# Consistency in Batting: An Analysis of Bat Speed and Swing Length in Major League Baseball

By Noah Tri, Manfred Tan

2025, January 15

**Abstract** This study analyzes bat speed and swing length to explore consistency in Major League Baseball batters during the 2024 season. We examine batters with consistent swings and seek to understand the conditions that significantly alter their swing. We define "consistent batters" as those with low interquartile ranges (IQR) for both bat speed and swing length, and “outlier swings” as a swing that is an IQR-defined outlier. Our analysis reveals that outlier swings are significantly influenced by factors such as ball-strike count, inning, and pitch type, with batters more likely to deviate from their typical swing when behind in the count or facing pitches like sliders. We found that a significant amount of outlier swings resulted in a much higher swinging strike rate—and much lower contact rate—than normal. More analysis must be done in the future to further explain the factors leading to outlier swings and to find ways to implement these factors in game plans.

## Introduction

Major League Baseball is one of the most well documented professional sports leagues in the world. It has pioneered the use of statistical analysis in professional sports and has one of the most extensive and accessible databases regarding baseball information and statistics. Each season, Baseball Savant collects comprehensive data on Major League Baseball plate appearances, utilizing Statcast tracking technology. For 2024, the Statcast tracking technology recorded data for 701,557 pitches from the 2024 MLB regular season and playoffs. This paper aims to explore newly released data regarding batter swing length and bat speed during the 2024 MLB season. In particular, the objective is to answer the following questions: If there are batters who are very consistent in their swinging, what kind of scenarios make them take extreme swings?

## Methodology

**Consistency in Batting** The first goal is to define consistency in batting and define what a consistent swing is. Given the variability in the nature of swinging, it is difficult to draw conclusions solely from a single variable. Each batter can have drastically different approaches and different swings depending on a multitude of factors. Instead of analyzing swing data as a whole, we group the data by batter in order to normalize the swing data before finding larger trends.

We will define swings using two main data points: bat speed (mph) and swing length, the distance (feet) the head of the bat traveled in X/Y/Z space. Theoretically, consistent batters should have a relatively low variability in both of these variables. Low variability means that most of their swings have similar swing length and speed. To accurately compare the variability between batters, an appropriate summary statistic is needed. Our aim is to find batters who are generally consistent but have some outliers, so the summary statistic should not be heavily skewed because of the outliers. Thus, the IQR (interquartile range) is the best statistic to compare bat speed and swing length across batters and define the batters with more consistent swings.

The Statcast tracking technology recorded at-bats for 651 batters during the 2024 regular season and playoffs, with a mean of 1078 pitches faced per batter and a median of 953 pitches faced. Taking only the pitches where bat speed and swing length were recorded (i.e. pitches where the batter swung), we found the interquartile range (IQR) for bat speed and swing length, respectively, for each batter. Batters who have consistent swings have lower variance in terms of bat speed and swing length. Ensuingly, we defined "consistent batters" as batters with a recorded IQR for bat speed in the lower 50% of all batters, in addition to a recorded IQR for swing length in the lower 50% of all batters. Although any subjective threshold number could have been used, our definition had 50% as a threshold because this means that the batter is technically better than the majority of batters in that metric. For our set of batters, the median for the IQR of bat speed was 6.1 MPH and the median for the IQR of swing length was 1.1 feet. Batters who had IQRs under both of these marks are considered consistent batters. Of our total of 651 batters, 136 fit this criteria. These 136 batters showed consistent behavior in terms of their swing. The wide majority of the 136 have an IQR for bat speed between 4 and 6.1 MPH and an IQR for swing length between 0.8 and 1.1 feet.

**Outlier Swings** From our consistent batter pool, we want to explore the "outlier" swings to analyze what factors—situationally and pitcher-wise—force a change in swing from our otherwise consistent batters. To define an "outlier swing", we first clean our data such that it only includes pitches where one of the 136 consistent batters swung. This reduced the original data set of 701,557 pitches down to 52,316 pitches. For simplicity and replicability, we defined an outlier swing from a consistent batter as a swing where the recorded bat speed and/or swing length was an IQR-defined outlier (less than 1.5 times the IQR below the first quartile or more than 1.5 times the IQR above the third quartile). These are swings that are significantly faster or slower or significantly longer or shorter. In essence, these are swings that deviate significantly from a consistent batter's typical swing.

Our consistent batters dataset has 3,301 bat swing outliers and 871 swing length outliers. Put together, outlier swings make up 3,415 swings from our 52,316 recorded swings for consistent batters, due to some expected overlap. Only 11% of the outliers are due to having a fast bat speed or high swing length. Meanwhile, 65% of the outliers have a significantly low bat speed only. 3% of outliers have a significantly low swing length only. The remaining 22% of outliers have a significantly low bat speed and a significantly low swing length.

**Potential Impactful Variables** From our newly defined dataset containing only batters with consistent swings, we look at relationships between outlier swings (i.e. outlier bat speeds and outlier swing lengths) and other variables that might impact batting consistency. Some of the possible variables may be game-situational: inning, number of outs, etc.. Other variables may be pitcher-influenced: pitch type and velocity and acceleration of the pitch in each dimension. We attempt to find what makes a batter more likely to deviate from his consistent swing. Also, we look at the relationship between outlier swings and the result of the swing.

**Analysis** The following variables were chosen to test against our hypothesis. Other variables may be significant, but may not have been chosen due to a variety of reasons, like a lack of clearly defined data points.

- Ball-Strike Count: the number of balls and strikes prior to the pitch.
- Pitch Number in At-Bat: the pitch number in the at-bat.
- Inning: the current inning of the at-bat.
- Outs: the number of outs prior to the pitch.
- Score Differential: the difference in score prior to the pitch, taken as the batter's team score minus the pitcher's team score. A positive number indicates the batter's team is ahead, a negative number indicates the pitcher's team is ahead, and 0 indicates a tie.
- Base States: the state of bases prior to the pitch. Formatted as X-Y-Z, where X,Y,Z are either 0 or 1, with 0 meaning no runner on base and 1 meaning runner on base. Formatted as first, second, and third base respectively.
- Pitch Name: the name of the pitch thrown.

Each of these variables were used as a test statistic to test against consistent and outlier swings. Thus, a Chi Square test is used on each of these variables. This is done to compare the frequencies between consistent and outlier swings assuming the null hypothesis is true. The null hypothesis is that there is no significant association between consistent and outlier swings and the given variable. For example, the inning variable looks at the frequency of consistent and outlier swings across all innings. A p-value below the threshold of 0.05 indicates that we reject the null hypothesis and that consistent/outlier swings are significantly associated with the inning.

## Results

The results of the Chi Square tests are tabulated below (Figure 1).

**TABLE 1.** *Results of Chi-Square Tests for Independence*

| *Variable* | *p-value* | *Result* |
|---|---|---|
| Ball-Strike Count | <2.2e-16 | Significant |
| Pitch Number in At-Bat | 0.05432 | Not Significant |
| Inning | 0.003521 | Significant |
| Outs | 0.3361 | Not Significant |
| Score Differential | 0.9022 | Not Significant |
| Base States | 0.2289 | Not Significant |
| Pitch Type (Name) | <2.2e-16 | Significant |

*Notes:* The threshold is 0.05. Individual Chi Square test results can be found in the attached code file.

**Ball-Strike Count** In conducting the Chi Square test for independence, it becomes apparent that counts considered "hitters counts" (3-0, 3-1, 2-1, 2-0, 1-0) are underrepresented in the outlier swings. The percent of the total population of pitches for each of the hitters counts is greater than the percent of the total outlier percent of pitches by at least 0.75%, with a 3-0 count being underrepresented by over 2%. On the flip side, each count considered "pitcher count" (0-1, 0-2, 1-2), except for the count 2-2, is overrepresented in outlier pitches by at least 1.75%. From the Chi Square test for independence, we reject the hypothesis that ball-strike count and outlier swings are independent of each other. Batters tend to be more likely to have an outlier swing when they are behind in the count, i.e. when the pitcher has "control" of the at-bat. Figure 2 shows the difference between the observed and expected outlier swings by number of strikes in the count. Counts with two strikes account for significantly more outlier swings than counts with zero and one strikes due to the observed number of outlier swings in 3-0, 3-1, 2-1, and 2-0 counts being significantly lower than expected and all two strike counts being pitcher counts.

While pitch number is not statistically significant with a 95% confidence interval, the relatively low p-value may be explained by the correlation with the ball-strike

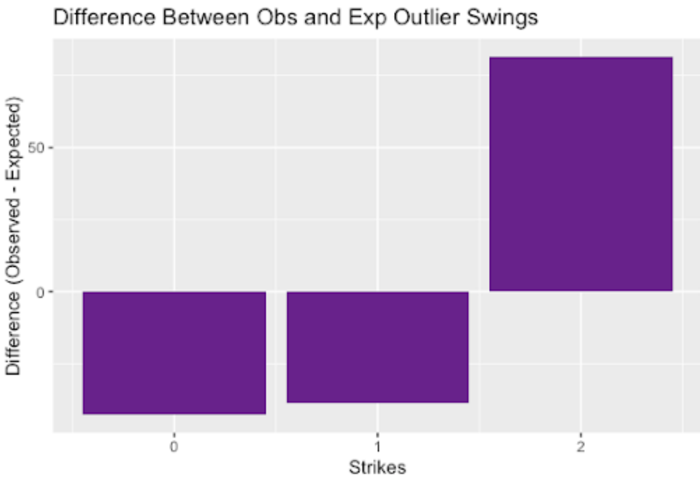


**FIGURE 2.** *Difference Between Observed and Expected Outlier Swings by Strikes*

count and the ball-strike counts statistical significance. The only pitch number overrespresented in the outlier data is the first pitch of the at-bat, which stems from the fact that 0-0 is typically considered a pitcher's count. The distribution for all the other pitch numbers for outlier swings is close to the population distribution, possibly resulting from the fact that the ball-strike count isn't known from just the pitch number of the at-bat after the first pitch.

**Inning** The inning is a significant variable for consistent and outlier swings. While the distribution of overall swings from consistent batters was roughly uniform across innings, the distribution for outlier swings across innings was not. Figure 3 highlights the difference between observed versus expected outlier swings based on the results of the Chi Square test. Note that the data included swings past the ninth inning, up to the fourteenth inning, but these extra innings were removed for the integrity of the Chi Square test due to lack of sufficient counts for the extra innings. The significant innings seem to be first, third, seventh, eighth, and ninth. The data shows that there are less outlier swings in the first innings than expected, and more outlier swings in the last three innings than expected. This follows the intuitive idea that batters are more disciplined during earlier at-bats and are less likely to have wilder swings, while batters are less disciplined or perhaps chasing score deficits during the later innings.

**Pitch Type** The pitch type, referred to by its name, is also a significant variable. The results of the Chi Square test indicate that the distribution of overall swings from consistent batters is significantly different to the distribution of outlier swings from

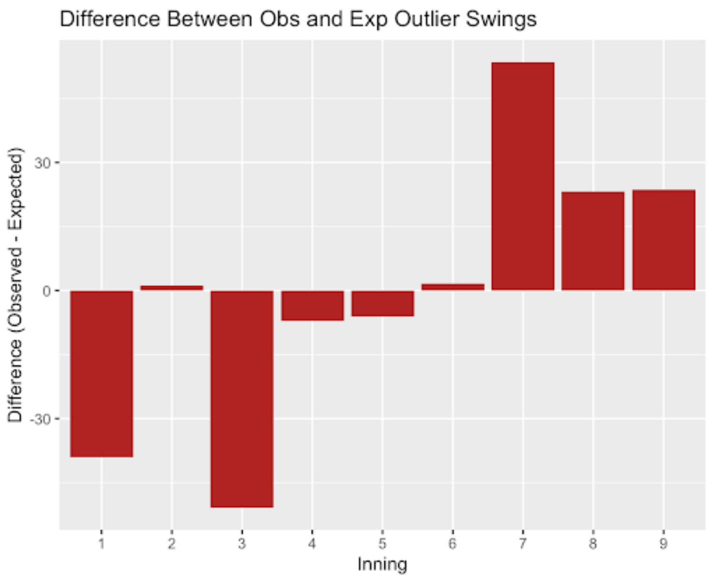


**FIGURE 3.** *Difference Between Observed and Expected Outlier Swings for each Inning*

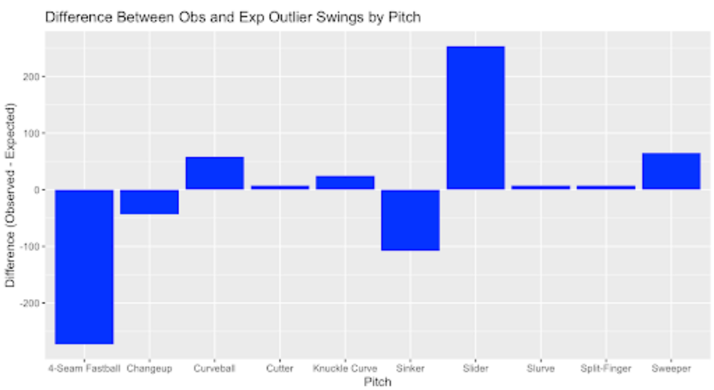


**FIGURE 4.** *Difference Between Observed and Expected Outlier Swings by Pitch*

consistent batters based on the pitch type. Figure 4 shows the difference between the number of expected versus observed outlier swings based on the pitch type. There are significantly less outlier swings for fastball pitches, and significantly more for slider pitches. This is a reasonable outcome, as a fastball is a high speed pitch that is difficult to react to, while a slider is a breaking pitch that relies on deceptive movement. Theoretically, a batter would approach a fastball with their standard swing while adjusting the timing, but they would need to change their bat speed or swing angle to make contact with a slider. This variation would therefore numerically classify as an outlier swing.

**Swing Result** After covering the relationship between outlier swings and some possible variables, we finish by examining the result of these outlier swings compared to our whole population of swings for the consistent batters. In particular, we focus on the proportion of swings that resulted in a swinging strike, a foul, or a ball hit into play. Taking into account the Chi Square tests done above—the ball-strike count in particular—, we want to test if outlier swings are employed to adapt to an unexpected pitch to foul the ball off or put it into play. Thus, we performed one-sided binomial test of proportions for each of our three main swing results, using the proportions of these three results from our whole population as the test proportion; the null hypothesis for the swinging strike binomial test is the proportion of outlier swings resulting in swinging strikes is less than or equal to the population proportion while the null hypothesis for the other two is the proportion of outlier swings for each of them is greater than or equal to their population proportions. The test results are shown in Figure 5.

**TABLE 5.** *Results for the Binomial Test of Distributions on Outlier Swing Results*

| *Swing Result* | *Outlier Swing Proportion* | $H_0$ | *P-Value* | *Result* |
|---|---|---|---|---|
| Swinging Strike | 0.450 | $p \leq 0.235$ | $< 2.2e-16$ | Reject $H_0$ |
| Foul | 0.340 | $p \geq 0.399$ | $< 0.00001$ | Reject $H_0$ |
| Hit into Play | 0.210 | $p \geq 0.366$ | $< 2.2e-16$ | Reject $H_0$ |

*Notes:* The threshold is 0.05. Individual Chi Square test results can be found in the attached code file.

The test results go against our initial hypothesis. The proportion of swinging strikes for outlier swings is actually greater than the population proportion of swinging strikes. Inversely, the proportion of outlier swings resulting in making contact with the ball (foul and hit into play) is less than the population proportion of swings. This tells us that batters significantly changing their swing don't seem to find success with fouling the ball off and staying alive in the at-bat. Outlier swings are less likely to make contact with the ball. We hypothesize that outlier swings are a result of successful deception by the pitcher, but innovative data collection and more analysis must be done in the future to test this hypothesis.

## Conclusion

While this analysis led to certain insights into outlier swings, it is important to note possible shortcomings of the methods used in the analysis. An aspect that may or may not have an impact on the swing analysis is the use of bunts. A bunt would classify as an outlier swing, which could theoretically impact the findings. Based on the data, they were not deemed impactful enough to be filtered out, as they only made up less than 0.05% of all swings. Nonetheless, it may be useful to further break down swings to account for bunting. Additionally, certain batters might inherently have more variability in their swing profiles due to their position (leadoff, cleanup, etc.). Further nuance is required before specific conclusions can be drawn for individual batters. In spite of this, the analysis of swing consistency shows promising potential for future exploration. More in depth analysis between pitching variables and swing consistency could yield useful information on how to approach certain batters as a pitcher, and vice versa.