Collin Hough
2023 Baseball Analytics Trainee Take Home
Part 2 A

Pitcher A should develop a cutter and Pitcher B should develop a circle changeup. My logic for these choices is to try and accentuate the strengths of each player. When analyzing both players, I wanted to see what pitch types each predominantly throws and the amount of success they currently have in terms of swing and miss rate. To do this, I gathered the swing and miss rate and usage rate for each of the pitch types in their arsenals. Swing and miss rate and usage rate can be defined as follows:

$$Swing\ And\ Miss\ Rate = \frac{Amount\ of\ swinging\ strikes\ per\ pitch}{Amount\ of\ times\ pitch\ was\ thrown}$$

$$Usage\ Rate = \frac{Total\ amount\ of\ times\ pitch\ type\ was\ thrown}{Total\ amount\ of\ pitches\ thrown}$$

After doing this analysis, I discarded any pitch types with a usage rate under 10% to remove any potential outlier results. Pitcher A had the most success throwing two-seam fastballs and sliders. Pitcher A threw a two-seamer ~53% of the time and achieved a swing and miss rate of ~14% with it. Due to their heavy usage of two-seam fastballs, a cutter would add more deception to their arsenal because of its opposing motion in comparison to the motion path of standard two-seamers.

Regarding Pitcher B, they also had a high fastball usage rate (~54%), but only a ~12% swing and miss rate. They achieved significantly more success throwing changeups and curveballs with swing and miss rates of ~18% and ~17%. In total, these two pitch types accounted for ~35% of Pitcher B's thrown pitches. Because of their strength throwing pitches with more finesse, I believe a circle changeup would fit perfectly into their arsenal to continue confusing opposing batters.

For predicting the swing and miss rates of each pitch type, I gathered the data on similar pitches in the dataset. To predict the swing and miss rate of Player A's cutter, I gathered the data of all two-seam and four-seam fastballs in the dataset. Then, I created a linear regression model and trained it using this dataset. Afterwards, I created a new feature dataset of randomized values of each feature and used this to predict the new swing and miss rate. I followed a similar procedure for predicting the swing and miss rate of Player B's circle changeup, but instead gathered the data on changeups, curveballs, and sliders.