

Justin Skehill and Haley Waiksnis

20 December 2024

Project Proposal

PARK SPECIFIC xwOBA CALCULATIONS

Brief Introduction:

For our project, we will be looking at park-specific xwOBA calculations, where we will be examining the differences in players' xwOBA at distinct MLB ballparks across the country. We find this to be an incredibly interesting topic on which so much information can be drawn from. We will first start with understanding certain MLB statistics such as wOBA which allow us to determine a hitter's overall offensive value. The wOBA statistic provides all of the certain types of hitting into one true value, in which they are weighted to their legitimate run value. Upon research we have also found out that wOBA is far more accurate than other metrics such as batting average, on-base percentage, and slugging percentage. The xwOBA of a player takes into a greater effect the skill of that certain player, bringing into concern measurements such as launch angle and exit velocity. We will look more into detail in our methods section of our page what xwOBA does and how it is calculated. Upon new calculations of xwOBA, we will then start to address park factors and how their differences can draw new light on xwOBA values. Park factors are a fascinating part of research for this particular topic because we can draw conclusions concerning certain parks. Certain parks allowed an advantage to a particular player or to the entire league overall. We look forward to researching and learning more about our project so that we can make significant progress towards this topic come the spring.

Rough Outline:

Overview

- **Overview of wOBA and xwOBA:**

- **What is wOBA:** wOBA stands for weighted on-base average. It is a baseball statistic that measures a player's overall offensive value by assigning different weights to each type of hit (like singles, doubles, home runs) based on their actual run value. It is a more accurate way to evaluate hitters than traditional statistics like batting average or on-base percentage. But it is also important to add that wOBA can also be determined on luck. Some hits could be bloopers that happen to land in no man's land on the field. There also seems to be different wOBA calculations for each MLB season. It would be interesting to investigate just how that difference is made. But the equation provided below is a relative example of how wOBA can be calculated.

$$wOBA = \frac{0.696 \cdot uBB + 0.726 \cdot HBP + 0.833 \cdot 1B + 1.244 \cdot 2B + 1.569 \cdot 3B + 2.004 \cdot HR}{AB + BB - IBB + SF + HBP}$$

Where:

- uBB = Unintentional Base on Balls (Walks)
 - HBP = Hit by Pitch
 - 1B = Singles
 - 2B = Doubles
 - 3B = Triples
 - HR = Home Run
 - AB = At Bat
 - BB = Base on Balls (Walks)
 - IBB = Intentional Base on Balls (Walks)
 - SF = Sacrifice Flies
- - **What is xwOBA:** xwOBA stands for expected weighted on-base average. It is a statistic that helps us understand how effective a hitter is at getting on base, not just with hits, but with the type of hits. XwOBA allows us to factor skill into the picture, factoring out those bloopers that were mentioned earlier. If we look at a certain player's wOBA compared to his xwOBA we can find in certain times, there are interesting differences in these values. For example, if we look at the 2024 data from baseball savant that shows wOBA and xwOBA, a player like

Kyle Tucker is interesting to look at. Kyle Tucker had a wOBA value of .419 which is the fourth highest in the league (behind the sports best offensive players Judge, Soto, Ohtani). But when we look at his xwOBA value it is .400, which drops him down to eighth highest in the league. We thought this data was interesting and would be fun to investigate within our research later.

$$xwOBA = \frac{0.696 \cdot uBB + 0.726 \cdot HBP + \sum_{i=1}^n xwOBACON_i}{AB + BB - IBB + SF + HBP}$$

-
- **Why it matters:** xwOBA gives a clearer picture of a hitter's value because it takes into account the quality of the contact (how hard and where they hit the ball). Unlike simple stats like batting average, xwOBA tells us if a batter is hitting the ball hard and where it lands. The xwOBA statistic is so rich and has endless possibilities when conducting data. We believe it has enough potential to work with for a whole semester.
- **How it is calculated:** xwOBA uses batted ball data like exit velocity (how hard the ball was hit) and launch angle (the angle at which it leaves the bat) to predict what kind of hit the batter will get. These numbers are compared to historical data to estimate how often a certain batted ball results in a hit. We also found that there are a number of different equations that are used online for calculating xwOBA and it would be in our best effort to determine which one would be the best for our aim.
- **Relevance of Park Factors:**
 - **What are park factors:** Park factors adjust stats like xwOBA based on the characteristics of each baseball stadium (like how deep the outfield fences are, the

altitude, or wind patterns). Each park affects the game differently, meaning that some help hitters more than others.

- **Why they are important:** Adjusting for these factors helps us understand if a batter's performance is consistent no matter where they play or if it is significantly affected by the park they are in. Ballparks can change the entire outcome of a single game or season. For example, looking back to the Wild Card Game in 2021 between the Red Sox and Yankees there was a certain hit that would have been a home run in every other ballpark in the MLB. We are talking about Giancarlo Stanton's blast that hit the top of the monster with a massive exit velocity. This could have very well propelled the Yankees to win this game but instead it was an extraordinarily long single.
- **Project Aim:** The goal of this project is to calculate and analyze how xwOBA changes when a hitter plays in different stadiums, to see if their performance is better or worse in certain parks. It would also be fun to look at the leagues xwOBA at different ballparks, we think this is another way we could go with this project.

Methods

- **Statistical Methods:**
 - **Overview:** We can use certain methods to calculate metrics like xwOBA that give us a better picture of player performance. These methods include linear regression, KNN and Park Factors.
 - **xwOBA:** The formula for xwOBA considers the type of hit and how often those hits lead to a run, based on historical data. The formula adjusts these probabilities based on the type of contact and the stadium the batter is playing in.

- **Park Factors:** Park factors adjust stats like xwOBA by taking into account the unique features of each stadium, things like wind patterns, how deep the outfield fences are, and the altitude. This helps us understand if a batter's stats would be better or worse if they played in a different stadium. I think we could approach this by using play by play data in which we could determine the outcome of each at bat, within each ballpark.
- **Statistical Methods:**
 - **Linear Regression:** This method helps us see how a batter's xwOBA changes when we account for the features of the stadium they are playing in. We look at things like how deep the fences are and how often the ball flies out of the park.
 - **KNN:** Upon research we found KNN could be a factor when looking at this type of data which could be remarkably interesting to use come the springtime.
 - **Park Adjustments:**
 - Use formulas to adjust xwOBA based on stadium characteristics (example: adjusting for home runs in a smaller park).
 - These adjustments show us how much a batter's performance might change if they played in a different park. For example, Yankee stadium for left-handed hitters is a gold mine. How might this affect a player's xwOBA value?

R Packages

- **CRAN Packages for Baseball Analytics:**
 - **lahman:** This package is useful for accessing detailed play-by-play data in R. It shows things like batted ball data, which is important for xwOBA calculations.

- **baseballr**: This package includes functions for analyzing baseball data, from basic stats to more advanced metrics. We can use this to manipulate data to find xwOBA and park adjustments.

ggplot2: This is used for visualizing data. We will use it to create graphs that show how xwOBA changes across different parks.

Datasets

- **Primary Datasets:**

- **Statcast Data**: This dataset will be used for information on every play in baseball, such as exit velocity, launch angle, and the speed at which the ball was thrown.
 - <https://www.mlb.com/glossary/statcast/expected-woba>
- **Hitter Data**: This dataset will gather data on how often batters hit for power (singles, doubles, home runs)
 - <https://www.mlb.com/stats/>
- **Park Factors Data**: We found FanGraphs and Baseball Savant which gave data on things like wind speed, outfield fence height, and stadium dimensions.
 - **FanGraphs**: this site provides baseball data, like advanced metrics like park factors, detailed player statistics, and batted ball profiles, which is essential for analyzing xwOBA and understanding the influence of different stadiums on players performance.
 - <https://www.fangraphs.com/>
 - **Baseball Savant**: This site provides play-by-play data and park adjustments. It will be great for accessing detailed batted ball data and understanding how different stadiums affect player performance.
 - https://baseballsavant.mlb.com/leaderboard/expected_statistics

- **Weather data:** We will use historical weather data to allow us to examine how external conditions, such as humidity or wind speed, influence xwOBA calculations.
 - <https://www.ncei.noaa.gov/cdo-web/>
 - We understand that weather data is difficult to factor into this research, but we think it would provide a lot of intriguing insights on how weather can affect a player or even a team's offensive performance. We are thinking of warm weather teams such as the LA Dodgers playing in cold environments come the wintertime. Do their statistics differ when it comes to these different environments?

Research Questions:

- **Main Questions:**

- How does park-specific xwOBA vary across different stadiums? Does a batter's performance really change from one park to another?
- What impact do park factors have on a hitter's overall performance metrics? Do players hit better or worse in certain stadiums?
- Can Park adjustments significantly influence the analysis of player performance? How much does it really matter where a batter plays?

- **Our Approach:**

- We will use linear regression models to see how park factors affect xwOBA. By comparing data from different parks, we will identify which factors are most important.

- Compare adjusted xwOBA values across different parks to see how much a player's performance might change.
- **More Ideas:**
 - More complex factors: weather conditions (wind speed, humidity) or changes in stadium features (new walls, changes in park dimensions). How do these factors affect player performance?

Summary:

The overall goal is to explore how park-specific adjustments to xwOBA can give a clearer understanding of a hitter's performance. By accounting for the unique characteristics of each stadium, such as outfield dimensions, altitude, and wind patterns, we can provide a more accurate analysis of a player's true skill level. This approach helps remove the bias of park conditions and allows for a fair comparison between players across the league.

Here are some (but not all) of the websites we have used in our research so far:

FanGraphs community research :

<https://community.fangraphs.com/exploring-batter-xwoba-and-its-applications-part-1/>

Baseball Savant wOBA and xWOB data:

https://baseballsavant.mlb.com/leaderboard/expected_statistics?type=batter&year=2024&position=&team=&filterType=bip&min=q&sort=12&sortDir=desc

MLB Technology Blog:

<https://technology.mlbblogs.com/an-introduction-to-expected-weighted-on-base-average-xwoba-29d6070ba52b>

Stat Corner:

<https://statcorner.com/mlb/stats/woba>