2018-2022 MLB Umpire Performance Analysis

Data Acquisition:

In this project, we are looking at the pitch call-strike accuracy for the MLB umpires during 2018-2022 season. We download the entire 5-year pitch information from pybaseball API and cut down to only called balls and called strikes. Then we apply two functions to check if umpires made the correct call during the game. First function checks if the ball landed in the home plate, and the second function runs 1000 simulations to test if umpire can make the call correctly 95% of the time. (data processing.py)

Strikes vs Balls:

Our data consists of over 1.6 million total pitches, 1.5 million of which are called correctly. This gives us a 5-year Bad Call Ratio of 7.6%.

A screenshot of a computer code

Description automatically generatedA close-up of a number

Description automatically generated

We then break down the data into balls and strikes. Roughly two thirds of the total pitches are balls and rest strikes. We see that strikes have a BCR of 13.6% and balls have a BCR of 4.6%.

A screenshot of a computer code

Description automatically generatedA close-up of a number

Description automatically generatedA close up of a number

Description automatically generated

This means that of if an umpire calls a pitch a strike, 13.56% of the time that this pitch is indeed a ball, which implies that umpire favors pitchers more than batters.

Pitch Types:

During the 5 years, there are total of 16 different pitch types and their corresponding pitch numbers and bad call ratios are as follows.

A screenshot of a computer

Description automatically generatedA white rectangular object with black text

Description automatically generated with medium confidence

After removing 5 outliers with the less than 1000 occurrence, we see that umpires have the highest bad call ratio at FF (0.082587) and SI (0.087315), while having the lowest bad call ratio at CH (0.057672) and FS (0.050976). The average bad call ratio for the 11 pitch types is 7%, and four-seam fastballs (FF) are 18% more likely than average to be called incorrectly and sinkers (SI) are 24.8% more likely to be called incorrectly.

Pitch Zones:

The home plate is divided into 9 equal parts, with zone 11, 12, 13, 14 outside the Homeplate on the four corners. Each pitch lands in one of the 13 zones and their corresponding occurrences and Bad Call Ratios are as follows.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

The thirteen strikes zones have a average Bad Call Ratio of 9.2%. Zones 1 and 3 are 2.22 times and 2.28 times more likely than average to be called incorrectly respectively. Zones 7 and 9 are 26% and 66% more likely than average to be called incorrectly. It appears that umpires are better at calling pitches that lands at the bottom of the home plate than the ones that lands at the top.

Sinkers and Fast Balls:

As previously stated, Sinkers and Fast Balls have the highest Bad Call Ratios amongst all the pitch types. Their respective Ball and Strike Bad Call Ratios are as follows.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

If a pitcher throws a Sinker, 5.7% of the chances that the called ball will be a strike, and 13.7% of the chances that the called strikes should be a ball. If a pitcher throws a Fast Ball, 5.3% of the chances that the called ball will be a strike, and 13.8% of the chances that the called strikes should be a ball.