By: Riva Lam and Leah Marcus

Polynomial Regression Analysis of Launch Angle

Summary: Baseball has in recent years increasingly used advanced statistical methods to evaluate players. Analyzing a player's advanced metrics can help scouts identify valuable players who might have been passed over using older and simpler assessments, and understanding these measures has helped players revitalize their careers. One of the most popular techniques has been the introduction of launch angle and its impact on exit velocity. By adjusting their swing path and thus increasing launch angle, batters increase their exit velocity which results in more hits and more power at the plate. Launch angle has a nonlinear relationship with exit velocity, as both negative and high launch angles result in lower exit velocities and fewer hits. Batters who can hit balls in a range from about 0 to 25 degrees have higher exit velocities and are more likely to get hits. Polynomial regression is a statistical tool that can model the relationship between launch angle and exit velocity.

Introduction: Polynomial regression is used when a model has a nonlinear relationship and thus cannot be accurately represented with standard linear regression. The predictors are raised to a degree, so the data can be better modeled using a nonlinear curve (James, 266). Polynomial regression can range from the most simple case of quadratic regression, which has the form $Y = B_0 + \beta_1X + \beta_2X^2 + \varepsilon$, to general polynomial regression of the form $Y = B_0 + \beta_1X + \beta_2X^2 + \beta_3X^3 + \ldots + \beta_kX^k + \varepsilon$, where k represents the degree of the model (Cannon, 122-125). The larger the value of k, the more flexible the model becomes, although usually k is no bigger than degree 3 or 4 because of over-flexibility of the curve. The degree of the polynomial function can be chosen by using an ANOVA table to do hypothesis testing to compare a degree 1(linear) model to a degree 2 (quadratic) model, and then quadratic to a degree 3 (cubic) model, and so on until all nested models have been tested. The degree is then picked based on the simplest model that sufficiently explains the data (James, 288-290).

One application of polynomial regression has been used when analyzing launch angles in baseball. Launch angle measures the degree at which a batter makes contact with a baseball with his bat. By determining the launch angle of a batted ball, exit velocity can be predicted. Exit velocity is essentially how hard a ball is hit by a batter and is a factor in determining whether or not the batter gets a hit. In general, balls batted at the right launch angle and at a high exit velocity are more often hits. This has led to what many baseball sabermaticians refer to as “the fly-ball revolution,” where batters are increasing the launch angles of their swings in an attempt to avoid ground balls, which generally are hit at a negative launch angle and are very likely to result in outs (Trueblood, “Lift”). However, balls hit at too high of a launch angle also have lower exit velocities and also often result in fly-ball outs (Alonso). The highest exit velocities, upwards of 100 mph, come from launch angles roughly between 0 and 25 degrees, with hard line drive hits in the air between 18 and 25 degrees (Trueblood, “Launch”). These are batters best chances of not only getting a hit but getting an extra base hit such as a double or home run because they are the hardest to field; the harder a ball is hit, the less time fielders have to react.

Methods: The data set used was for Mike Trout from his 2015 Statcast data (Albert). All of Trout’s batted balls in play were graphed and separated by color based on whether they were hits or outs, for simpler practical analysis of how launch angle impacts exit velocity and the outcome of the batted ball, with launch angle on the x-axis and exit velocity on the y-axis. The graph appeared to be best fit by a polynomial function rather than a linear model, as the data set exhibited curvature. The degree of the polynomial function was then decided by hypothesis testing using ANOVA and nested models, comparing two sets of functions from a set that consisted of functions from a linear model to a degree ten polynomial. The degree was chosen based on the simplest polynomial function that sufficiently described the data.

Results:

We fit ten different models and compare the simpler model to the complex model. The degree-6 polynomial appears to provide a reasonable fit to the data:

Table 1. Hypothesis Tests.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
383	82686.47	NA	NA	NA	NA
382	70762.24	1	11924.22435	64.19309194	1.449942e-14
381	70583.98	1	178.26850	0.95969396	3.278973e-01
380	70566.16	1	17.81221	0.09589059	7.569904e-01
379	70513.33	1	52.83689	0.28444309	5.941205e-01
378	69851.64	1	661.68613	3.56213347	5.988582e-02
377	69539.65	1	311.98728	1.67955813	1.957824e-01
376	69517.16	1	22.48861	0.12106561	7.280774e-01
375	69498.33	1	18.83801	0.10141291	7.503184e-01
374	69472.58	1	25.74343	0.13858769	7.099000e-01

Degree-6 Polynomial, where x = Launch Angle:

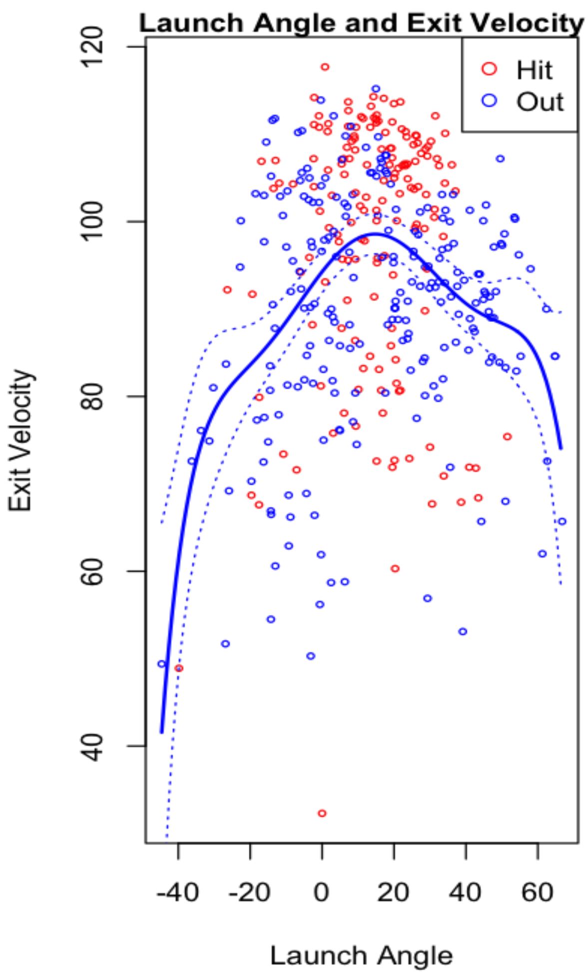
Table 2. Coefficient Table for 6-Degree Polynomial.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	93.244156	0.6928066	134.5890048	2.005907e-321
poly(Launch_Angle, 6)1	24.743503	13.5938478	1.8201986	6.951945e-02
poly(Launch_Angle, 6)2	-109.198097	13.5938478	-8.0329056	1.223982e-14
poly(Launch_Angle, 6)3	13.351723	13.5938478	0.9821886	3.266352e-01
poly(Launch_Angle, 6)4	4.220451	13.5938478	0.3104678	7.563764e-01
poly(Launch_Angle, 6)5	7.268899	13.5938478	0.5347197	5.931581e-01
poly(Launch_Angle, 6)6	-25.723261	13.5938478	-1.8922722	5.921864e-02

Exit Velocity = 93.244156+ 24.743503x - 109.198097(x^2) + 13.351723(x^3)+ 4.220451(x^4) + 7.268899(x^5)-25.723261(x^6)

The final scatter plot of all of Mike Trout's batted balls in play, fit with the degree six polynomial function:

Figure 1. Scatterplot with Degree Six Polynomial.



Discussion: Since the highest exit velocity is produced from a range of middle degrees and lower exit velocities are the result of both negative launch angles and the highest launch angles, polynomial regression appeared to be a better fit for fitting a model where the predictor variable x = launch angle and y = exit velocity than a linear model. After it was determined that a degree six polynomial function was the best fit, based on the p-value of .05989 (Table 1) being the most significant and corresponding to a comparison of a degree six model to a degree five model, the scatter plot was fit with the model (Table 2 and Figure 1 respectively). While none of the models had a p-value of .05 or under, the degree six model was by far the closest and was at least close to .05. As a final note, as expected from previous research, a majority of Trout's hits came from launch angles from approximately 0 to 30 degrees and resulted in his higher exit velocities, and balls hit outside of this range were often outs at lower exit velocities.

Acknowledgments: A batter's launch angle is not the only factor that determines exit velocity. For example, a pitcher's own velocity, type of pitch thrown, and a batter's own strength all impact exit velocity. A batter's strength can be accounted for by only looking at individual batter's statistics, but breaking down exit velocities based on the type of pitch hit and pitch velocity requires a much more in depth analysis. Additionally, exit velocity cannot by itself predict whether a batter will get a hit, as that is affected by numerous other variables, such as the quality of the other team's fielders, the pitcher, and even the time of day. Exit velocity does serve as a simple way to measure a batter's ability, as higher average exit velocities generally translate to better hitters.