EFFECT OF COVID-19 IN THE 2020 MLB PLAYOFFS

Jose Nicolas Currea Vergara
Information Technology and Math Department

Abstract

Several studies have been performed to explore how the most recent pandemic, COVID-19, has affected the different sports in aspects such as player performance, team performance, player motivation, and franchise performance. In this poster we focus on player and team performance, and how they were affected by this pandemic. In our research, we consider Sabermetrics which are statistics that measure both individual and collective in-game activity in baseball. Since there was a format change for the 2020 MLB Playoffs, in this research we compare the ALCS, NLCS and World Series which are 7-game series, and this has not changed since 1994. For this research we use statistical and analytical topics discussed in our Sports Data Analytics course such as the use of Sabermetrics, and some computer science tools such as web scraping in python and data modeling. Our findings include how percentages calculates remain somewhat similar, meaning there is no big effect on percentages such as OBP, Batting Average, and OPS. However, we really see the impact in the on-pitch performance.

Introduction

The Major League Baseball (MLB) is a professional baseball association in the United States. A total of 30 teams play in the MLB and they are divided evenly between the NL (National League) and AL (American League), where 15 teams play in each. The National League was founded in 1876 and the American League was founded in 1901; the first World Series was held in 1903. In this World Series, the best team from each league began competing against each other in the World Series. Teams have switched from the NL and it remained an 8-team league until the New York Mets and the Houston Astros joined the league.

Now, Sabermetrics is a term widely used in the data analysis field to refer to the analysis of baseball statistics. Sabermetrics are widely used to forecast results by making predictions based on previous data, analyze player performance by recording and evaluating aspects of the game, and assist in decision making by offering insights into player's performance, matchups, and scouting prospects. Bill James started to use this term in 1980, who was a renowned baseball analyst. However, the analysis of baseball did not begin in 1980, baseball started to be analyzed through statistics since the 1840's.

Materials

To perform this research, we used the MLB statistics databases. These databases can be found online for free and can be either web scrapped or downloaded into a .csv or .xcls format.

In order to perform the analysis, we used our knowledge on Sabermetrics as well as our knowledge on python to build our data tables and construct graphs.

The python libraries used in this research were pandas, which offers data structures and operations for manipulating numerical tables and time series.. The numpy library was also used. This library adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Lastly, the matplotlib library was also used. This library provides an object-oriented API for embedding plots into applications

Methodology

This research started with one big question, what was the effect of COVID-19 on sports? However, this question was very broad. Together with Dr. Robert Shumaker, we decided to narrow it down to: what was the effect of COVID-19 in baseball? We chose baseball since statistics in baseball have had a huge evolution since 1845. Statistics in baseball, better known as Sabermetrics, are arguably the most developed statistics in all sports. The use of statistics in sports is still evolving, however, Sabermetrics are the most developed so far.

Our question was still broad, and we were not taking into account the format change that occurred in 2020, where the only 7 game-series were the NLCS, ALCS, and World Series. Therefore, we got to our research question, what was the effect of COVID-19 in the 2020 MLB playoffs.

Having our question ready, we moved forward to perform the analysis. We chose to make a comparison on particular Sabermetrics from the year 2017 through 2020. We first looked up the data tables, which we found in mlb.com. Having this, web scrapping was done and converted these data sets into .csv format. Having the data sets stored locally, we performed the analysis using the python libraries mentioned above. The results of this research are surprising.

Results

We decided to analyze the offensive team's statistics. We analyzed some of the most important Sabermetrics used. We analyzed offensive teams with respect to Batting Average, Average OBP, Average OPS, and Average homeruns in the series. Batting Average is the rate of hits per at bat, OBP is the rate at which a batter reached base in his plate, OPS is the sum of OBP and Slugging, and Home Runs in a whole series. The following graphs sum up our findings
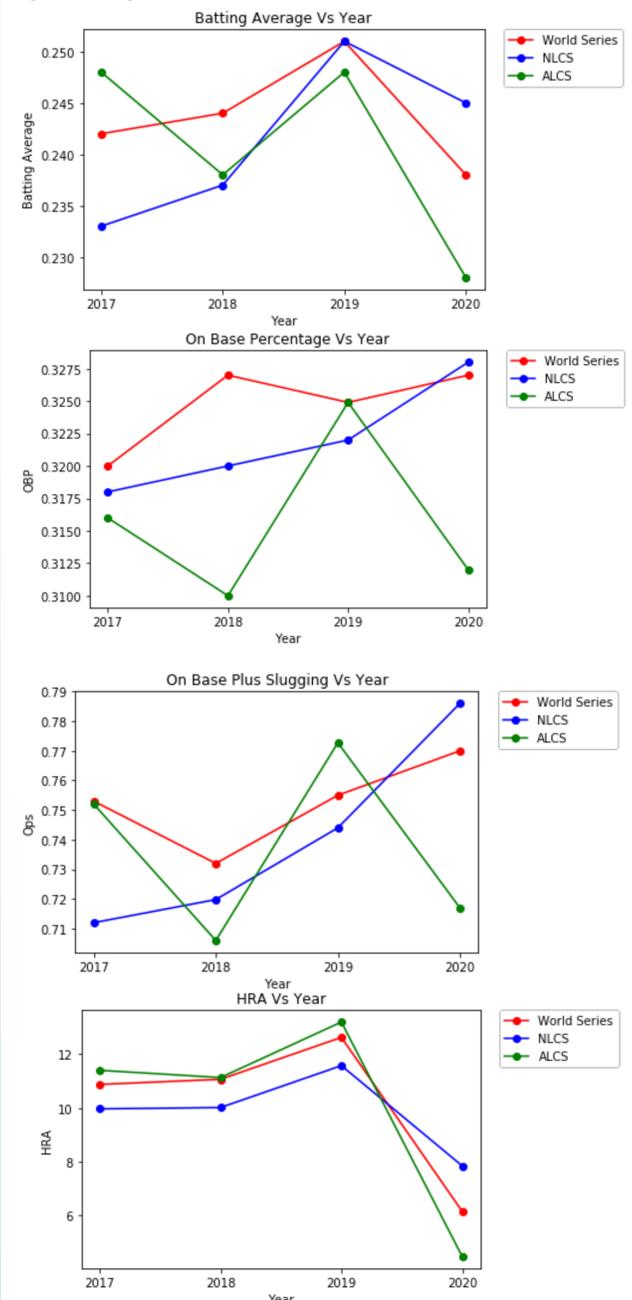
Conclusion

On the graphs, it is evident how the percentages such as OBP and OPS do not show a clear trend such as increase or decrease for the year 2020. This happens since these percentages are calculated by using different on-pitch statistics such as batting average, 1B, 2B, 3B, HR, among others.

Now, analyzing Batting Average and HR Average, which are the on-pitch statistics, we see a clear trend for the year 2020. They all have a significant decrease in comparison to previous years.

We could think about endless numbers of reasons of why this happens. However, the main reason may be related to the quarantine stage that we all had to go through. The lack of exercise facilities, practice areas, and constant gameplay may have been some of the main reasons for the "poorer" performance of players and teams in 2020.

Further Research Options

Some further research ideas can be how COVID-19 affected defensive teams, this being pitching and catching Sabermetrics. Also, researchers may look into how did the Pandemic affected the teams as a busines, meaning income, player salaries, staff salaries, among others.

Regarding other sports, similar research topics can be used. However, the data gathering may be somehow tougher considering the development level of statistics that other sports have

Acknowledgements