CMPSC 301
Data Analytics
Spring 2020

Lab 4, Part1:
Group Work to Explore The Pitcher-Batter Relationship
in Baseball Using Data

Objectives

The objectives of this group work are to gain practice in asking exploratory questions concerning the pitcher-batter relationship in baseball, and to use data analysis to respond to these questions and to extract new relationships among data variables. During the first part of the lab you will obtain the data, that you will pre-process as needed, and then will do exploratory data analysis. Then, during the second part of the lab, to be released next week, you will conduct a more in-depth statistical analysis of your data to respond to your questions.

Specifically, in the first part of the lab you are to answer your exploratory questions using exploration of two or more variables in baseball data. The data, which will be analyzed using R code, will be obtained from online sources. After you have your data, you may decide to wrangle your data, i.e., to transform the obtained data in to some suitable format for your work. Finally, you are to conduct the analysis and provide plots of the variables from your study of data that respond to your research question(s). Notably, your plots are to argue that there is some *non-random event* happening that may be used to explain the pitcher-batter relationship. In your report, you will justify your observations and thinking.

Figure 1: Baseball analytics helps the coaches plan better strategies for winning games. Note: A single research question is often impossible to directly respond to the underlying phenomenon that makes up the pitcher-batter relationship in baseball. To determine a broader view of how to understand the relationship to make informed decisions about placing players in your game, you will often need to approach this central question by responding to several smaller questions which are still directly related.

Reading Assignment

Please read Chapters assigned for this week's lessons which you will find in the class slides, in addition to reviewing your notes. Please take some time to gain experience with using Markdown to write your work. See *Mastering Markdown* `https://guides.github.com/features/`

HANDED OUT: 24th FEBRUARY 2020

`mastering-markdown/` for more details about Markdown. In addition, you may consult your notes on the talk given by Assistant Coach Dean Peterson of Allegheny Baseball who discussed motivations and methods for the study of this data.

GitHub Starter Link for Groups

<div align="center">

STOP! STOP!
Not everyone will be clicking this link at this time!
Only the team leader will be clicking the link to create the repository!!
`https://classroom.github.com/g/JxrCNON4`

</div>

Creating Your Repository

We will use a group assignment functionality of GitHub Classroom for this assignment. For group assignments **only one person will be creating the team while the other team member will join that team.** Please form a team of **no more than two people** and select one person to create the repository.

The selected person of the team should go into the link to the lab in the assignment sheet. Copy this link and paste it into your web browser. Now, you should accept the laboratory assignment and create a new team with a unique and descriptive team name (under "Or Create a new team").

Now the other members of the team can click on the assignment link and select their team from the list under "Join an Existing Team". When other team members join their group in GitHub Classroom, a team is created in our GitHub organization. Every team member will be able to push and pull to their teams repository.

To use this link, please follow the steps below.

- Click on the link and accept the assignment

- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab,

- Clone this repository (bearing your name) and work locally

- As you are working on your lab, you are to commit and push regularly. The commands are the following.

    - `git add -A`
    - `git commit -m ''Your notes about commit here''`
    - `git push`

Introduction

Imagine that that a coach for a major league baseball coach comes to you to ask you how you could help his team have a better strategy when playing baseball. As mentioned in Figure 1, this strategy would have to come by way of an analysis of data of seasons of game-play...

In baseball, a pitcher throws a ball to a catcher who stands behind a batter. The batter attempts to hit the flying ball with a bat before the catcher can trap the ball. To prevent the batter from hitting the ball and scoring points for his or her team, the pitcher attempts to throw the ball in a way that complicates the batter's ability to negotiate a ball-hit. The pitcher and the batter in this scene are the actors that perform in the pitcher-batter relationship play. This relationship may be one of the most important factors to determine a winning outcome in a game and so the types of pitchers and batters are especially chosen by each team to play each other where each has some form of presumed advantage over the other. Here we will spend some time to investigate this scene using data obtained from free-sources of baseball data online.

Work has been completed to compare the pitching biomechanics between left- and right-handed professional pitchers. In Solomite *et al.* [1] and Diffendaffer *et al.* [2] some of the kinematic and kinetic differences that exist between left- and right-handed pitchers were studied. In their work, both research groups they find that there were notable biomechanical differences stemming from elbow movements, in addition to foot and leg position, stance, shoulder movements and the position of other body parts, that contrasted between the pitches of left- and right-handed players.

On the other hand, left-handed batting in baseball has been shown to provide batters an advantage in winning batting titles over right-handed batters [3] and discussed in the *SportsRec* blog at `https://www.sportsrec.com/5950567/left-handed-batter-vs-a-right-handed-pitcher`. Interestingly, seven of the top 10, and 19 of the top 30, are left handed. Discussed in Walsh [4], most batters hit right-handed and the observation that a majority of these top-hitters (making the leaderboard ranks) hit left-handed is of great interest. This observation may be explained by the phenomenon that batters tend to hit better when facing an opposite-hand pitcher, however, further exploration needs to be completed in order to completely understand how this pitcher-batter *incompatibility* can be harnessed into a working strategy for baseball.

An Analysis of the Pitcher-Batter Relationship

The above discussion caters to a specific interest in baseball analytics where the batter-pitcher relationship has been reduced to factors of left- and right-handedness. There are likely to be many, many other types of investigations into baseball data which may explain other incompatibilities between the batter and pitcher and may help to inform strategies for placing batters in game against particular pitchers.

In this lab, your task is to investigate the **pitcher-batter relationship** using exploratory data analysis. In particular, you are to ask **deep, far-reaching**, questions of the data and to use your analysis skill in efforts to answer some of these questions. Note, that you will have to find this data and then examine it to be sure that your questions can actually be answered by the analysis of the data. Your analysis and results will then be used to inform the hypothetical coach (from above) to help him to consider the advantages of particular pitchers and batters, in effort to win games. Note: your work does not have to concentrate on the left- and right-handedness of players, and you are encouraged to pursue seemingly any type of investigation of your data to study the pitcher-batter relationship.

HANDED OUT: 24th FEBRUARY 2020

Links to Data Sources

You are to spend some time researching sources of data online which will be helpful to your investigation of the pitcher-batter relationship. Remember, if the questions cannot be answered from the dataset, then either change your approach to answering your over-all research question, or change the data set that you are using for your investigation. Below are some suggestions of sources of data. If you find another source of data, please be sure to cite the source.

- `http://www.seanlahman.com/baseball-archive/statistics/`

- `https://Baseball-Almanac.com`

- `https://sabr.org/how-to/statistical-databases-and-websites`

- `http://m.mlb.com/statcast/leaderboard#exit-velo,r,2019`

- `https://www.kaggle.com/`

Required Deliverables

1. Report, `writing/report.md`: Your report of three or more (well-thought out) research questions which can be answered by the data you have obtained. In this report, you are show and explain your plots that you used to respond to your questions using at least two variables from the study of the pitcher-batter relationship. You are to explain how the plots are able to study some part of this phenomenon behind your research question.

2. Dataset, `data/*`: You are to store your versions of data in this directory to preserve it for part2.

3. Source code, `src/analysis.r`: Your code that can be run to load the data files and to produce the plots of your work. Please add documentation to your code to help the instructor understand what it is doing on a line-by-line basis.

When you have finished, please ensure that the GitHub web site has your pushed work by visiting your repository at the site. Please see the instructor if you have any questions about assignment submission.

References

[1] M. J. Solomito, J. V. Ferreira, and C. W. Nissen, "Biomechanical differences between left-and right-handed baseball pitchers," *Sports biomechanics*, vol. 16, no. 2, pp. 143–151, 2017.

[2] A. Z. Diffendaffer, G. S. Fleisig, B. Ivey, and K. T. Aune, "Kinematic and kinetic differences between left-and right-handed professional baseball pitchers," *Sports biomechanics*, vol. 18, no. 4, pp. 448–455, 2019.

[3] D. M. Brown, Z. A. Poucher, M. Myers, J. D. Graham, and J. Cairney, "Sinister right-handedness provides canadian-born major league baseball players with an offensive advantage: A further test of the hockey influence on batting hypothesis," *PloS one*, vol. 14, no. 8, 2019.

[4] J. Walsh, "The advantage of batting left-handed," *Retrieved October*, vol. 25, p. 2010, 2007.