Zach Landry

Atlanta Braves R&D Questionnaire

Analysis Questions

Please limit each answer to a maximum of 500 words.

1. On 8/24/2021, the Cardinals trailed the Tigers 4-3 going into the top of the 9th. To begin this inning, Daz Cameron doubled, Akil Baddoo struck out, and Jonathan Schoop grounded out, moving Cameron to 3rd. The batter is now Robbie Grossman. Assume Luis Garcia will pitch through the 5th spot in the batting order, Jeimer Candelario. Should the Cardinals intentionally walk Grossman? Describe what your process would be to determine whether to pitch to him. The following link contains the box score information for this game: https://www.mlb.com/gameday/tigers-vs-cardinals/2021/08/24/632781#game_state=final,lock_state=final,game_tab=box,game=632781

Determining whether to pitch to a batter is a classic example of how leveraging data can aid in-game decision making. Based on the following analysis, the Cardinals should intentionally walk Grossman and pitch aggressively to Zack Short, who entered as a pinch runner the previous inning. First, let's assess the situation to see how an intentional walk would affect the game from the Cardinals' perspective. Already trailing by a run late in the game, an intentional walk would create a force at first or second base and perhaps shift the fielders into a no-doubles defense. Looking at the career matchup statistics for the next three hitters versus Garcia, Grossman and Candelario are a combined 5-8 with a 1.450 OPS, while Short had yet to face Garcia and was batting below the Mendoza line in his first MLB season. In the month of August, Grossman and Candelario were carrying a .857 and .790 OPS, respectively, while Short was having his worst month of the season at .328. In the previous week, none of the three hitters were hitting well, but Short was the coldest going 3 for his last 17. Garcia had an above-average sinker and slider combo with a usage approaching 90% of the time in 2021, and Short was by far the worst performing hitter against these two pitches in terms of wOBA.

In terms of win probability, the Tigers only gained half a percent from intentionally walking Grossman – a wise concession for a much more advantageous matchup with Short. It's also important to note that Grossman had a career-high 20 steals in 2021 and above-average sprint speed according to Statcast, and the situation would in warrant a no-doubles defense to prevent Grossman from scoring on an extra-base hit. Garcia, however, was extremely effective at holding runners with 0 stolen bases allowed in 2021 so the likelihood of Grossman swiping second is assumed to be relatively low. Additionally, Short wasn't a power threat toting a mere .243 wOBA, reducing the need to be as aggressive with the defensive positioning.

Coming off the bench late in the game and batting in a high-leverage situation is one of the hardest things to do in baseball. Based on the splits and matchup statistics, the Cardinals' best option is to pitch to Short rather than Grossman. This scenario is an ideal use case for a model that takes these factors into consideration and produces a real-time decision.

Zach Landry

2. You are running a generic mid-market team and are exploring the idea of signing Cody Bellinger this offseason. What contract would you be willing to offer him? Please explain your thought process and discuss any important considerations.

After several lackluster and injury-marred seasons for the 2019 NL MVP, Bellinger had a resurgent 2023 campaign and will be one of the most sought-after free agents this offseason, assuming he declines his mutual option for 2024. With this option worth $25 million, it's almost certain that Bellinger and Boras Corp. will be looking for a more lucrative, long-term deal given his age and the season he just had. Identifying contracts of comparable players in terms of statistical production and age is necessary in order to gauge his market value (data obtained from FanGraphs' RosterResource Free Agent Tracker):

FA Year	Name	Position	Age	Prev. WAR	Years	Total Salary	AAV
2022	Nick Castellanos	RF/LF/DH	30	3.7	5	$100M	$20M
2022	Kris Bryant	3B/LF/RF	30	3.1	7	$182M	$26M
2023	Brandon Nimmo	CF/RF/LF	30	5.2	8	$162M	$20.25M
2022	Michael Conforto	RF/LF	29	1.4	2	$36M	$18M
2022	Kyle Schwarber	LF/1B/DH	29	2.9	4	$79M	$19.75M

According to Baseball Savant, Bellinger was above the 80th percentile in batting, baserunning, and fielding run value in 2023. Considering his dynamic skillset, young age, and a bull market for outfield free agents, his projected average annual value could very well be in the $25 to $30M range. Therefore, a possible contract offer from a generic mid-market team could be in the **5-year, $137.5M** range with performance-based incentives and a 6th-year option. Of course, the duration and AAV of this contract depends on how Bellinger fits into the team's short and long-term plans and how aggressive they are willing to be to compete with big-market teams this offseason. Considering all the above factors, this contract structure balances risk and reward, providing the opportunity to sign a versatile, high-ceiling player while mitigating the risk associated with his injury history and inconsistent performance.

3. Pitcher A walks half the batters he faces and strikes out the other half. Pitcher B doesn't walk or strike out any of the batters he faces. Which pitcher would you prefer? What ratio of strikeouts to walks would make you indifferent between the two pitchers?

To fully answer this question, more context is needed such as how Pitcher A and B are used (starter vs. reliever), how effective Pitcher B is at inducing weak contact, the quality of the defense, or the game situation. However, by making the following assumptions, an empirical answer can be derived using a mathematical approach based on probabilities and expected value:

- Probabilities of ball-in-play events are calculated on a league-average, per-plate-appearance basis. For example, the league single rate in 2023 was 1B / PA = 26,031 / 184,104 = 0.141, so this value will be used as the probability of a single.
- Expected values are determined using Tom Tango's 2023 linear weights which, according to FanGraphs, "properly measure different kinds of events [by] taking the average run expectancy impact of each type of event (singles, doubles, etc) and finding their average."

- The only ways to reach base are walks, hit by pitch, singles, doubles, triples, and homeruns since these are the only events in which linear weights are provided. The problem doesn't state that Pitcher B doesn't hit any batters he faces, so this is included in the set of possible outcomes. Other ball-in-play events such as errors, ground into double play, and sacrifice flies are not considered.

Hence, the expected runs per plate appearance, or expected runs per batter (EPB), can be calculated for the two pitchers below:

$$EPB_A = P(BB)*wBB + P(K)*wK$$
$$= (0.5)(0.696) + (0.5)(0)$$
$$\approx 0.348$$
$$EPB_B = P(HBP)*wHBP + P(1B)*w1B + P(2B)*w2B + P(3B)*w3B + P(HR)*wHR$$
$$= (0.011)(.726) + (0.141)(.883) + (0.045)(1.244) + (0.004)(1.569) + (0.032)(2.004)$$
$$\approx 0.259$$

Based on the above assumptions and calculations, it's clear that Pitcher B is preferred as his expected runs per batter is less than Pitcher A. To determine a K:BB ratio that creates a neutral preference, we can adjust P(BB) for Pitcher A so that $EPB_A = EPB_B$. Setting x = P(BB):

$$0.696*x = 0.259 \Rightarrow x \approx 0.372$$

With P(BB) ≈ 0.372, it follows that P(K) = 1 – P(BB) = 0.628. Therefore, the ratio of strikeouts to walks that would create indifference between the two pitchers would be 0.628 / 0.372 ≈ 1.688. Rephrasing the answer in terms of the question, Pitcher A would need to strikeout approximately 62.8% and walk approximately 37.2% of batters faced.

4. Briefly explain how you would go about estimating the effect of catcher framing at the major league level? Assume you only have access to the identities of the people involved, information about the pitch (location, characteristics, etc.), and information about the game (count, inning, score, etc.).

A supervised machine learning model such as mixed-effect logistic regression would be an appropriate method of estimating the effect of catcher framing at the MLB level. This type of model can incorporate random effects for pitchers, catchers, and umpires to control for their individual tendencies and fixed effects for pitch characteristics and game situations to control for these factors as well. Using pitch-level data, each batter's strike zone coordinates would first be determined to create a binary variable indicating whether the pitch was in the strike zone before encoding a new binary dependent variable for whether the umpire called the pitch a strike (1) or a ball (0). The model would then predict the probability of the pitch resulting in a called strike. The catcher's framing ability would be quantified by extracting the catcher random effect estimates from the model while controlling for other factors. To assess the accuracy and robustness of the model, the data could be split into training and test sets. The results of this model could give insight into which catchers are the most effective framers in addition to which umpires are more susceptible to framing effects.

Zach Landry

Modeling Questions

For these exercises, you will be using your knowledge of R or Python to answer a few baseball-related questions.

Use the attached Trackman pitch by pitch data of Braves pitchers from the 2018 season to answer the following prompts. It is not necessary to utilize every column in the attached file; only use those you feel are necessary. There is a GLOSSARY defining the columns in the PitchData.csv file on the second page of this document.

This exercise should not take more than a few hours. Please include all of your code with your responses. If you do not know how to complete one or more of the questions, feel free to leave them blank.

https://github.com/zlandry11/CB-Whiff-Predictor.git

1. **Create TWO models to predict the likelihood of a swing and miss based on the characteristics of a curveball. Evaluate and compare the performances of your models using any method(s) you'd prefer. Explain your results in 500 words or less.**

2. **Using your preferred model from Question #3, create a visualization to display the most important characteristics of a curveball in recording a swing-and-miss. Explain your visualization in 500 words or less.**

Note: Models in this exercise will be less accurate due to small samples of pitches and pitchers, so proceed with your evaluations and conclusions as if there were a complete set of 2018 data.

Zach Landry

PitchData.csv Glossary

Variable	Definition
Pitcher_ID	The pitcher's MLBAM ID
Pitcher	The pitcher's full name
Pitcher_Throws	The pitcher's handedness
Batter_ID	The batter's MLBAM ID
Batter	The batter's full name
Batter_Hits	The batter's handedness
Game_Date	The date the game occurred
Top_Bot	Whether it is the top or bottom of the inning (1 signifies the top and 2 signifies the bottom)
Inning	The inning the pitch was thrown
Balls	The number of balls when the pitch was thrown
Strikes	The number of strikes when the pitch was thrown
Outs	The number of outs when the pitch was thrown
Pitch_Outcome	The outcome after the pitch was thrown
Pitch_Type	The pitch type (4-Seam and 2-Seam are grouped as fastballs)
release_speed	The pitch's velocity (mph)
x_movement	The pitch's horizontal movement (inches)
z_movement	The pitch's vertical movement (inches)
release_spin_rate	The pitch's spin rate (rpm)
spin_dir	The pitch's spin axis (degrees)
release_pos_z	The horizontal release point for that pitch (ft)
release_pos_z	The vertical release point for that pitch (ft)
release_extension	The release extension for that pitch (ft)
plate_x	The horizontal location of the ball when it crosses home plate (ft)
plate_z	The vertical location of the ball when it crosses home plate (ft)