Investigating the Presence of Mechanical Deviations in Baseball Swings Over Time

Khushi Shah and Yunus Mouline

The University of North Carolina at Chapel Hill

January 15th, 2025

Abstract

In professional baseball, every swing reflects a decision influenced by pitch characteristics, game context, and a batter's intent. However, deviations from expected swing mechanics may indicate a breakdown in consistency throughout a game. This study investigates the presence and progression of deviations from expected swing mechanics throughout a game, focusing on two key metrics: swing length residuals and the relationship between bat speed and launch speed. Using pitch-level data from Major League Baseball, we apply regression models and residual analysis to quantify these deviations. A logistic regression indicated a 26.4% increase in unintended swing likelihood per additional swing in an inning ($p < 0.001$). Swing length analysis revealed increasing mechanical variability, with mean squared error increasing significantly with swing count (Estimate: 0.00176, $p = 0.00467$). These trends suggest a pattern of mechanical inconsistency, potentially linked to fatigue or cognitive strain. This research provides a foundation for further exploration into batter performance trends and offers insights for coaching strategies aimed at injury prevention and fatigue management.

1 Introduction

A batter's swing is a complex biomechanical action influenced by pitch characteristics, game context, and player intent. Maintaining consistent swing mechanics is crucial for performance, as deviations can affect contact quality and overall success at the plate. However, as a game progresses, physical fatigue or cognitive strain may lead to subtle yet impactful inconsistencies in swing mechanics, raising questions about how batter performance shifts over time.

This study investigates how deviations from expected swing mechanics evolve throughout a game, focusing on two key metrics: swing length residuals and the relationship between bat speed and launch speed. By modeling expected swing length and launch speed based on pitch characteristics and bat speed, respectively, we aim to measure deviations from intent and assess whether these deviations become more pronounced as the game progresses.

Our objective is to provide a data-driven analysis of batter performance by quantifying swing length residuals and the bat speed-launch speed relationship, offering insights into mechanical consistency. Identifying when a batter's mechanics begin to break down can inform coaching strategies for in-game adjustments, player fatigue management, and injury prevention. Beyond baseball, this research contributes to a broader understanding of skill variability under physical and cognitive strain across sports and physical activities.

2 Methodology

Research Problem and Objectives

The primary research question guiding this analysis is: *How do deviations from intended swing mechanics change throughout the game, and what factors might drive these changes?* Specifically, we aim to determine whether mechanical deviations increase as the game progresses by analyzing both swing length residuals and the correlation between bat speed and launch speed. Swing length residuals measure the difference between observed and predicted swing length, while bat speed variations are compared against launch speed as an indicator of mechanical consistency.

Our hypothesis is that both metrics will show increasing variance over time, suggesting a potential link to fatigue or reduced mechanical precision. This hypothesis is tested using regression models and residual analyses to measure deviations across game stages.

Relevance and Importance

Understanding the relationship between swing mechanics and game progression is crucial for player performance analysis, coaching strategies, and injury prevention. Consistency in swing length and bat speed directly affects contact quality, with mechanical deviations potentially leading to reduced success at the plate.

Identifying patterns of increased deviations can help coaches make informed decisions, such as adjusting player workloads or modifying training regimens to improve late-game endurance. Additionally, monitoring mechanical breakdown could assist in the early detection of injury risk, as fatigue-driven inconsistencies often correlate with muscle strain and overuse injuries.

Data Processing

The dataset used for our analysis contains performance data from 651 unique batters. To ensure robust and reliable results, we implemented a series of data processing steps to prepare the dataset for modeling and analysis. Initially, the data showed a wide range of observations per batter, with an average of 1078 observations per batter and a median of 953. The distribution was highly variable, with a minimum of just 1 observation and a maximum of 3126. To reduce potential bias from underrepresented batters and ensure a sufficient sample size for statistical modeling, we restricted our analysis to batters with a minimum of 1000 observations. This threshold retained a more stable subset of 315 batters while discarding those with limited data points that could distort the models.

Next, we filtered the dataset to focus only on instances where the batter swung at the pitch. Since our goal was to model expected bat speed and swing length, non-swinging pitches would not provide meaningful information for the analysis. Removing these observations allowed us to focus exclusively on the mechanics and outcomes of swings, enhancing the clarity and relevance of our models.

Finally, we refined the dataset by excluding variables that were not relevant to our analysis. The final set of variables retained for analysis included contextual variables such as game date (game_date), batter identifier (batter), inning (inning), balls count (ball_count), and strikes count (strike_count). Additionally, variables required for the bat speed and swing length analyses were also included, such as launch speed (launch_speed), bat speed (bat_speed), swing length (swing_length), and all relevant pitch characteristics, including vertical pitch location (plate_z), horizontal pitch location (plate_x), release speed, pitch type, release spin rate, arm angle,release position (release_position_z and release_position_y), release extension, pfx_x, pfx_z, and spin axis. These variables were chosen based on their potential influence on mechanical consistency throughout the game.

Analysis Methods

To investigate how deviations from batting intent change over time, we developed a two-part methodology focused on modeling intended launch speed and swing length.

<u>Launch Speed:</u>
Defining intent required isolating factors controlled by the batter while minimizing external influences. Launch speed appears to have constant mean and spread throughout the game (see Appendix for plot). However, the assumption that this reflects a constant intended launch speed is flawed, as intended launch speed can vary significantly based on game context, such as pitch type, game pressure, and strategic considerations. Therefore, we focused on bat speed, a more direct indicator of batter control and intent, as it reflects the batter's effort and desired outcome in the swing.

We began by examining the correlation between bat speed and swing length using a Pearson correlation test, which provided insight into their relationship and helped us justify the use of bat speed in modeling intent. Based on this analysis, we developed a model to predict launch speed using bat speed as the single predictor. This model generated predicted launch speed values, which serve as a proxy for the batter's intended launch speed and was added to our dataset for further analysis.
To measure deviations from intent, we calculated the residuals between the actual and predicted launch speeds using the formula: **Residual = Actual Launch Speed - Predicted Launch Speed.**

Observations falling below the lower bound of the 95% confidence interval for predicted launch speed were identified as unintended swings. These points, as shown in red in **Figure 4**, represent mishits where the batter likely failed to achieve the desired contact quality. Notably, it was less common for launch speed to significantly exceed predicted values since players generally aim to maximize exit velocity.

With the deviation metric established, we proceeded to analyze how the proportion of unintended swings fluctuates over time within a game and inning. Since our dataset already captured swing-level data, we grouped observations by player, game, and inning while numbering each swing event to reflect the order of swings taken by a batter in a given inning. This process allowed us to track how deviations from intent evolved during a player's appearance at the plate.

To evaluate the trend, we created a binomial plot showing the proportion of unintended swings relative to total swings over the course of a game and inning. To model this relationship statistically, we applied logistic regression, which provided a fitted curve for the probability of an unintended swing occurring as time progressed. To assess the significance of the observed trends, we used a likelihood ratio test and confidence intervals for coefficients, ensuring the results were statistically significant.

<u>Swing Length:</u>
In addition to bat speed, this study employed a structured analytical approach to investigate deviations in swing length throughout a baseball game. A multiple linear regression model was developed using pitch-level data, including vertical pitch location, release speed, pitch type, horizontal pitch location, spin rate, and arm angle. A stepwise regression approach refined the model, retaining only significant predictors while minimizing overfitting.

Residuals (the difference between observed and predicted swing length) were calculated to measure deviation from expected mechanics, with Mean Squared Error (MSE) being computed as the measure for deviation magnitude. To assess whether these deviations changed throughout the game, swings were ordered chronologically by swing number within each game, and the proportion of significant deviations and average MSE were calculated per swing. Linear regression models were applied to test whether deviations increased over time.

To visualize these trends, **Figure 1** shows pitch locations overlaid with swing length quantiles, while **Figures 2 and 3** display the relationship between MSE and swing number, with Figure 3 including a linear regression line and a 95% confidence region. This combined approach of modeling, residual analysis, and visualization provides insights into batter performance variability across a game.

3 Results

<u>Launch Speed</u>

Our analysis examined the relationship between bat speed and launch speed to identify deviations from expected performance as the game progressed. A Pearson correlation test revealed a correlation coefficient of **0.34**, indicating a positive relationship between bat speed and launch speed, suggesting that higher bat speeds generally result in higher launch speeds due to efficient energy transfer during the swing.

To assess whether this relationship deteriorates over time, we developed a logistic regression model predicting the likelihood of an unintended swing, defined as abnormal launch speed, based on the number of swings taken in an inning. The model showed a statistically significant effect, with a p-value of $\mathbf{< 1e{-}6}$ and an odds ratio of **1.264**, indicating that for each additional swing in an inning, the likelihood of an unintended outcome increases by **26.4%**. Residual analysis further confirmed this trend, with the proportion of unintended swings increasing from **1.52%** at the start of an inning to **4.05%** at the 8th swing in the inning. See **Table A** for all proportions per inning.

A binomial plot, as seen in **Figure 5**, and the fitted logistic regression curve visually supported the observed trend, showing a gradual rise in the probability of unintended swings with increasing swing count. While the effect size remained **modest**, the statistically significant results suggest that even professional athletes experience subtle declines in mechanical efficiency during extended at-bats.

<u>Swing Length</u>

The swing length analysis revealed significant findings regarding mechanical deviations with each passing swing. The multiple linear regression model developed for swing length yielded an Adjusted R^2 of **0.4352**, with predictors such as plate_z, release speed, and various pitch types significantly influencing swing length. Lower pitch locations were associated with longer swings, as shown in Figure 1, which maps pitch locations with a color gradient representing swing length quantiles.

Residuals from the model were used to calculate the average mean squared error (MSE) across swing sequences. **Figure 2** and **Table B** illustrate how MSE generally increased with swing number, suggesting growing mechanical variability over time. A linear regression confirmed this trend, as swing number significantly predicted MSE with a p-value of **0.00467**. **Figure 3** reinforces this upward trend, displaying a regression line and a 95% confidence interval, with an Adjusted R^2 of **0.4893**.

These findings suggest a measurable increase in mechanical inconsistency as the game progresses, supporting the hypothesis that fatigue or cognitive strain may contribute to declining swing stability. This emphasizes the importance of monitoring swing mechanics for both performance optimization and fatigue management.

4 Discussion

The findings of this study provide valuable insights into how deviations from expected swing mechanics evolve throughout a baseball game. Our primary objective was to investigate whether mechanical inconsistencies, measured through both swing length deviations and the relationship between bat speed and launch speed, increase over time. To address this question, we developed two complementary models: one examining swing length deviations and the other focusing on the relationship between bat speed and launch speed, with both serving as proxies for intent, and with deviations from predicted values serving as indicators of unintentional swings.

The results revealed statistically significant positive trends, with both the average mean squared error (MSE) for swing length and the correlation between bat speed and launch speed deteriorating as the number of swings increased. The data showed a measurable decline in mechanical consistency, aligning with the hypothesis that physical fatigue or cognitive strain accumulates as the game progresses. However, the effect sizes were modest, suggesting that while mechanical inconsistencies do emerge, professional athletes maintain a high degree of control even under physical stress.

Our data processing methods played a key role in ensuring the robustness of these findings. The dataset was filtered to include only batters with a minimum of 1000 observations and swings where the batter actively engaged with the pitch, helping isolate batter-driven variability and minimize external influences such as pitch strategy.

The positive relationship between swing order and mechanical deviations supports the theory that fatigue affects swing mechanics. As shown in **Figures 2 and 3**, both swing length deviations and declining bat speed correlations suggest reduced mechanical efficiency as the game progresses.

Comparison with Previous Studies

This research builds on prior studies examining fatigue in sports performance but focuses more directly on mechanical variability rather than outcome-based metrics like batting average or exit velocity. While previous studies have linked fatigue to scoring efficiency or movement accuracy in other sports (e.g., basketball shooting accuracy or swimming stroke efficiency), our focus on swing mechanics offers a more precise measure of physical fatigue. Using residual-based metrics and logistic regression, this study emphasizes the importance of monitoring underlying mechanics rather than performance outcomes alone.

Practical Applications

The findings of this study have several important implications for both short-term performance management and long-term player development. Monitoring swing length deviations and launch speed inconsistencies could serve as early indicators of fatigue, helping coaches make real-time adjustments such as strategic substitutions or targeted feedback on swing adjustments. Additionally, training regimens could be adjusted to focus on improving late-game endurance and mechanical consistency. Identifying mechanical breakdown patterns may also aid in injury prevention, allowing teams to better monitor workload and reduce strain-related injuries by intervening when fatigue begins affecting mechanics.

Limitations and Future Directions

While the findings offer valuable insights, this study has some limitations. Restricting the dataset to batters with a minimum of 1000 observations limits the generalizability of results to part-time or developing players. Additionally, the residual-based approach assumes that deviations from predicted launch speed directly reflect fatigue, which may not fully capture psychological factors or strategic adjustments during a game.

Future research could expand these findings by analyzing a more diverse range of players and examining additional external factors. For example, incorporating biomechanical data like bat path deviations or wrist mechanics could further clarify the physical effects of fatigue on swing consistency.

5 Conclusion:

This study provides evidence that deviations in both swing length and the relationship between bat speed and launch speed increase as the game progresses, suggesting a gradual decline in mechanical stability over time. While the magnitude of these deviations was modest, the statistically significant trends emphasize the importance of monitoring swing mechanics alongside traditional performance metrics. By focusing on mechanical efficiency rather than just game outcomes, this research offers a deeper understanding of how fatigue influences professional baseball performance. These insights can be valuable for optimizing training strategies, managing in-game fatigue, and reducing injury risk, contributing to a more well-rounded approach to performance management in professional baseball.

6 Appendix:

Figure 1: Heatmap of Swing Length Quintiles by Plate Location

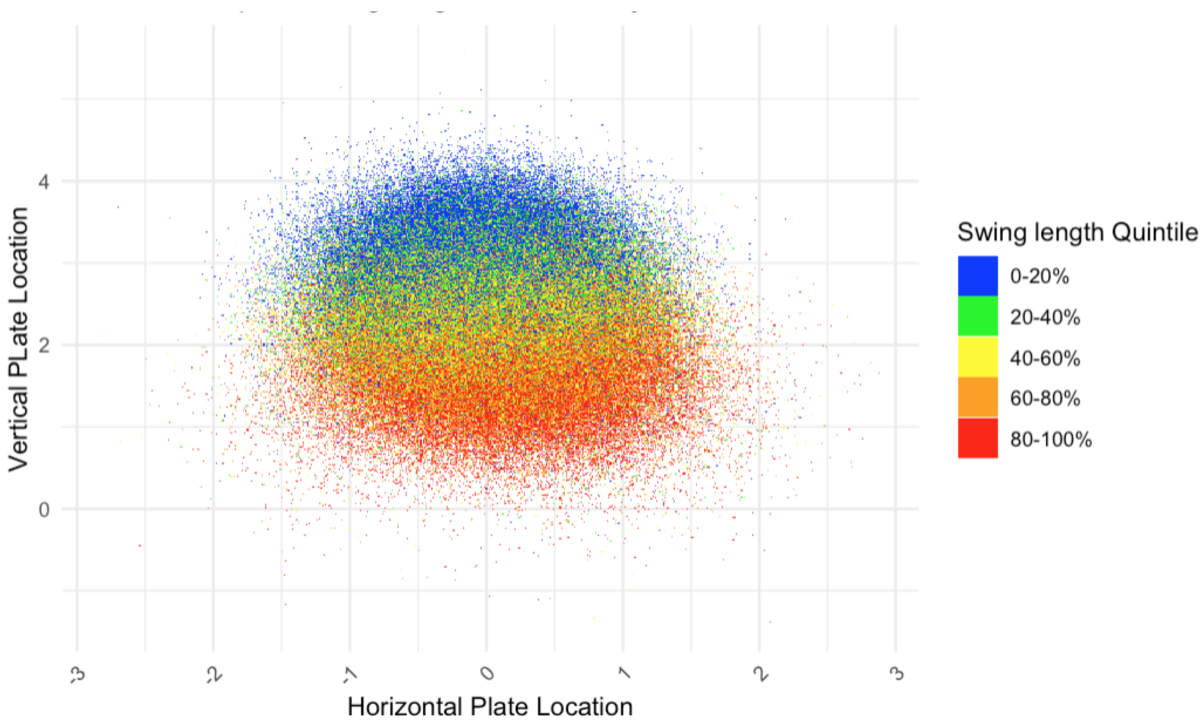
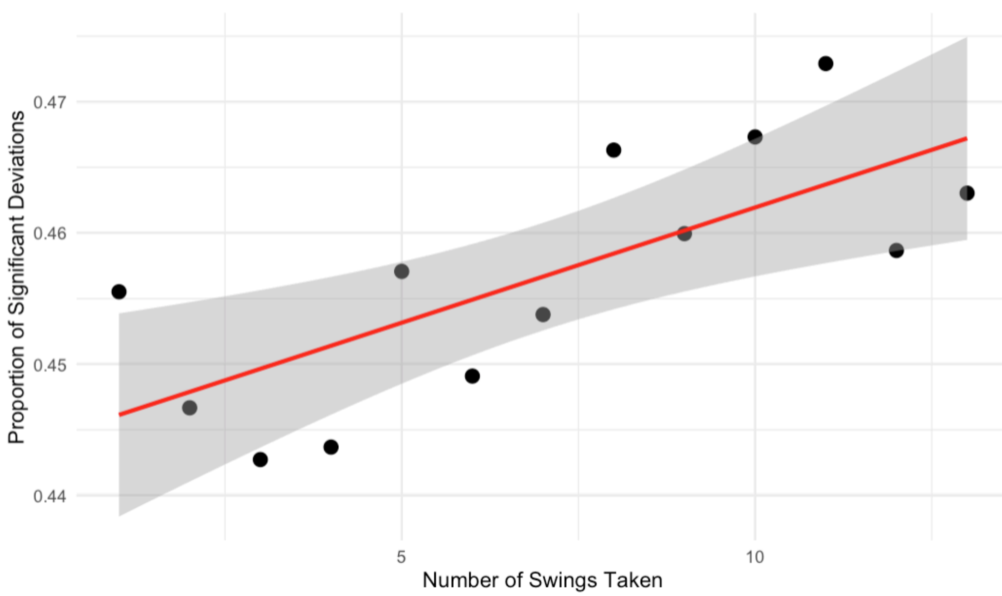


Chart of overall pitch locations from baseball games in the 2024 season, overlaid with a color gradient representing swing length quantile.

Figure 2: Swing Length Deviations by Swing Number based on MSE

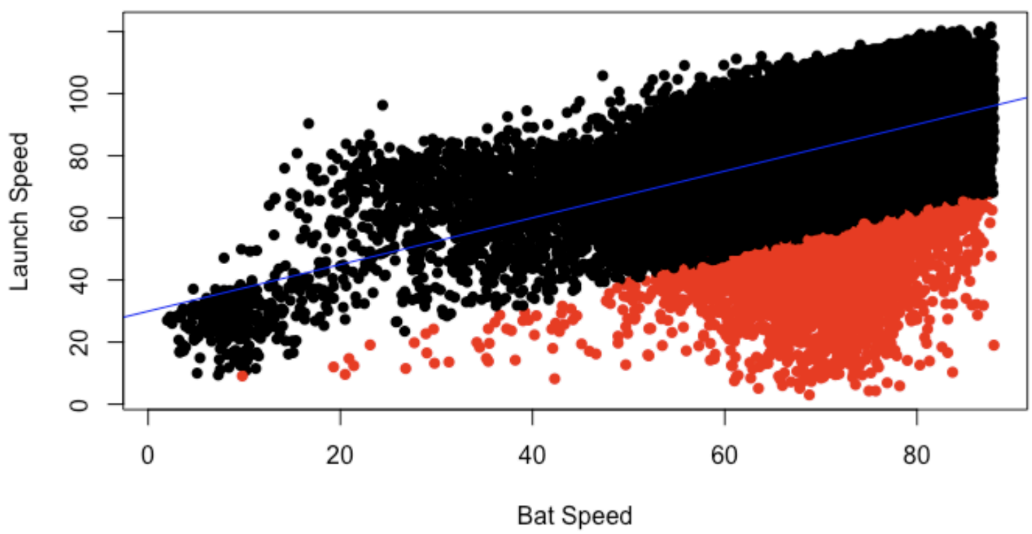
Relationship between average mean squared error from the swing length model and the number of swings a player has taken in a game.

Figure 3: Trend of Proportion of Significant Deviations Over Swings



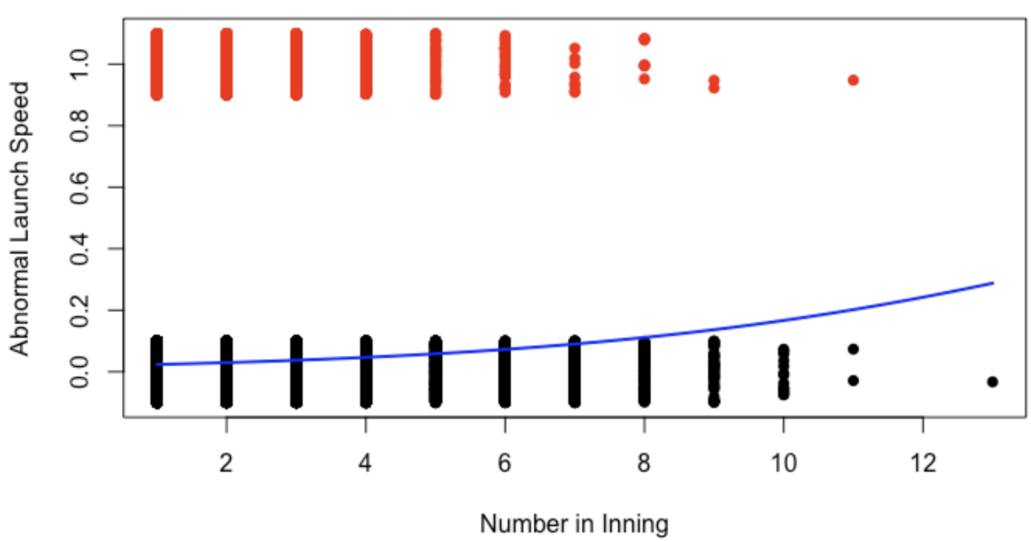
Relationship depicted in Figure 2, overlaid with the best-fit line and 95% confidence region.

Figure 4: Launch Speed vs. Bat Speed



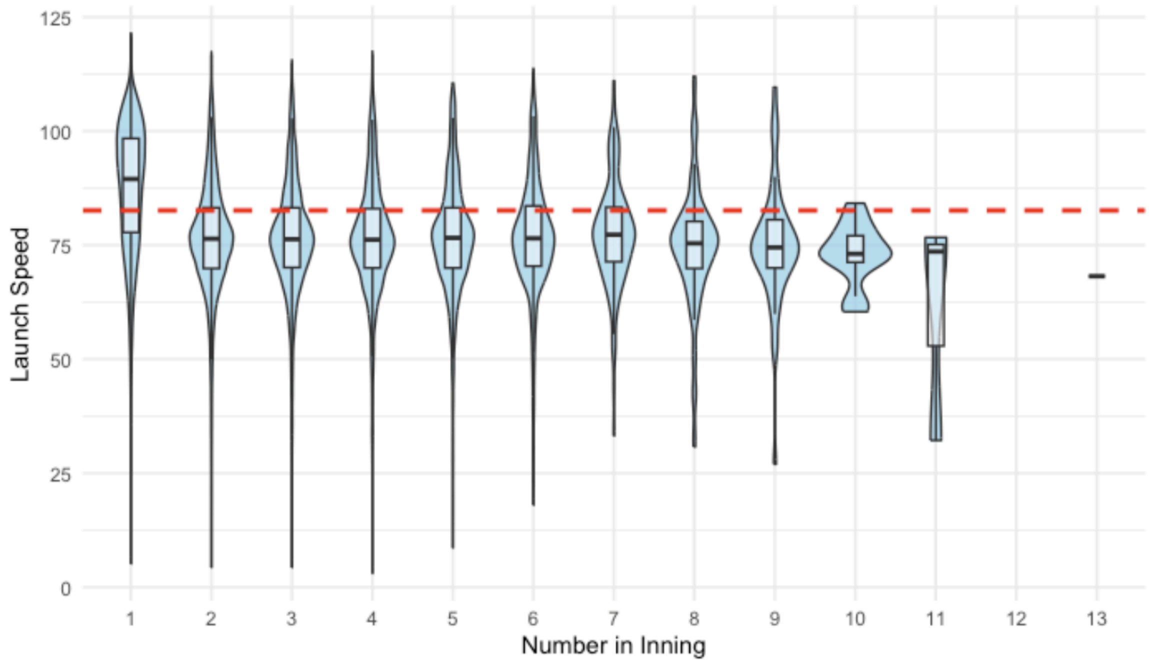
Relationship between bat speed and launch speed, highlighting swings outside the 95% confidence interval in red for clearer identification of subset.

Figure 5: Abnormal Launch Speed vs. Number in Inning



Odds of abnormal launch speeds in swings throughout an inning.

Figure 6: Launch Speed vs. Number in Inning



Average launch speed per swing for each batter throughout an inning.

Table A: Proportion of unintended launch speeds in swings throughout an inning.

swing in inning	mishits	total	proportion
1	2075	136580	0.0151925611363304
2	1990	76288	0.026085360738255
3	840	34174	0.0245800901269971
4	279	11621	0.0240082609069787
5	97	3957	0.0245135203436947
6	34	1388	0.0244956772334294
7	8	474	0.0168776371308017
8	7	173	0.0404624277456647

Table B: Deviations from predicted swing length in swings throughout an inning.

	swing number	total pitches	average mse
1	1	46984	0.455514256474769
2	2	45274	0.446669268025544
3	3	43045	0.442724026145851
4	4	39908	0.443674017273095
5	5	35405	0.457066797185625
6	6	29477	0.449087086120882
7	7	22763	0.453775972466305
8	8	16417	0.466312309904773
9	9	11190	0.459944170530311
10	10	7215	0.467317978002636
11	11	4490	0.472907647744553
12	12	2703	0.45866105513189
13	13	1623	0.463032238608496