MLB Pitcher Luck 2024

By: Garrett Johnson

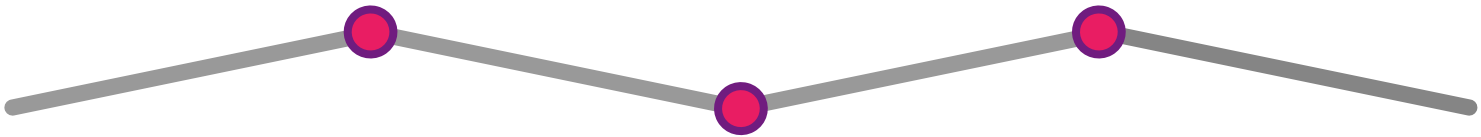
Outline

- Data was gathered from **Baseball Savant** and **Fangraphs**
- The r markdown file and csv with all of the data can be found via Github using this link: [2024 Pitcher Luck]
- Along with the 64 starting pitchers analyzed, the stats used were: **barrel rate**, **hard hit rate**, **BABIP**, **k rate**, **fly ball rate**, **ground ball rate**, **line drive rate**, **ERA**, and **xERA**

Manipulation Steps

Standardize column names so my column heads look cleaner and easier to read

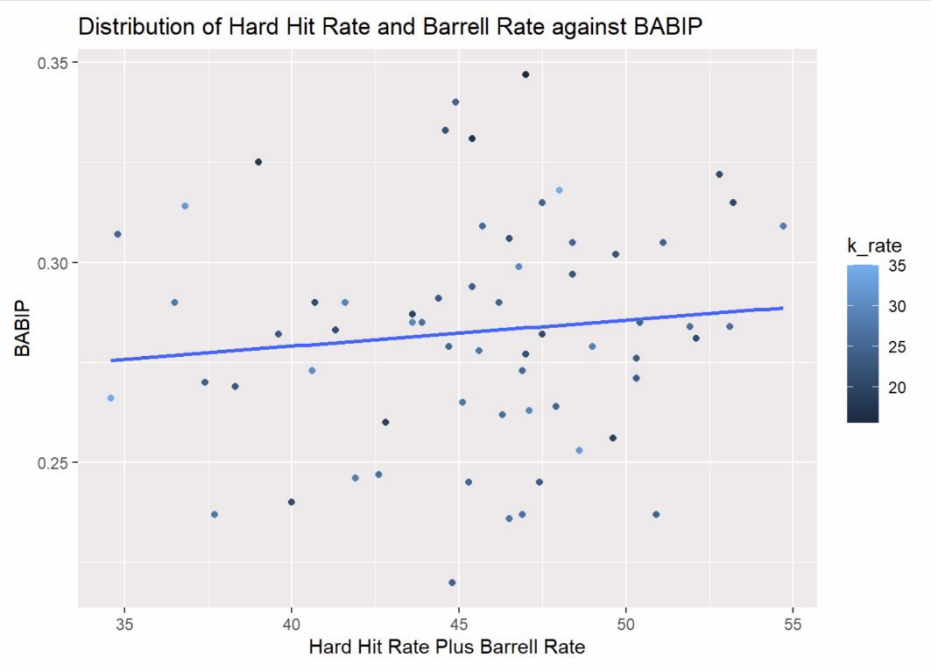
Compute the difference in ERA and xERA in excel to create the difference column



Combine barrel rate and hard hit rate in excel to get the created sum: barrell_hard_hit

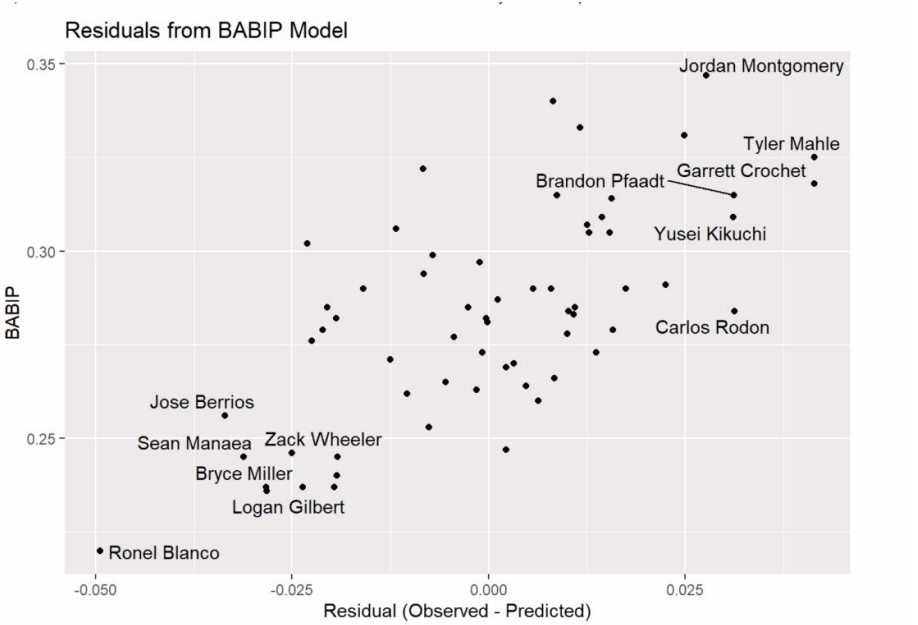
Scatterplot Model #1

- Compares BABIP against barrel rate and hard hit rate combined.
- Strikeout rate is factored by color to see which pitchers potentially have less balls in play.
- In the plot, typically the harder and more barreled pitchers get hit, the worse their BABIP is based on the trend line



Linear Regression

- Used linear regression to make predictions on pitcher BABIP using independent variables such as: barrel rate, hard hit rate, ground ball rate, fly ball rate and k rate.
- Then took residuals from the model (the predicted subtracted by the actual observed and plotted against BABIP to see who's actual was far away from the observed

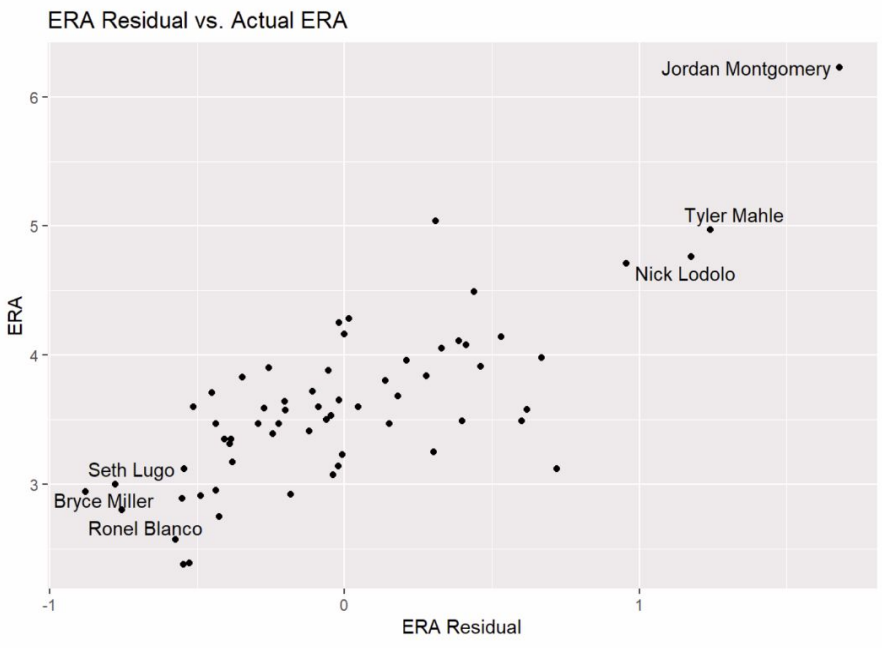


Residual Model Analyzed

- In taking a look at BABIP we are only trying to predict the balls in play that a pitcher is giving up.
- The residual model is centered at 0, so the further a pitcher is away from the middle, the further away they are from their actual BABIP based on the linear regression model.
- The further to the left a pitcher is, the luckier they are.
- The further a pitcher is to the right, means they are more unlucky.
- Based on the model, Ronel Blanco is the luckiest pitcher in the dataset and Jordan Montgomery is the unluckiest.

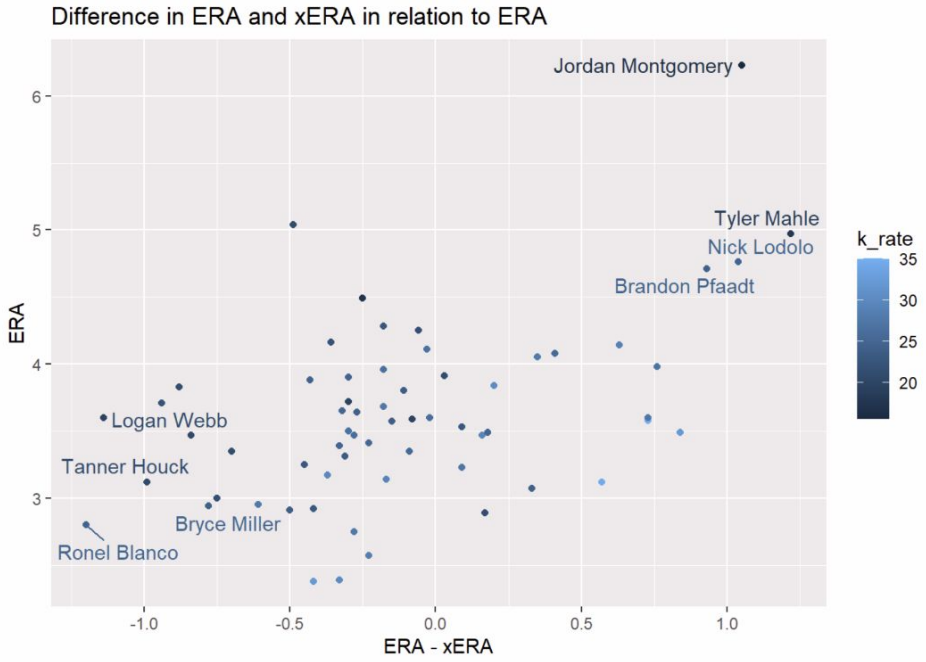
Cross-Reference Linear Regression Scatterplot

- This time we used linear regression on ERA with the same independent variables to see if we see the same lucky/unlucky pitchers as the first model.
- The residuals of ERA are as shown in the figure to the right.



Final Cross Reference

- xERA is a stat that tries to encapsulate expected outcomes rather than what actually happens on the field.
- By finding the difference in ERA and xERA, we basically have our own residual based on stats already given to us.
- In graphing the difference against ERA we can see how accurate our own linear regression model was.



FInal Cross Reference Analyzed

- Factoring in k rate we can see which pitchers are getting lucky/unlucky with less opportunities of balls in play or lucky/unlucky with more opportunities of balls in play.
- From the data it showed us some similar names that are far away from their actual outcomes such as Ronel Blanco and Jordan Montgomery, further boosting confidence in the regression model.

Summary

- With comparisons of expected stats and the residual scatterplot models, I concluded that Ronel Blanco and Bryce Miller were among the luckiest pitchers in baseball in 2024.
- The unluckiest pitcher, while still posting a very high ERA and xERA was Jordan Montgomery.
- This matters because raw numbers don’t always tell the full story in baseball. With linear regression and expected outcomes, we can better understand pitcher performance.