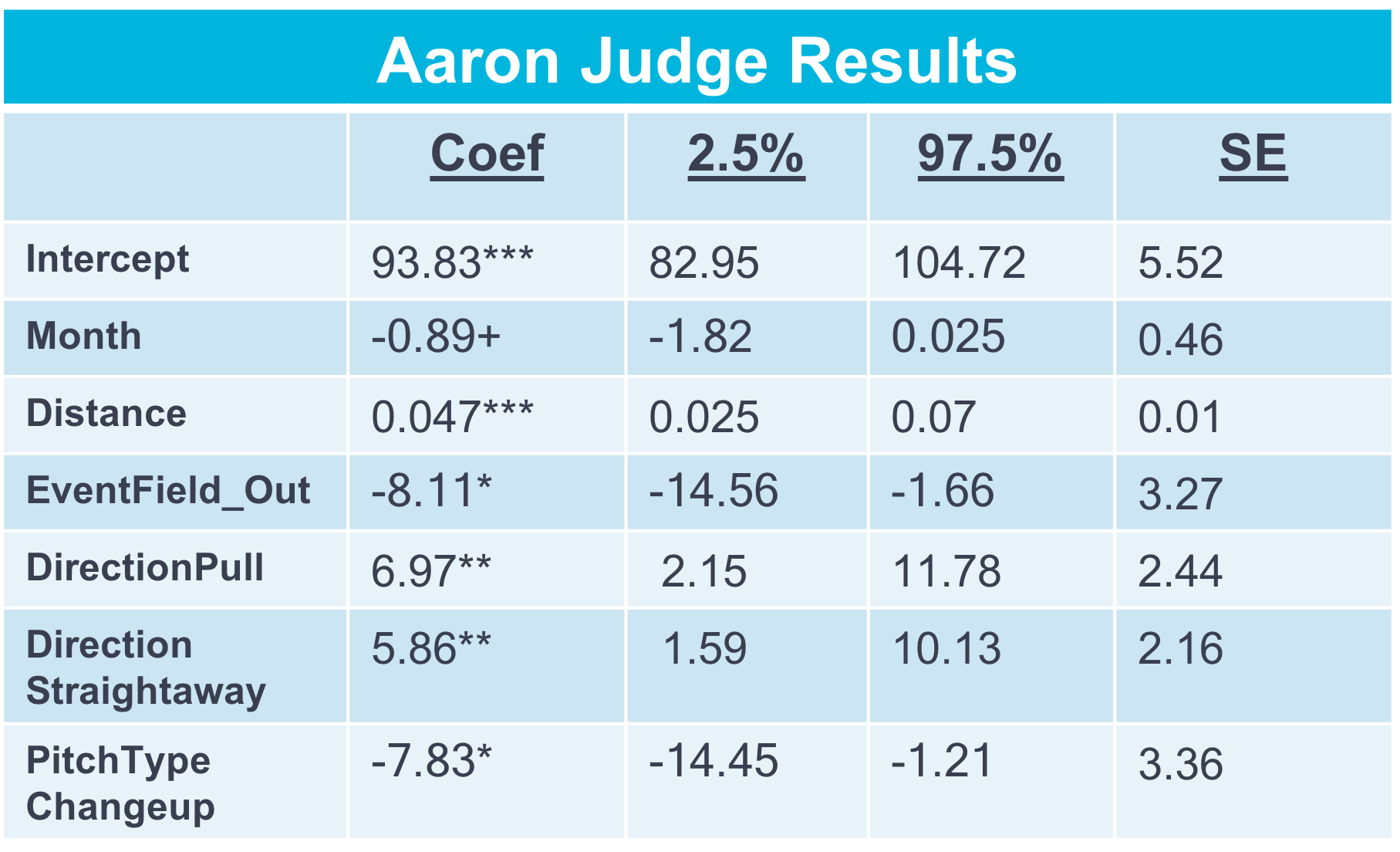
****

Table 7: Yelich OLS Regression Summary

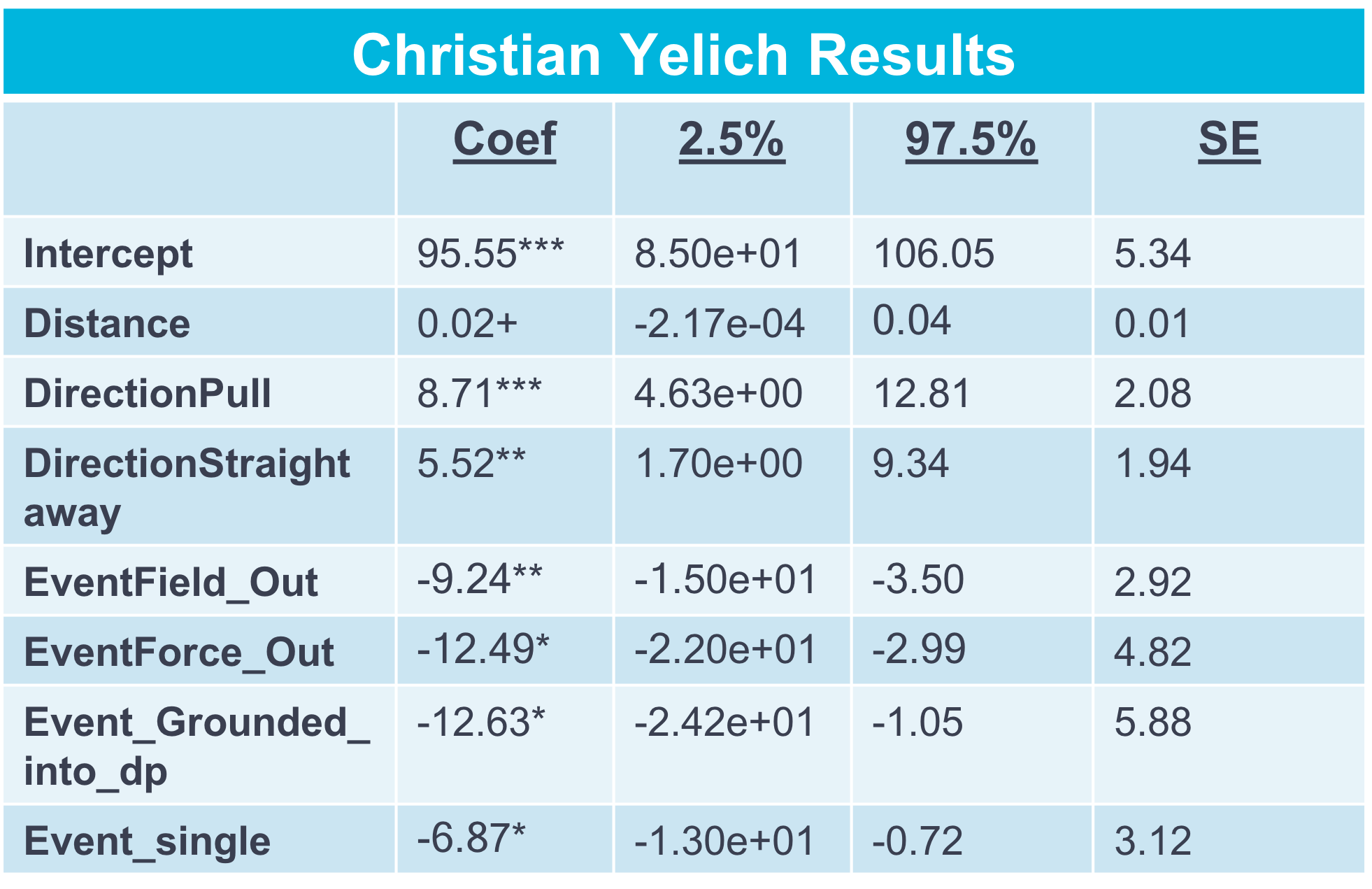
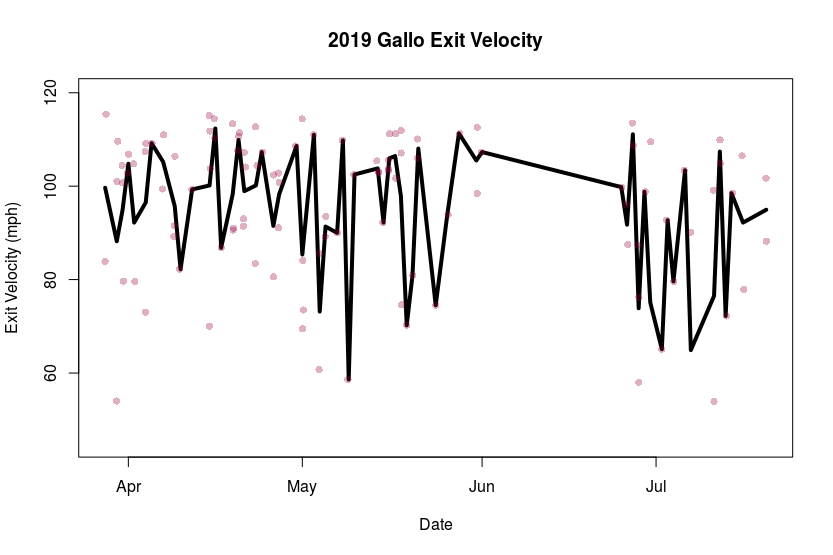
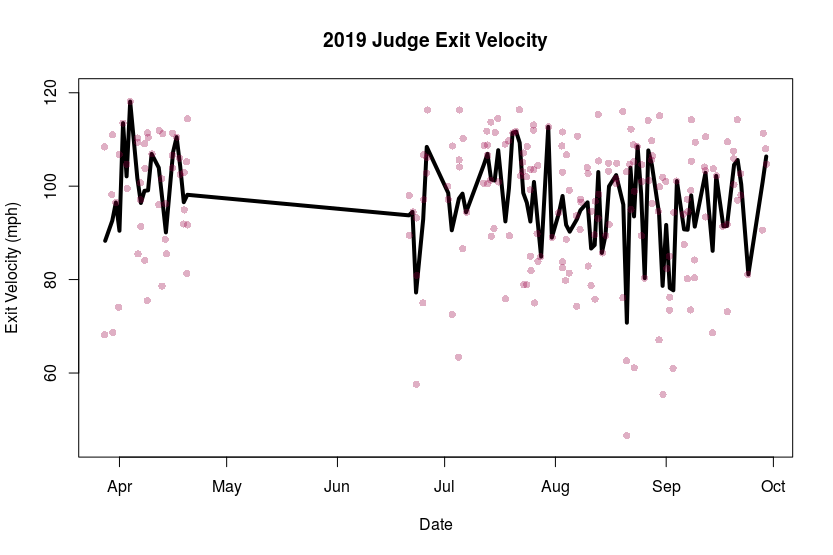


Table 8: Ozuna OLS Regression Summary

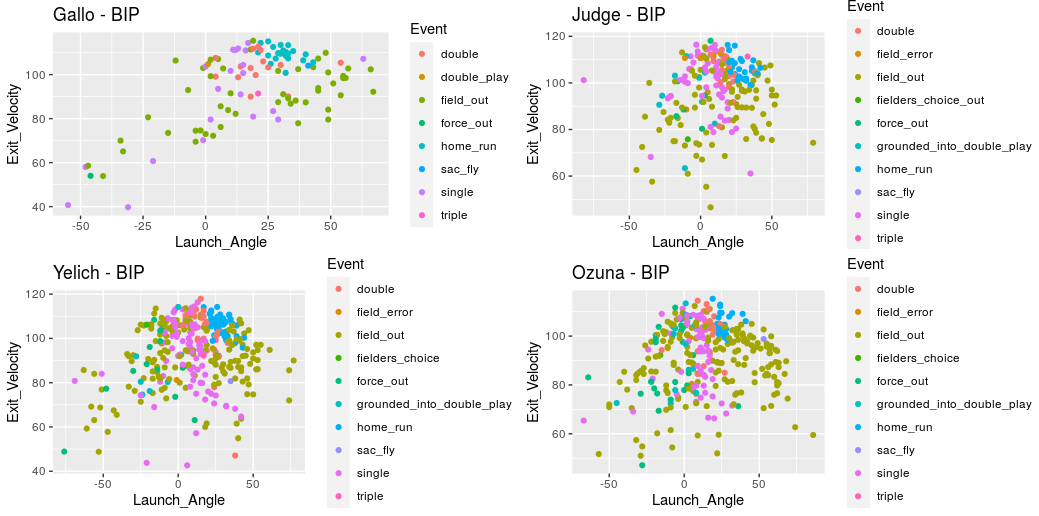
****

Our hitters all vary in what is significant in affecting each of their BEV’s. But there were factors that were exhibited across all four hitters: Distance, Direction(Pull), Direction(Straightaway), and Event(field\_out). All that aside let’s look deeper into these predictors by graphing some of them. Let’s begin with the month in which our player performed his at bats. Two of our players, both in the American League had negative significance with the month predictor. We can display the batter’s exit velocity over the course of the season to gain insight into our batter’s BEV trends over the course of the season. Let’s look specifically at Gallo and Judge who had negative significance with the month.

*Figure 2.* Game by Game Time Series ****

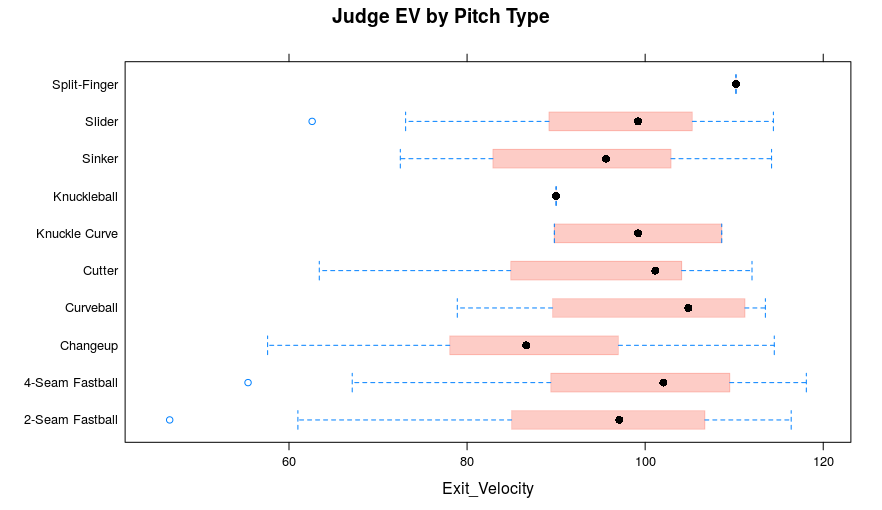
There is without a doubt that Gallo and Judge negatively struggled month to month. Gallo once again shows the inconsistency reaching BEV as high as 115 mph and lows in the 60 mph range. He struggled especially in May and July with huge fluctuations in his BEV. After that he was hurt for the rest of the year. Judge although was more consistent in keeping his exit velocity constant, but struggled in late August. On a side note, the flat lines displayed are the timeframes in which Gallo and Judge were hurt. As you can see these plots are extremely useful in identifying trends in our hitter’s BEV over the course of the whole season. A suggestion may be to tell our player to focus more time and effort into hitting in the months that he has previously shown dramatic drops in his exit velocity.

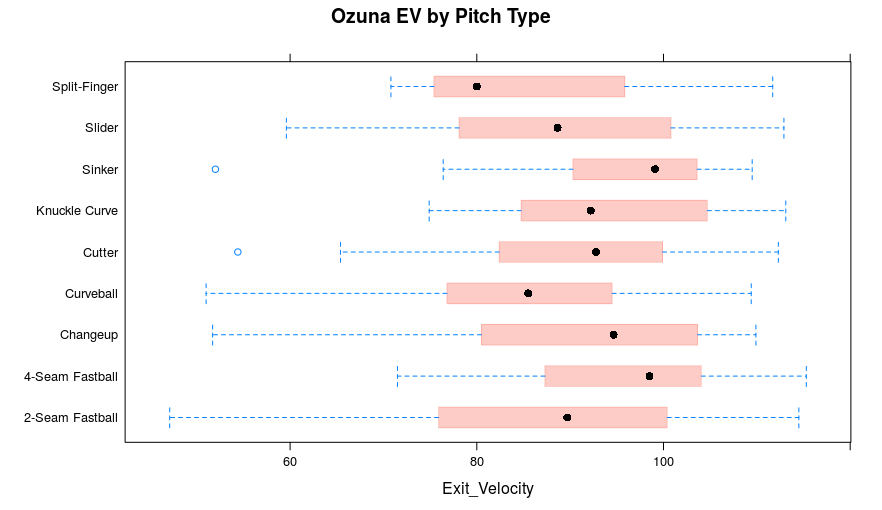
The four scatterplots below provide an in depth look at which intersection of launch angle and exit velocity produces which outcome. The home runs in the light blue color all seem to be in the same vicinity. While other outcomes appear to be more spread out over the grid. The significant predictor field\_out seems to be strewn all over the graph, which gives us reason to believe that provided significance. Now we can instill meaning into the junction of exit velocity and launch angle as we know the end results. Everyone nowadays wants to hit the long ball, so observing the range of exit velocity and launch angle for each hitter to be able to hit a home run proves valuable.



*Figure 3*. Launch Angle (degrees) against the Exit Velocity (mph) for all of hitter’s balls put in play where the color of the point corresponds to the outcome.

Finally, OLS discovered that for both of our right-handed hitters the pitch types they faced had a positive or negative effect on their exit velocity. For Judge every Changeup had a negative effect of -7.83 mph in his exit velocity. While Ozuna positively hit almost every pitch well with a positive effect on his exit velocity. He hit the sinker at the highest clip and the fastball came in at a close second. Other additional pitches that he hit fairly well were the cutter, knuckle-curve, and slider.





*Figure 4.* Box and Whisker Plots highlighting Judge and Ozuna’s Exit Velocity by Pitch Type

Both of these plots reaffirm the regression results in regards to the changes in exit velocity in relation to different pitch types. Judge’s changeup boxplot sticks out like a sore thumb and makes sense as he is hitting the fastball lights out. Some worthy insights for coaches to take away from this would be targeting each of these hitters based on their weakest exit velocity against certain pitches. Against Judge I would be focused on throwing pitches with movement. Such as Changeups, Two-seamers, and Sinkers. In Ozuna’s case it is a lot more difficult to assess where he should be targeted. He like Judge is hitting the fastball extremely well, but in terms of a recommendation I would pitch hanging pitches like curveballs and splitters that break. These graphs highlight both our batter’s strong points and their weak ones.

**5. Conclusion**

The factors having a significant impact on a hitter’s exit velocity that were displayed across all four hitters were Distance, Direction(Pull), Direction(Straightaway), and Event(field\_out). In addition to these universal factors were predictors that came up in a couple of our hitters. The month, launch angle, outcomes, and pitch types proved to be of significance in relation to some hitters. Moreover, the speed of the pitch when alongside other predictors had little to no impact on BEV. Overall, is the fact that there is a need to personalize exit velocity as every batter’s has strengths, weaknesses, and tendencies that can be enhanced or exploited.

In addition to recognizing the characteristics most likely to increase exit velocity, the insights in this analysis may be utilized to optimize a batter’s exit velocity. This is a significant note that can be brought to coaches to be able to work and smooth out their exit velocity line to go up rather than drop. Competition as well could use this information in game to target certain areas of the strike zone most likely produce a certain launch angle and exit velocity. In consideration of all these factors coaches, hitters, and competition have the opportunity to positively or negatively impact a hitter’s exit velocity.

Going forward, several different analyses could improve upon these findings. Does the position and age of a player influence their exit velocity? Clustering could be utilized to include more hitters to discover further information useful for decision making. Additionally, are there any differences between pro and collegiate hitters in obtaining significant factors the optimal exit velocity and launch angle for maximal hit probability? Overall, these findings have the potential to dramatically affect practice time allocation, in game decisions, and lineup positioning.

**References**

[1] Albert, J. (2016, August 15). Exploring Statcast Data from Baseball Savant. Retrieved from https://baseballwithr.wordpress.com/2016/08/15/exploring-statcast-data-from-baseball-savant/

[2] Arthur, R. (2016, May 26). Who's Hitting The Ball Harder This Year, And Who's Just Getting Lucky? Retrieved from https://fivethirtyeight.com/features/whos-hitting-the-ball-harder-this-year-and-whos-just-getting-lucky/

[3] Arthur, R. (2016, April 13). The New Science Of Hitting. Retrieved from https://fivethirtyeight.com/features/the-new-science-of-hitting/

[4] Baseball Savant: Trending MLB Players, Statcast and Visualizations. (n.d.). Retrieved from https://baseballsavant.mlb.com/

[5] Dutto, M. (2019, October 1). Debunking Bat Speed Myths. Retrieved from https://www.drivelinebaseball.com/2019/05/debunking-bat-speed-myths/

[6] Nathan, A. (2016, April 6). Going Deep on Goin' Deep. Retrieved from https://tht.fangraphs.com/going-deep-on-goin-deep/

[7] O'Rourke, S., & Judge, J. (2016, May 17). Prospectus Feature: The Need for Adjusted Exit Velocity. Retrieved from https://www.baseballprospectus.com/news/article/29210/prospectus-feature-the-need-for-adjusted-exit-velocity/

[8] Petti, B. (2016, June 14). Using Statcast Data to Predict Hits. Retrieved from https://tht.fangraphs.com/using-statcast-data-to-predict-hits/

[9] Stanton/Judge Project. (2018, July 6). Retrieved from https://www.kaggle.com/jidbro1/stanton-judge-project

[10] Statcast: Glossary. (2016). Retrieved from http://m.mlb.com/glossary/statcast

[11] Swanay, S. (2017, November 16). Baseball Analytics in the Statcast Era. Retrieved from https://www.casact.org/community/affiliates/cagny/1117/Swanay.pdf

[12] Hastie, Trevor and Tibshirani, Robert. (1990), Generalized Additive Models, New York: Chapman and Hall.

[13] Eilers, Paul H.C.; Marx, Brian D.; Durbán, Maria. «Twenty years of P-splines». SORT : statistics and operations research transactions, Vol. 39 Núm. 2 (July-December 2015), p. 149-186.

[14] Dias, R. (n.d.). Nonparametric Regression: Lowess/Loess. Retrieved from http://www.ime.unicamp.br/~dias/loess.pdf

[15] Wahba, G., & Craven, P. (1964, January 1). Smoothing noisy data with spline functions. Retrieved from https://link.springer.com/article/10.1007/BF01404567

[16] Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning with 293 figures and 78 tables*. New York: Springer.

[17] “readxl” package,  [https://CRAN.R-project.org/package=readxl](https://cran.r-project.org/package=readxl)

[18] “devtools” package, [https://CRAN.R-project.org/package=devtools](https://cran.r-project.org/package=devtools)

[19] “lattice” package, [https://CRAN.R-project.org/package=lattice](https://cran.r-project.org/package=lattice)

[20] “gam” package, [https://CRAN.R-project.org/package=gam](https://cran.r-project.org/package=gam)

[21] “corrplot” package, [https://CRAN.R-project.org/package=corrplot](https://cran.r-project.org/package=corrplot)

[22] “mgcv” package, [https://CRAN.R-project.org/package=mgcv](https://cran.r-project.org/package=mgcv)

[23] “tidyverse” package, [https://CRAN.R-project.org/package=tidyverse](https://cran.r-project.org/package=tidyverse)

[24] “leaps” package, [https://CRAN.R-project.org/package=leaps](https://cran.r-project.org/package=leaps)

[25] “ggvis” package,  [https://CRAN.R-project.org/package=ggvis](https://cran.r-project.org/package=ggvis)

[26] “ggpubr” package, [https://CRAN.R-project.org/package=ggpubr](https://cran.r-project.org/package=ggpubr)

[27] “PerformanceAnalytics” package, [https://CRAN.Rproject.org/package=PerformanceAnalytics](https://cran.r-project.org/package=PerformanceAnalytics)