**Investigating the Effect of Temperature on Exit Velocity and Home Run Distance**

**Austin Leonard**

**December 12, 2019**

**SDS 358**

****

**Introduction**

**Objectives:** There is a common prevailing wisdom in baseball discussion that hot temperatures create a favorable hitting environment due to reduced air pressure/density. The objective of this project was to investigate the veracity of this claim through the lens of exit velocity and home run distance. Covariates include air pressure, precipitation, pitch speed, pitch type and launch angle.

**Hypotheses:** Temperature will significantly moderate the effect of exit velocity on home run distance – at lower temperatures, balls hit a certain speed will travel shorter than they would at hotter temperatures.

**Methods**

**Sample:** Data represents 359 home runs and their distances (feet) hit by Atlanta Braves players in the regular season and playoffs at SunTrust Stadium in Atlanta from 2015-2019. Additionally, data contains other quantitative variables such as exit velocity (mph), launch angle (degrees), precipitation (inches), pitch speed (mph), temperature (Fahrenheit), and air pressure (atm). Pitch type was included as a categorical variable with two levels: Fastball and Off-speed. Analysis found 10 observations with abnormal Cook’s Distance values; these values were removed before running a final model.

**Analysis Method:** Using R, a multiple regression model was run which included a quantitative-by-quantitative interaction term.

**Descriptives**

**Response Variable**

|  |  |  |
| --- | --- | --- |
|  | **Median** | **Standard Deviation** |
| **Home Run Distance** | 400 | 23.87 |

**Explanatory Variables:**

|  |  |  |
| --- | --- | --- |
|  | **Median** | **Standard Deviation** |
| **Exit Velocity** | 103.3 | 4.2 |
| **Temperature** | 78 | 7.48 |

**Results**

|  |  |  |  |
| --- | --- | --- | --- |
| **Simple Slope** | **Estimate** | **T-statistic** | **P-value** |
| * 1 SD | 3.84 | 13.55 | 1e-33 |
| Mean | 4.15 | 19.33 | 1.12e-56 |
| +1 SD | 4.46 | 14.92 | 4.74e-39 |

**Results table:**

**Model Summary:**

Overall Fit: F(8, 339): 53.03, p < 0.05. , adjusted

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coefficient** | **Estimate** | **SE** | **t-value Statistic\*** | **P-value** |
| **Pitch Speed** | -0.159 | 0.228 | -0.696 | 0.49 |
| **Pitch Type: Off-Speed** | -0.447 | 2.71 | -0.165 | 0.87 |
| **Launch Angle** | -0.059 | 0.178 | -0.33 | 0.742 |
| **Air Pressure** | -2.26 | 8.63 | -0.262 | 0.79 |
| **Precipitation** | -1.32 | 2.87 | -0.461 | 0.65 |
| **Temperature** | -0.025 | 0.115 | -0.221 | 0.83 |
| **Exit Velocity** | 4.15 | 0.21 | 19.3 | <2e-16 |
| **Temp\*EV** | 0.042 | 0.026 | 1.58 | 0.11 |

A close up of a map

Description automatically generatedA screenshot of a map

Description automatically generated

**Assumptions**

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated**Assumptions:** In this analysis, I checked the assumptions of multicollinearity and homoscedasticity. The former was checked by examining Variance Inflation Factor and satisfied by mean-centering the Variables of Interest. The latter was examined using a Residuals vs. Fitted plot. I also ensured there were no outliers in the final model by examining a Cook’s Distance plot.

**Discussion**

**Interpretation:** The overall model is significant, F(8,339) = 53.03, p < 0.05. However, the interaction term was not significant, , p = 0.11. The variable of interest, exit velocity, was found to be significant at the mean of the moderator, temperature, as well as one standard deviation above and below the mean of the moderator. The region of significance was found to be quite large: [-43.73, 408.73]. This completely encapsulates the range of observed data, and thus the interaction between home run distance and exit velocity is always significant for my sample. This contradicts my pre-analysis hypothesis.

**Limitations:** While the data passed all assumptions and the overall model was significant, the interaction was not significant. Perhaps the data shouldn’t have excluded Braves away games; that may have improved predicting power. Additional analysis should also consider the composition of baseball, which may have a larger effect on the flight of the ball.

**Implications:** This analysis has important implications in the greater context of the current MLB landscape. Much fuss has been made about “juiced” baseballs and a historic increase in the amount of home runs, and the goal of this analysis was partially to investigate if that spike was due to altered baseballs or if it was just a natural fluctuation due to a rise in the amount of great hitters. The peak of this controversy, and the inspiration for this study, came in the playoffs, when balls were soaring out of stadiums in freezing temperatures. If I were to run this analysis again, I would include more teams; preferably some based in colder climates such as Toronto or Minneapolis. I would also look at overall fly ball distance instead of home run distance – by including non-home run observations, the model might gain better insight into the overall effect of temperature on the flight of baseballs, and perhaps may even find a significant interaction. I would be sure to include Colorado in a future study – the Mile High City is known for artificially inflating offensive numbers due to its exceptionally low air pressure. As stated above, the composition of the baseball would also be an interesting factor to consider.

**References:** Baseball data from Baseball Savant: <https://atmlb.com/2NIwJZM>

Temperature data from NOAA: <https://www.ncdc.noaa.gov/cdo-web/search>