Analysis of the Potential Implementation of an Automatic Strike Zone

Devon Goetz, Zack Kopstein, Simran Pabla, Claire Wichman

Project Overview

The MLB currently possesses the ability to implement an Automatic Strike Zone, eliminating the need for an umpire to be calling balls and strikes. Before using it, however, it is important to know how this addition might change the game.

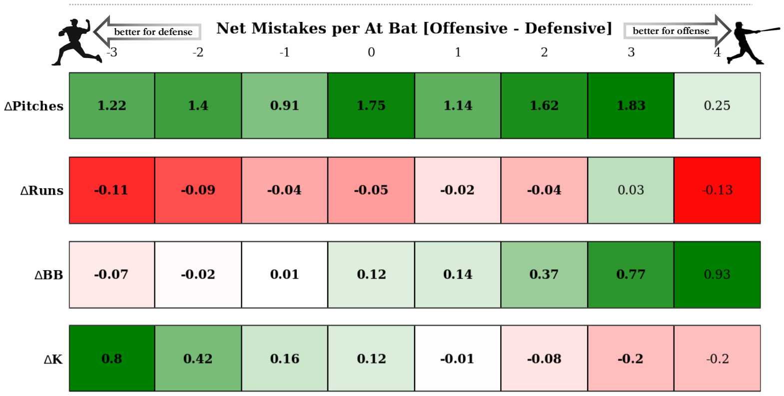
The main objective is to quantify the impact of the automatic strike zone. In order to accomplish this task, our team analyzed the difference between existing innings, grouped by their mistake characteristics, and created a predictive model to allow us to analyze how a game, if called correctly, may have ended differently.

Methodology & Approach

I. Existing Discrepancies: After determining that out metrics of focus would be **strikeouts**, **walks**, **runs scored**, and **pitches thrown**, we grouped at bats by the number of offensive and defensive mistakes they contained. **Offensive mistakes** (true strikes that had been called balls) are advantageous for the batting team, or while **defensive mistakes** (true balls that have been called strikes) aid the fielding team. This breakout allowed us to compare the average number of each metric with the average number in a perfect, mistake-free at bats in order to quantify the impact

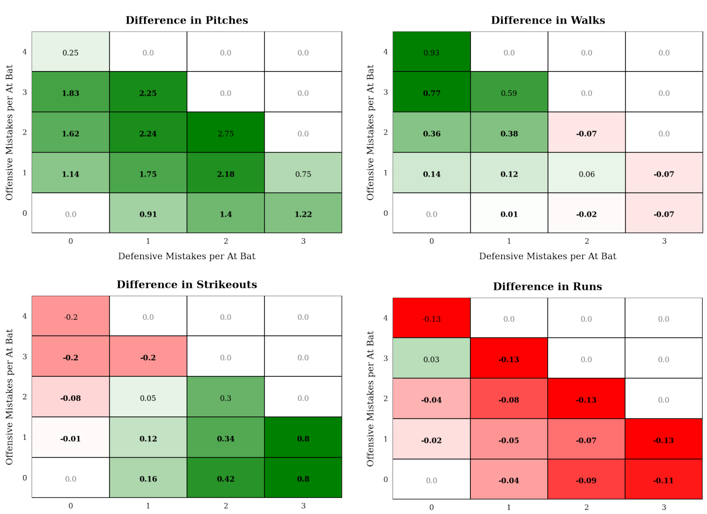
II. Predicted Changes: We created a predictive model that steps through each pitch. When it finds an umpire error, it corrects the state (ex. if the third pitch took the count to 2-1, but it was an incorrectly called ball, the new count would be 1-2). From there, it predicts the next most likely state for the game to enter, and so on, until we can begin referencing reality again. The model returns a full, error-free game. We compared our "corrected" games with the original versions to determine the difference the mistakes made in the metric outcomes.

Results: Existing Discrepancies



Note: Δ values calculated as [X Mistake Inning Average - 0 Mistake Inning Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

For each at bat, we took the net mistakes (offensive minus defensive) in order to get a broad strokes picture of the impact. Strikeouts and walks had clear and opposite trends, while runs and pitches had weaker patterns.
In order to increase granularity, we then classified at bats by the combination of offensive and defensive mistakes they contained, so we could get a clearer picture of how each type of mistake was changing the outcomes.



Results: Predicted Changes

The following represent extremely preliminary results. We have yet to incorporate offensive mistake corrections into our model, and are still running into some bugs in our final tabulations. These are more to represent placeholders than anything else.

The original innings typically had more strikeouts, and when we corrected defensive mistakes, we see a significant decrease in how many strikeouts are occurring per inning. [There will be more analysis about other metrics here when our model is complete.]

Conclusions & Moving Forward

[At this point, we will make a conclusion about our predictive model and what it can offer.]
Moving forward, this type of analysis could be used to examine how we can manipulate the strike zone, and how game outcomes might change. That research might be limited, as a new shape would drastically change batter and pitcher behavior, but for non-hit pitches it could serve an interesting way to begin that line of investigation.