Evaluating Pitcher Performance by Time Through Order

Sebastian Symula

Department of Statistics

University of Connecticut

March 24, 2025

Abstract

In this paper, we investigate the decline (or lack thereof) of an MLB pitcher throughout a game. To accomplish this, we use Statcast data that includes every pitch from the 2024 MLB season. Through methods like t-tests, ANOVA, and linear regression, we try finding the significant factors that lead to an increase in base runners later in the game.

Keywords: keyword1, keyword2, keyword3.

1 Introduction

In baseball, pitching is arguably the single most important aspect. Each team is allowed a roster size of 26-28 (26 in the regular season and 28 in playoffs), and has 13-14 pitchers on their active roster at any given time. Half of each team's roster is dedicated to their pitching, as well as large sums of money for the starting pitchers. During the course of a game, the starting pitcher's performance will inevitably decline. Their pitches will get slower, and their

placement will become sloppy, leading to at bats resulting in base runners at a higher rate. At this point, the starting pitcher is pulled from the game and a reliever is brought in, but the damage has already been done. The starting pitcher has already allowed a couple of base runners and the usually less reliable reliever has been put in a tough spot.

The transition between the starting pitcher and the reliever is where games are won and lost. If teams wait too long to pull the starter, there may be an irreversible cascade of base runners, leading to runs scored. On the other hand, if they pull the starter too early, they may tire out their bullpen for future games. Additionally, if they pull the starter too early, the reliever's performance may be worse than the fatigued starter's. If there were a way to optimize pitcher performance, allowing starting pitchers to last longer into each start, while also knowing the correct time to pull them before it's too late, we should see a drastic decrease in the number of runs scored late in games.

Baseball is a very statistically driven sport. As such, many papers have been written on similar questions. For example (Brill et al., 2023) evaluates the Time Through the Order Penalty, the idea that batters will become more familiar with a pitcher's style of game-play. This leads to more hits, the more times a particular batter faces the starting pitcher. This paper specifically evaluates the Third Time Through the Order penalty, which is a common cut-off point for pitchers. (Brill et al., 2023)'s analysis concludes that although there is a steady decline in pitcher performance throughout the game, it may not necessarily be due to the Time Through the Order Penalty. While this does have relevance to my question, it goes less in depth. I want to evaluate the specific factors that may lead to a pitcher's consistency or lack thereof.

There has also been work done on models to predict when a starting pitcher should be pulled from the game. Ganeshapillai and Guttag (2014) paper created a model that decided when a starting pitcher should be removed based on whether it was predicted that a run would be surrendered in the following inning. Although this method is pitcher-specific, it was done from game to game. In some cases, that approach may lead to unreliable results

due to small sample size. This method utilizes historical data (prior to the game), but as for making the in-game decision, there will always be limited data. This is why I wanted to use averages for each starting pitcher, the results will be more representative. Their approach does sound very interesting and may be worth revisiting at a later time.

The rest of the paper is organized as follows. The data will be presented in Section 2. Section 3 describes the methods. The results are reported in Section 4. A discussion concludes in Section ??.

2 Data Description

Use this section to describe the data that helps to answer your research questions. Some descriptive statistics in tables or figures are suggested here to summarize data. See Notes 4 and 5 for details.

The dataset is pitch-by-pitch data from the 2024 MLB season and was collected from Baseball Savant, obtained through the CSAS Data Challenge webpage. There were originally 701557 rows and 113 columns but this changed greatly during the pre-processing stage. Since each row represents a pitch, but my question concerns pitcher specific trends throughout the course of a game, the data needed to be compressed. Using the pitcher column, I filtered the dataset for just starting pitchers and split that into three separate datasets, one for each time through order.

The result of this pre-processing was three datasets with 304 rows each (one for each starting pitcher) and about 20 columns containing the pitcher's averages for the given time through the order for relevant variables like pitch speed,on base percentage, most used pitch type, most used pitch percent, arm angle, and others.

In the future, I plan on finding the differences between these averages for each time through the order and making datasets for these. I will make three of these as well: 1st time minus 3rd time, 1st time minus 2nd time, and 2nd time minus third time. Finding

these differences should balance out talent differences between pitchers. We are interested in aspects that contribute to a pitcher's gradual decline throughout the game, not what makes the pitcher good in the first place. Finding these differences will allow us to put a magnitude on the decline (or incline) of a pitcher throughout the game measured by on base percent.

Currently, the response variable is the on base percent allowed by the pitcher for the given time through the order, which would make it a continuous variable. We would expect that in most cases, we would see on base percentage positively associated with times through the order. However, there may be certain pitchers who actually allow fewer base runners the more times through the order they go. In the future, if these cases exist, I may try binning this response into a binary positive or negative response so that it may be treated as a classification problem. I.e., what separates pitchers who allow fewer base-runners further into the game from pitchers who allow more base-runners further into the game.

Each table and figure included must be explicitly referenced and commented upon in the text. For example, Table **??** summarizes some distributional features for some of the variables in our dataset. The `[!t]` specification forces the table to be located at the top of the page.

2.1 Variables

Some important variables:

release_pos_x,release_pos_y,release_pos_z: These are the 3d coordinates of the pitcher's release point of the pitch. Since these will all be averages, we will essentially get each pitcher's average release location when we put them together. X is horizontal direction from catcher's perspective, z is vertical direction from catcher's perspective.

plate_x,plate_z: Coordinates of pitch over the plate. X is horizontal direction from catcher's perspective and z is vertical from catcher's perspective.

pmv: Pitch Mix Variation. A measure of a pitcher's diversity with respect to the kinds of pitches thrown and the proportions. A higher pitch mix variation (PMV) would indicate

first that a pitcher has a relatively diverse mix of pitches and, second, throws each pitch roughly as much as any other. A lower PMV would indicate a pitcher has fewer pitches and relies on just one or maybe two of those the vast majority of the time (Perry, 2021). The equation given by

$$PMV = 1 - \left(\sigma_\tau + \frac{max(\tau)}{\nu}\right) \tag{1}$$

,

where τ denotes pitch-use proportions and ν represents the number of unique pitch types.

pitch_variety: Another measure of pitch variation that I made. Calculated by taking dividing unique pitch types by pitch type proportions and then taking the log for scaling reasons.

in_box: The proportion of pitches that were in the strike zone.

is_strike: Proportion of pitches that result in a strike.

arm_angle: Angle of pitcher's arm at release point.

p_side_adv: Proportion of pitches where the pitcher has advantageous handedness matchup with batter. This happens when the two are of the same handedness.

pitcher_fav_pitch: The pitcher's most thrown pitch type. Encoded for regression.

Some Visualizations:

3 Methods

3.1 T-Tests

To first answer the question of whether pitchers allow more runners on base the more times through the order they go, I can perform t-tests on the datasets that contain the differences in OBP, and test with H_0: $\Delta_{OBP} = 0$ and H_a: $\Delta_{OBP} > 0$. It may also be a good idea to test H_0: $\Delta_{OBP_{1-2}} = \Delta_{OBP_{2-3}}$ and H_a: $\Delta_{OBP_{1-2}} \neq \Delta_{OBP_{2-3}}$. The point of this would be to see if

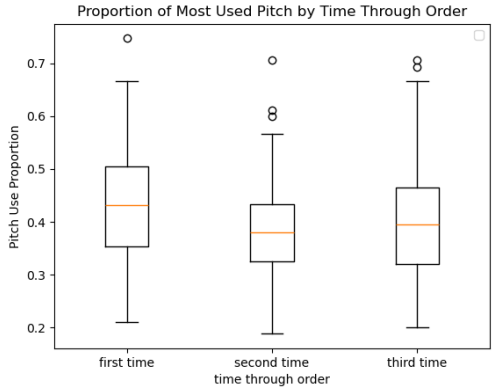

Figure 1: Most Used Pitch Proportion By Time Through Order

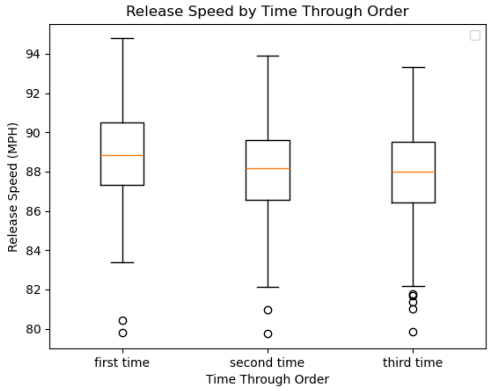

Figure 2: Release Speeds By Time Through Order

the drop in pitcher performance is gradual or not. If the change from OBP_1 to OBP_2 and from OBP_2 to OBP_3 is not statistically significant, this would suggest a gradual decline in performance.

3.2 Binning and Grouping

I may assign more groupings based on binary variables to see how those impact the results. For example, I may group by home and away for the pitcher to see if pitchers tend to hold up longer when at home, or if the factors that strongly correlate with consistent performance are different from the correlations when a pitcher is away. I could also group by handedness combinations between the pitcher and batter. It is a widely accepted fact that it is advantageous for a pitcher to be throwing against a batter who is the same handedness as him.

3.3 ANOVA

There are some good opportunities for ANOVA with our given dataset. We have a few important categorical variables, as well as several continuous variables that can be binned into categorical variables. I plan to run an ANOVA on change in OBP and use favorite pitch

type as groupings. This is following the thought that some pitches may be inherently more difficult to hit due to speed or movement of the pitches. Some pitch types are slower than others, so maybe pitchers who use a larger proportion of these slower pitches can keep up a high level of performance further into the start than those who use faster pitches like the four-seam fastball. I can also bin the proportion of most used pitch to see if pitchers with more lop-sided pitch type proportions tend to fair worse as the game goes on due to being more predictable.

3.4 Linear Regression

My general goal is to find the most important factors that contribute to changes in OBP between times through order. To evaluate the importance of these factors, linear regression analysis is a good place to start. I will use the dataframe containing the differences from OBP_1 to OBP_3 and run a multiple linear regression regressing the change in OBP onto release_pos_x,release_pos_z, plate_x,plate_z,release_pos_y,arm_angle,age_pit,p_side_adv,in_box, is_strike, pitcher_fav_pitch (encoded),pitch_variety,pmv,avg_rel_speed, fav_pitch_order, and fav_pitch_prop_order. Variables may need to be removed to keep the model parsimonious, which is especially important in our case because we only have 304 rows (one for each pitcher) in our dataframes. Additional variables may be engineered however, if there aren't enough significant ones already in the model. I plan to use LASSO to assist in this variable selection.

4 Results

4.1 Linear Regression

A preliminary linear regression model was fitted onto the data, which was split roughly 2/3 - 1/3. These variables all represent the aggregate difference for each pitcher from the first time through the order to the third time through the order Through this preliminary linear regression model, we find that is_strike,plate_z,avg_rel_speed, and in_box are all significant

predictors of OBP at $\alpha = 0.05$. The resulting R-squared was 0.345

This is just a preliminary model, and many of the 19 features may be dropped due to being insignificant. In addition, transforms or interaction terms may be applied to existing features to make them more relevant. Other, new features may be engineered from the existing dataset as well.

References

R. Brill, S. Deshpande and A. Wyner (2023). "A bayesian analysis of the time through the order penalty in baseball." pp. 1–36.

G. Ganeshapillai and J. Guttag (2014). "A data-driven method for in-game decision making in mlb." *MIT Sloan Sports Analytics Conference* **8**, 1–6.

B. Perry (2021). "Pitch mix variation and ways to measure it."