Assessing MLB Umpire Accuracy

Michael Hessler

Department of Mathematical Sciences
Montana State University

May 8, 2024

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Michael Hessler

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Ian Laga Writing Project Advisor
Date	Katie Banner Writing Projects Coordinator

Abstract

Major League Baseball (MLB), like all professional sports, has officials to make calls and enforce the rules of the game. MLB on-field officials are called umpires. There's one umpire that holds more influence than any other, the home plate umpire. Their job is to determine whether each pitch thrown, if not swung at by the batter, is located within the strike zone. The primary goal of this paper is to determine which factors influence an umpire's accuracy when calling balls and strikes. In this paper, we will first give a brief overview of the rules of baseball, a history of the strike zone, discuss advancements in pitch-tracking technology, give an overview of the data used for our analysis, detail the methods of our analysis, discuss which factors influence an umpire's accuracy when calling balls and strikes, and share future improvements to the working model.

Contents

1 Introduction

In Major League Baseball (MLB), the home plate umpire is subject to a lot of scrutiny, perhaps more than any other individual on the field. The home plate umpire needs to determine if every pitch that a batter doesn't swing at is a ball or a strike based on its location. One can imagine why an umpire would be subject to such scrutiny given how substantial of an impact they can have on the outcome of a game. The addition of strike zone boxes on television screens in recent years have made things even tougher on umpires, as fans at home now have the ability to know instantly whether a pitch was truly a strike or not, a luxury umpires don't have. Umpires need to call pitches live from behind home plate without the aid of any technology. Strike zone boxes are electronic boxes that are put on your television screen that show whether a pitch was a strike as it crosses home plate.

In this paper, our primary goal is to determine which factors influence an umpire's accuracy when calling balls and strikes. This paper will discuss the following: the rules of baseball, the history of the strike zone, advancements in pitch-tracking technology, data used for our analysis, the methods used for modeling our data, which factors influence an umpire's accuracy when calling balls and strikes, and future improvements to the working model.

2 Rules of Baseball

In baseball, the goal of the hitter (also referred to as a batter) is to reach all four stations on the field (bases) to score points for your team. The fourth station (home plate) is where the batter stands. The batter's goal is to go around all the bases and get back to home plate to score a point (called a run) for your team. The objective of the pitcher and his teammates in the field is to prevent that from happening. One way to reach base is to hit the ball and get to a base without the ball being caught in the air or beating you to the base. Not every time a batter comes up to hit ends with the ball being hit. This is where the home plate umpire comes in. A batter gets three strikes to hit the ball or else they are out. A ball thrown from the pitcher to the batter (called a pitch) that crosses an invisible plane over home plate at the correct height, is called a strike. The home plate umpire's job is to determine whether the pitch crossed through that invisible plane or not. If a pitch doesn't go through that plane, the pitch is a ball. If a pitcher fails to throw three strikes to a batter and the batter doesn't hit the ball before a pitcher throws four balls, the batter gets to go to first base. With how important ball and strike calls are towards getting on base and scoring runs, one can see how influential the home plate umpire is towards the outcome of a game.

3 The Strike Zone

The strike zone is quite a bit more complicated than the casual observer would think. In fact, the official top and bottom of the strike zone has seen many changes throughout the history of the game. In 1876 (the beginning of the organization that would come to be known as the MLB), batters were allowed to tell a pitcher where they wanted the pitch to be thrown. Balls and strikes would be determined based on whether the pitcher threw the ball where the batter requested (MLB, 2024b). The rule allowing batters to request pitch locations was changed in 1887, giving rise to a strike zone similar to what we see today. The strike zone was defined as being from the shoulders to the knees at its inception. In 1950, the top of the strike zone was lowered slightly to the batter's armpits. In 1963, the top of the zone went back to the shoulders before once again returning to the armpits in 1969. Also in 1969, the bottom of the strike zone was defined as the top of the batter's knees. Finally in 1988, the top of the strike zone adjusted to its current level, and in 1996, the bottom of the zone had its final adjustment. According to the MLB rule book, “The official strike zone is the area over home plate from the midpoint between a batter’s shoulders and the top of the uniform pants – when the batter is in his stance and prepared to swing at a pitched ball – and a point just below the kneecap” MLB (2024b). The top and bottom varies fractionally with every pitch because a batter’s stance will always be a little different. The left and right edges of the strike zone are fairly straightforward because any pitched baseball that crosses the 17 inch wide home plate, even if it’s just part of the baseball, is considered a strike (assuming its height is correct). All of these slight variations in the strike zone from pitch to pitch, coupled with the inherently tough task of determining whether an object traveling at close to 100 miles per hour crosses through an invisible three-dimensional box, while being obstructed by an individual (the catcher) directly in front of you, makes it almost impossible for an umpire to make it through even one game without making an incorrect call.

4 Technological Advancements

With all of the issues related to human umpires and significant advancements in pitch tracking technology in recent years, there is currently a push to replace umpires with automated strike zones. As of the writing of this paper, some Minor League Baseball leagues are giving automated strike zones a trial run. In 2023, the AAA level, which is the highest level of Minor League Baseball, implemented automatic ball-strike (ABS) systems league-wide (Olney, 2023). With this technology, it would seem that deciding to implement the ABS system in the MLB would be an easy decision. However, there is a vocal contingency of fans, players, and personnel that object to this technology because it takes out the “human element” of the game. They argue that part of what makes baseball special is the lack of perfection. Despite the objections of many, it seems inevitable that we will see ABS systems reach the MLB at some point in the near future.

5 Data

Another benefit of the advancements in technology across the MLB is the ability to measure many quantifiable aspects of every pitch of every game. Founded in 2015, Statcast is a tracking technology system used by the MLB in every ballpark to take these measurements, including horizontal movement on a pitch and arm strength of an outfielder (MLB, 2024a). Statcast was originally made up of two tracking systems, Trackman for radar and ChyronHego for video (Kagan, 2020). In 2020, the MLB replaced those systems with the Hawk-Eye camera tracking system. This is the technology that is used in tennis. One advantage of Hawk-Eye is that it does all of its tracking by 100 frame-per-second video (Kagan, 2020). This significantly improved the measurement accuracy of existing metrics and added many other metrics that were previously unable to be measured, many of which centered around pitch tracking data. For example, the magnitude and direction of spin on a pitch can now be determined, such as sidespin or backspin. Previously, only the magnitude of the spin (overall spin rate) could be measured. The pitch tracking data from Statcast was of particular interest for this paper.

For this paper, we used pitch data from the 2023 MLB season (Baseball-Savant, 2023). Our dataset consisted of every pitch that wasn't swung at by the batter. Pitch metrics that were of interest to us were velocity (mph), pitch zone, pitch coordinates, pitch type, and spin rate (rpm). For pitch type, we only included the seven most commonly thrown pitches for the 2023 season (fastball, sweeper, changeup, curveball, slider, cutter, and sinker). These seven pitches accounted for over 95 percent of all pitches thrown in 2023. Fastballs, cutters, and sinkers tend to be thrown with much

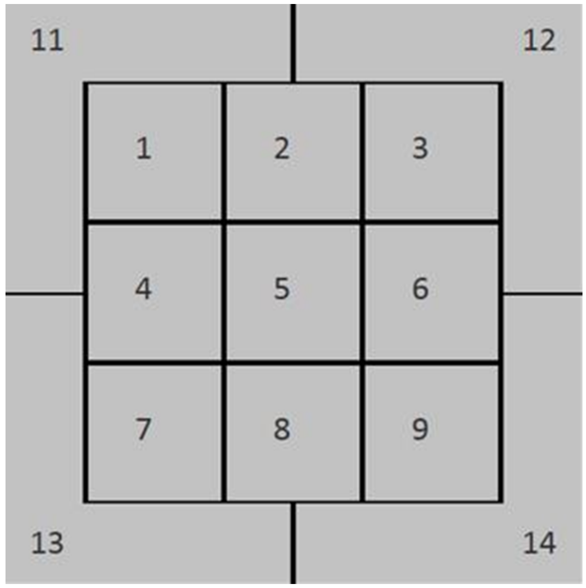
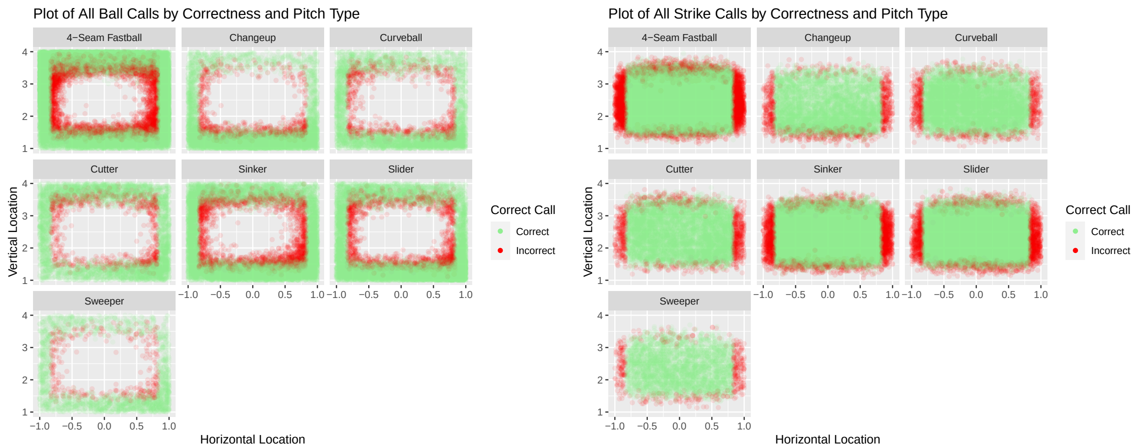


Figure 1: Pitch Zones from Umpire's Perspective

more velocity than sliders, curveballs, changeups, and sweepers. Pitchers generally throw pitches in different locations depending on the pitch type to be most effective. We will discuss this more in depth later. Perhaps the most important aspect of the pitch tracking data is location. Every pitch is placed into a zone based on its x-y coordinates. Zones 1-9 are for pitches in the strike zone, and zones 11-14 are for pitches out of the strike zone. Figure 1 shows the orientation of the zones from the perspective of the umpire behind home plate.

Other variables associated with the pitch that we were interested in were inning number, total pitches thrown in the game, inning top/bottom, score difference, batter and pitcher handedness, and count to batter. We condensed batter and pitcher handedness into a single variable with four possible combinations (e.g., right-handed batter and left-handed pitcher would be one combination).

Some initial data visualization yielded some interesting findings. Looking at all calls by whether the umpire's call was correct and pitch type in Figure 2 shows some interesting trends in pitchers' strategies. Slower pitches tend to be thrown at a much lower height while there isn't much of a noticeable trend in the location of faster



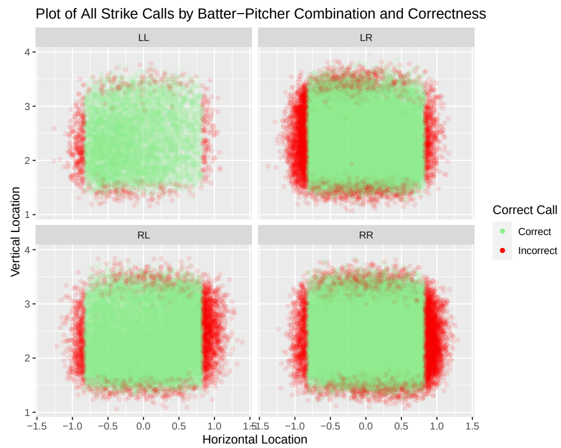
(a) Called Balls (b) Called Strikes

Figure 2: Plot of All Pitches by Correctness of Call and Pitch Type

pitches. Keeping slower pitches down is a strategy employed by most MLB pitchers since their downward movement makes them most effective when thrown lower. Not surprisingly, the majority of missed calls for slower pitches are near the bottom of the strike zone since that's the most common location to throw those pitches. There are less obvious trends in locations or missed calls of faster pitches.

Another interesting finding was in regards to where umpires tend to miss calls on the left and right side of the strike zone depending on the handedness of the batter and pitcher. Figure 3 shows that umpires tend to incorrectly call pitches as strikes when they are truly balls at a higher rate on pitches that are farther away from a batter. In fact, umpires incorrectly call outside pitches that are truly balls as strikes at a rate of about 6.4 percent, while they incorrectly call inside pitches that are truly balls as strikes at a rate of about 4.5 percent. There are many pitches that are well off the outside edge of the strike zone that are being called strikes. The possible explanation for this has to do with the alignment of umpires. Umpires position themselves on the inside shoulder of the catcher (shoulder of catcher that is closest to the batter). That means they have a much better view of pitches on the inside edge of the plate then pitches on the outside edge.

(a) Called Balls



(b) Called Strikes

Figure 3: Plot of All Pitches by Batter-Pitcher Handedness and Pitch Type

6 Methodology

To assess umpire accuracy, we modeled the binary variable indicating whether the umpire made the correct call or not. Pitches that were in zones 1-9 that were called a ball by the umpire were marked as incorrect, and pitches that were in zones 11-14 that were called a strike by the umpire were marked as incorrect. To account for pitch location, we created another variable that measured distance from center of the strike zone for each pitch. Combining the zone that a pitch was in with the distance from the center of the strike zone of that pitch, we were able to know the approximate location of the pitch.

We modeled umpire accuracy using two different models, one for pitches in the strike zone and one for pitches out of the strike zone. Splitting into two models allows us to make more inference about how different pitches affect umpire accuracy depending on whether they are true strikes or balls. We can learn more about umpire tendencies splitting up the data this way. Instead of just calling all pitches correct or incorrect, we know what kind of correct or incorrect call the umpire made (e.g., called ball when truly strike).

For determining the best model to use, we chose to use LASSO regression.

LASSO stands for least absolute shrinkage and selection operator. LASSO adds a penalty term to a model that makes it so that insignificant predictors will be shrunk to 0 (shrinkage) (Fahrmeir et al., 2013). Models tend to be simpler with this method to avoid over-fitting, and prediction will often be more accurate than ordinary linear regression. Furthermore, one can decide how much shrinkage they want to occur by choosing the value for a regularization parameter, λ, that will control the level of shrinkage that occurs. This regularization parameter is part of the penalty term. Ultimately, LASSO attempts to minimize the mean-square error (MSE) in a model (Fahrmeir et al., 2013). Often, one doesn't choose a λ value. Common practice is to choose a range of λ values to search over and perform a selection method such as cross validation to determine which model is best. The formula for the LASSO estimator is as follows:

$$\hat{\boldsymbol{\beta}}_\lambda = argmin\left(||y - \boldsymbol{X\beta}||^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

where $\sum_{j=1}^{p} |\beta_j|$ is the penalty term (Meier et al., 2008). With the regularization parameter, λ, being multiplied by the βs, one can see how the choice of λ influences the amount of shrinkage that occurs.

LASSO is also useful when there is multicollinearity between predictor variables, which we have with our data (Fahrmeir et al., 2013). Many variables we considered have multicollinearity issues. A simple example is the relationship between pitch type and velocity. Fastballs will almost always be faster than any other pitch that's thrown, and curveballs will generally be the slowest. This clearly would lead to multicollinearity between those predictors.

LASSO regression removes some levels of a factor while leaving others if it deems that some levels are insignificant. This is fine for prediction but can create issues for interpretation. To handle this issue, we decided to use a variation of LASSO called

Group LASSO. Group LASSO allows you to group factors (or levels of a factor) together, so that either all factors or levels in a group are kept or none are. (Meier et al., 2008) The algorithm runs very similarly as regular LASSO, but now entire factors need to be considered instead of individual factors or levels. We utilized this in our model to ensure that all levels of variables were kept together or removed together. The Group LASSO estimator, which is quite similar to regular LASSO is given by

$$\hat{\boldsymbol{\beta}}_\lambda = argmin\left(||y - \boldsymbol{X\beta}||^2 + \lambda\sum_{g=1}^{G}||\boldsymbol{\beta}_g||\right)$$

where we now see the penalty term having the predetermined groups of parameters instead of individual parameters. $\boldsymbol{\beta}_g \in \mathbb{R}^{\mathrm{df}_g}$ is the vector of predictors of length df_g for the g^{th} group of variables (Meier et al., 2008). We also needed to combine Group LASSO with logistic regression since we have a binary response in our model. Recall that logistic regression assumes

$$logit(p_\beta(\boldsymbol{x}_i)) = \beta_0 + \sum_{g=1}^{G}\boldsymbol{x}_{i,g}^{\boldsymbol{T}}\boldsymbol{\beta}_g,$$

where $p_\beta(\boldsymbol{x}_i) = P_\beta(Y = 1|\boldsymbol{x}_i)$, β_0 is the intercept, and $\boldsymbol{\beta}_g$ is the g^{th} vector of predictors. The Logistic Group LASSO estimator is as follows:

$$\hat{\boldsymbol{\beta}}_\lambda = argmin\left(-\ell(\boldsymbol{\beta}) + \lambda\sum_{g=1}^{G}s(df_g)||\boldsymbol{\beta}_g||\right)$$

where $\ell(\boldsymbol{\beta})$ is the log-likelihood function for logistic regression, λ still determines the amount of shrinkage that occurs, and $s(df_g)$ is the square root of the number of parameters in group g (Meier et al., 2008).

7 Discussion

For both the balls and strikes models, we performed k-fold cross-validation with k=5 folds over a range of reasonable λ values to determine the best model. The models that gave us the lowest MSE were $\lambda = 7.8$ for strikes and $\lambda = 31.8$ for balls. Figure 4 shows the values of our logistic regression coefficients for the strikes model, and Figure 5 shows the logistic regression coefficients for the balls model. Not surprisingly, the further away from the center of the plate a strike was (and conversely, the closer to the plate a ball was), the lower the probability of it being correctly called was. The highest values for the distance from edge variable for the strike model were in the far corners of the strike zone, while the lowest values for the ball model were just off the edge of the plate, which explains the lower probability of the calls being made correctly. Additionally, our model shows that umpires tend to make less correct calls on pitches in the strike zone and more correct calls on pitches outside the strike zone when the batter is behind in the count (e.g., 0 balls and 2 strikes, 1 ball and two strikes, etc). This means that the strike zone tends to shrink in those situations. Umpires tend to not want to call a batter out on strikes on borderline pitches. The converse is true when the pitcher is behind in the count (e.g., 3 balls and 0 strikes). In those situations, umpires are much more likely to make the correct call on pitches inside the strike zone, while they are less likely to make the correct call on pitches outside the strike zone. This means that the strike zone tends to expand in those situations.

Some other interesting observations were that umpires are likely to shift the strike zone down from its true dimensions, especially at the bottom of the strike zone. There was a very high probability of pitches in zones 7, 8, and 9 (bottom of the strike zone) being called a strike and a decreased probability of pitches in zones 13 and 14 (just below the strike zone) being called strikes when they are actually balls. There wasn't much of a difference in the correctness of calls based on pitch type,

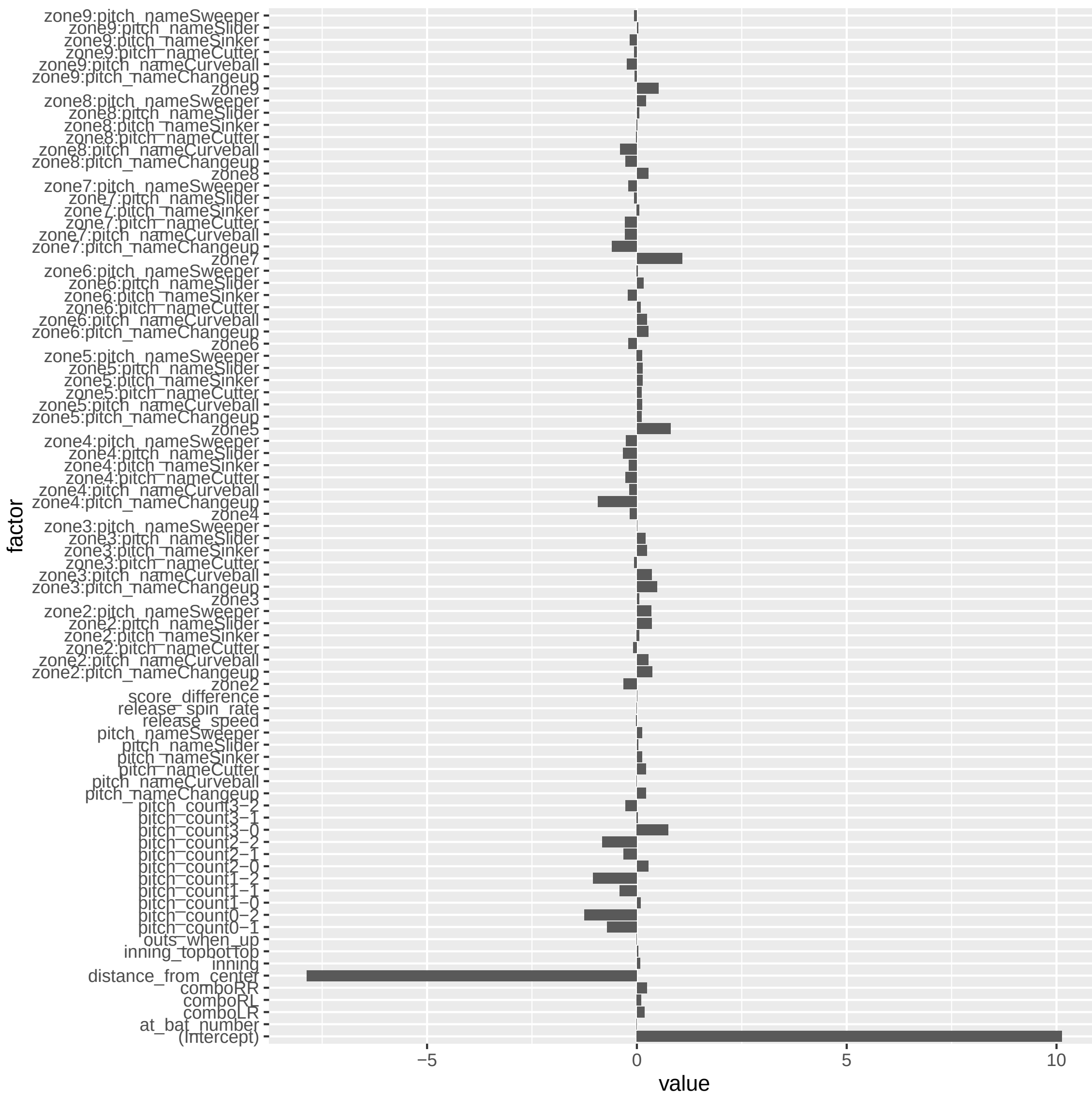


Figure 4: Strike Model Coefficients

but there were some interesting interactions between pitch type and zone. A couple notable ones were that changeups in zones 1, 4, and 7 (left side of home plate) are expected to be called incorrectly at a higher rate than anywhere else, and curveballs are expected to be called as strikes at a much lower rate near the bottom of the strike zone than anywhere else. Our model also predicted that right-handed batters against right-handed pitchers would get the most accurate calls (although not by a large margin) made by umpires. A potential explanation for umpires being best in

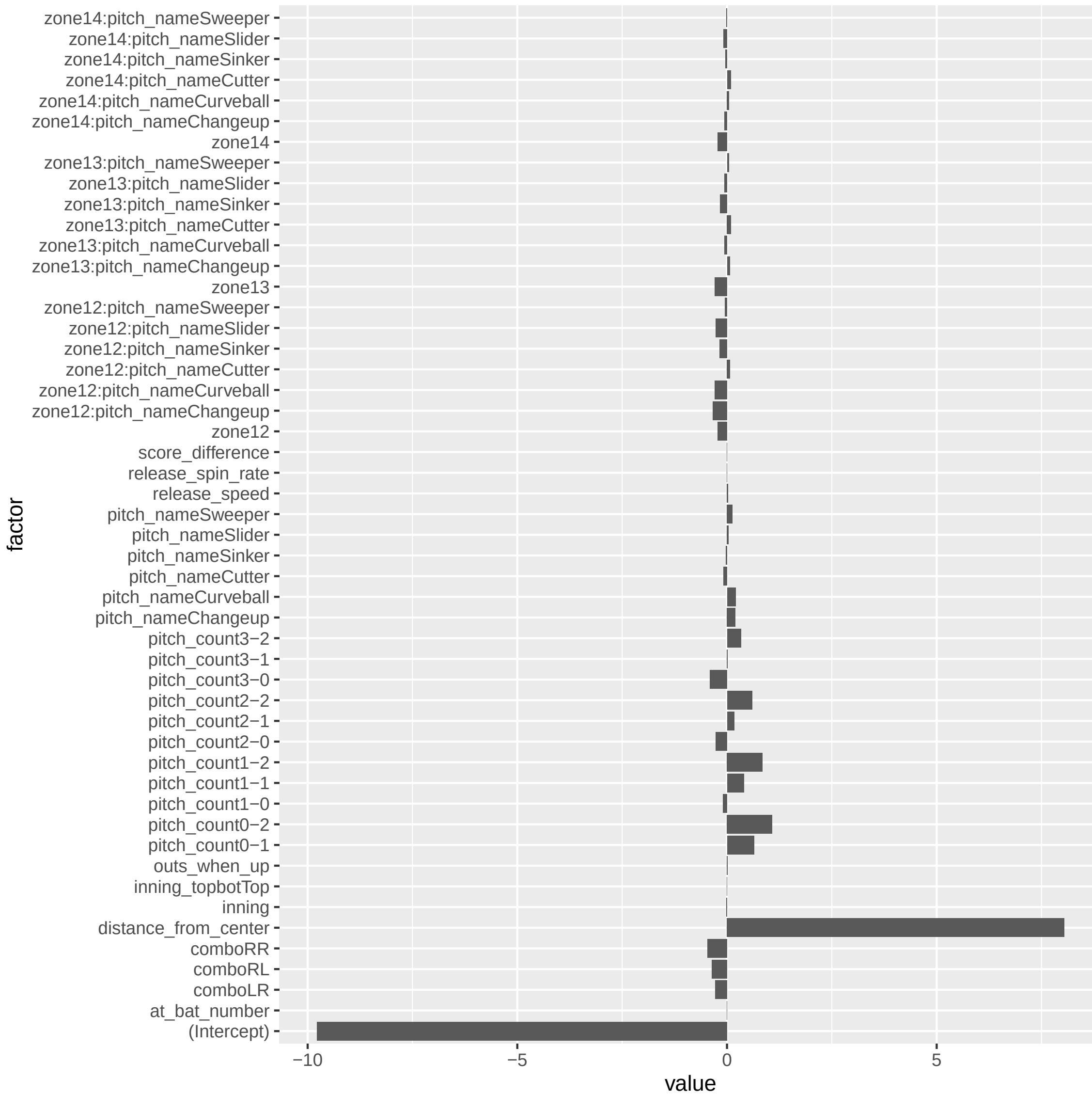
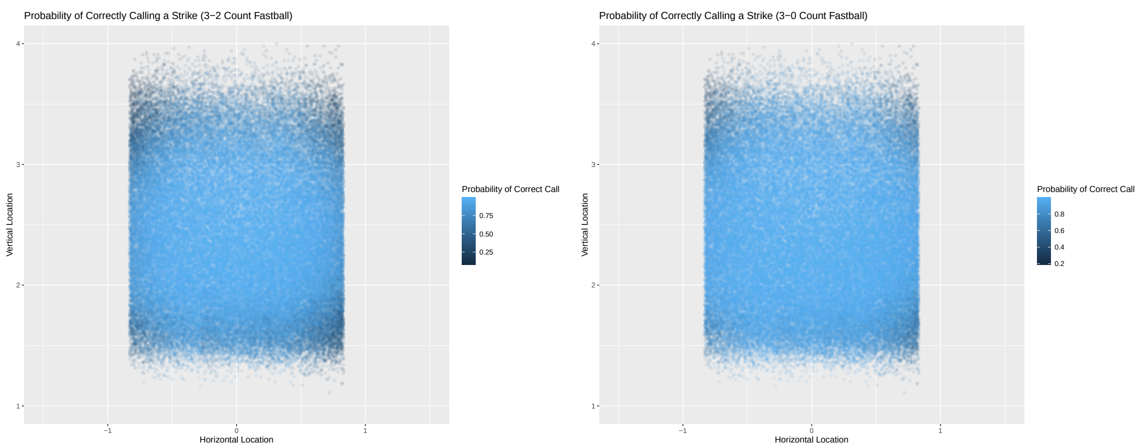


Figure 5: Ball Model Coefficients



(a) 3-2 Fastball (b) 3-0 Fastball

Figure 6: Heat Map of a 3-2 Fastball and a 3-0 Fastball

those scenarios is that right-handed players are most common in baseball, so umpires are most used to calling pitches in those scenarios.

Another way to visualize these umpire tendencies is by creating heat maps to illustrate where they are most likely to make the correct calls. Figure 6 shows an example of this. The left side of the figure shows a heat map of a fastball in a 3 ball and 2 strike count, while the right side shows a heat map of a fastball in a 3 ball and 0 strike count. This heat map illustrates what we know already from our model: there is a higher probability of a pitch in the strike zone being correctly called a strike when the count is 3-0 than when the count is 3-2.

8 Conclusion

For this paper, we investigated which factors influence the calls of MLB umpires. With the help of Statcast data, we were able to analyze every pitch from the 2023 MLB season to look for trends in the data. Using Group LASSO logistic regression, we were able to create two models, one for pitches in the strike zone and one for pitches not in the strike zone, to answer our research question. The most significant finding was that umpires tend to shrink their strike zone when the batter is behind in the count and expand their strike zone when the batter is ahead in the count. We also found they tend to shift the strike zone downwards, especially for pitches near the bottom of the zone. There were some other findings of interest in pitch type, batter and pitcher handedness, and the interaction between pitch type and location; however, none of those were as significant as the other findings.

In the future, it would be worth considering other interactions between some of our factors. One interaction of particular interest would be the interaction between pitch type and batter-pitcher handedness. Pitchers tend to throw different pitches to different batters depending on the handedness of the two parties, which could influence umpire calls. Another possible method to consider would be to reflect all the horizontal pitch locations to left-handed batters across the y-axis. This would allow all pitch locations to be the same in terms of how the umpire lines up and in relation to the batter. Looking at individual umpires, pitchers, and hitters could provide some interesting insight into how umpires are influenced. It would be worthwhile to investigate whether umpires are biased towards star players as well as how different the pitch calling tendencies are between individual umpires. A final area of interest would be whether umpires perform better or worse in certain stadiums. Every stadium backdrop is different, which could lead to differences in umpire accuracy in different places.

Given the rapidly changing landscape in sports technology, it remains to be seen if umpires will soon become a relic of the past and be replaced by computers. For the time being, however, the umpire is still one of the most important people on the field. In a sport as analytically driven as baseball, every quantifiable piece of information matters. As long as umpires are still in use, players and teams will want to use every bit of information they can about every individual on the field, including umpires, to gain a competitive advantage.

References

Baseball-Savant (2023). Statcast, trending mlb players and visualizations. "https://baseballsavant.mlb.com/".

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer, Berlin, Germany.

Kagan, D. (2020). There's a lot of physics to do now that hawk-eye is up and running. "https://tht.fangraphs.com/theres-lots-of-physics-to-do-now-that-hawk-eye-is-up-and-running/".

Meier, L., van de Geer, S., and Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):53–71. http://www.jstor.org/stable/20203811.

MLB (2024a). Statcast: Glossary. "https://www.mlb.com/glossary/statcast".

MLB (2024b). Strike zone: Glossary. "https://www.mlb.com/glossary/rules/strike-zone".

Olney, B. (2023). Sources: All aaa parks to use electronic strike zone in '23. "https://www.espn.com/mlb/story/_/id/35434317/sources-all-aaa-parks-use-electronic-strike-zone-23".

Appendix I: R Code

R code for the paper can be found at: https://github.com/mhessler35/Hessler-Writing-Project.git