Predicting an MLB Player's Offensive Performance

Cristian Hendriks (hendriks@sonoma.edu), RJ Leon (leonr@sonoma.edu), Eric Polverari (polverae@sonoma.edu) & Matthew DiLeonardo (dileonar@sonoma.edu)
Dr.Newman

Background

wOBA (weighted on base average) is a stat many consider to best asses a baseball player's batting performance. Using MLB players' offensive statistics, such as straight away percent and pop-ups percent, from 2015-2019, the best predictor variables to predict a player's wOBA were analyzed through linear regression. We specifically focused on predictor variables that affected a player's wOBA percentage.

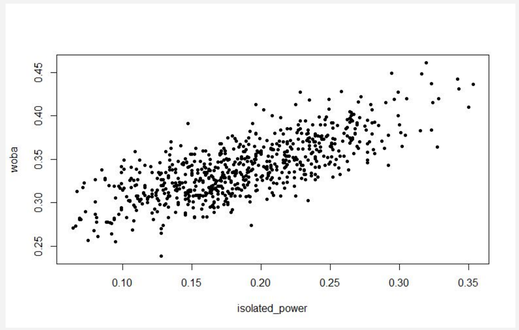
- Isolated_power - measures raw power of a hitter
- sprint_speed - a players top running speed
- b_hit_line_drive - number of line drives hit
- Popups_percent - represents the percentage of balls hit characterized as pop-ups
- Straightaway_percent - rate of hits going towards the middle of the field
- in_zone_swing - number of times a player swings at a ball in the strike zone
- The goal of this project was to construct a model that can predict a players overall offensive performance.

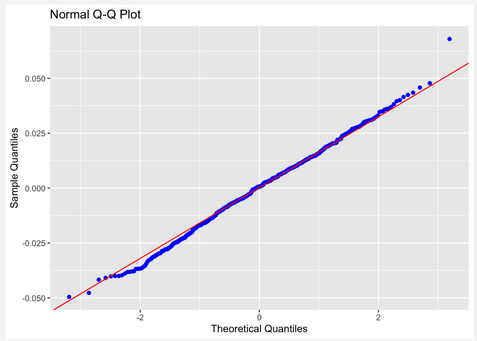
Significance Of This Study

- The significance of our study was to see how effective a wOBA Stat could be used for GM's, Coaches, and Scouts to determine a player's contribution to the team.

Research Question/Hypotheses

- What relationship is there between a player's offensive performance and their individual hitting statistics?
- Our group was interested in finding variables that many not be commonly associated with offensive production.





Methods

- 708 data points from MLB players
- Specific to players' batting stats
- Data source: Statcast
- Scatterplots to assess linearity
- Checked that model satisfied assumptions of regression inference
- Multivariate regression using Rstudio

Results

Based on our results, we found that our model is useful for predicting a player's wOBA based on the batting statistics described in our background. Above all else, a player's isolated power number is the best predictor of wOBA. We have a residual standard error of 0.01715, meaning 95% of the wOBA values fall between 2 standard errors of the predicted value. The model's R square is 0.7482, which means that roughly 75% of a player's wOBA can be predicted by our model. We have a significant p-value <2.2e-16, which means we have evidence to conclude that this model is useful to predict a player's offensive performance for a season by their different statistics.

Discussion

- We intentionally avoided variables used in the wOBA formula as that would skew the results and provide little other insight.
- The variables we came up with are not in the wOBA formula, but can be good predictors of wOBA which is considered a good overall single stat to evaluate a player's performance.
- Keeping these things in mind, what we have done is found things that scouts/general managers can use to evaluate players.