Significance and Effect of Handedness in Baseball

Matthew Loftus

Introduction

For my STAT 499 senior project, I looked at baseball data regarding handedness in batting versus handedness in pitching. It has been traditional in baseball to have batters in your lineup that swing from the side of the plate that is opposite from the pitcher's throwing hand. For example, if the starting pitcher is right-handed, the conventional wisdom is that it would be more beneficial for the team to start as many left-handed batters as possible. This idea stems from a few scientific reasons. First, when a pitch from a right-handed player goes to a right-handed batter, the ball seems to be coming directly at the batter, which may cause a slight hesitation. This slight hesitation against a pitch might be enough to disrupt the reaction time needed to hit it. The same pitch coming from the opposite side, however, gives the batter a better view across the body. Second, a curve ball from a same-handed pitcher will typically move away from the hitter, causing the batter to reach across the plate (Chu). These reasons may be why so many teams have left-handed batters on their rosters. Around 35% of all 2018 Major League Baseball (MLB) batters were left-handed, compared to 10% left-handers in the general population. My main goal for this project was to see if there is a significant difference in performance between batters facing a same-handed pitcher versus an opposite-handed pitcher.

Data Collection

I used data from the MLB main website (www.mlb.com) for every batter in the three most recent MLB regular seasons (2017, 2018, and 2019). The three key variables tested in this project were batting average, on-base percentage (OBP), and slugging percentage (SLG). Some important baseball terms are listed here:

At-bats: Plate appearances minus sacrifice flies, walks, and hit by pitches
Batting Average: Hits/At-bats
OBP: Percentage of time a batter ends up on base; i.e.,
(Hits + Walks + Hit by pitch)/Plate Appearances
SLG: The measure of the power of a hitter; i.e., Total Bases/At-bats

After collecting the data, I had 5766 data points. However, after doing initial graphical analysis of batting averages, I found that the data was not normally distributed, but rather right-skewed, and contained a large amount of batting averages equal to zero. This non-normality was because there were many data points with less than five at-bats, making it far more likely for a batter to have no hits and resulting in a batting average of zero. To fix this, I removed all data points with less than 30 at-bats, since that is roughly where batting averages stopped being zero. Additionally, I had many batters who were switch hitters, making it impossible to know when they were batting on the right or left side of the plate. To fix this, I created a new variable that coded switch hitters under the assumption that they would bat on the opposite side of the pitcher's dominant hand (ex. batting right when faced with a left-handed pitcher). After this cleaning, I had 2933 data points remaining, with a distribution of batter/pitcher handedness shown in the table below.

	BAT		
PITCH	L	R	Total
L	349 (11.89%)	904 (30.82%)	1253 (42.72%)
R	734 (25.03%)	946 (32.25%)	1680 (57.28%)
Total	1083 (36.92%)	1850 (63.08%)	2933

Methodology

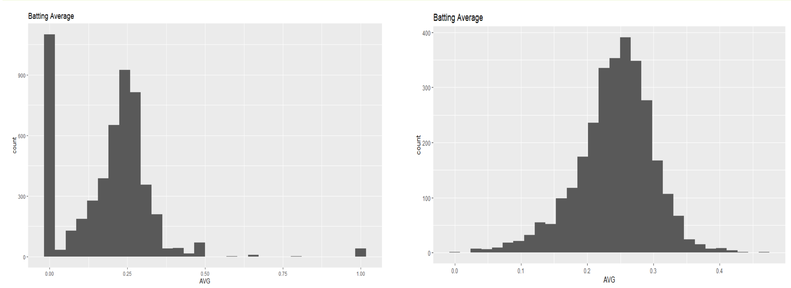
For each of the three key variables of batting average, on-base percentage, and slugging percentage, I ran a Two-Way ANOVA test using the statistical software R. Each Two-Way ANOVA test investigated the following questions:

1. Is there a significant difference between left-handed batters and right-handed batters?
2. Is there a significant difference between left-handed pitchers and right-handed pitchers?
3. Is there a significant interaction between the handedness of the batter and the handedness of the pitcher?

A p-value less than 0.05 was considered statistically significant. When checking the assumptions of the tests, the normality and constant variance assumptions were reasonably satisfied, but the independence assumption was not. This lack of independence was due to the fact that the data set contains multiple observations on the same players. Hence, the results that follow must be viewed cautiously due to the violation of the statistical assumptions.

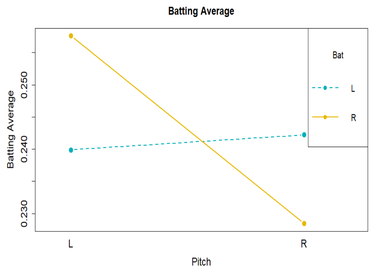
Batting Average Results

The histograms below display the distribution of batting average before (left) and after (right) data cleaning.



From the interaction plot below, we observe that right-handed batters have significantly better batting averages against left-handed pitchers than against right-handed pitchers, while left-handed batters have slightly better batting averages against right-handed pitchers than against left-handed pitchers. Overall, left-handed pitchers tend to do better than right-handed pitchers, while there is not a significant overall difference in batting averages between left- and right-handed batters.

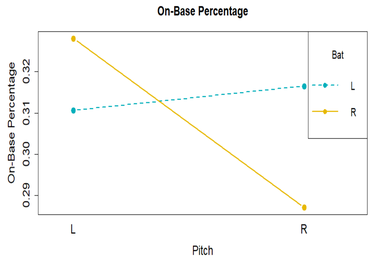
	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Pitch	1	0.238	0.23816	75.996	< 2e-16
Bat	1	0.002	0.00229	0.731	0.393
Pitch:Bat	1	0.155	0.15547	49.611	2.33e-12
Residuals	2929	9.179	0.00313		



On-Base Percentage Results

From the interaction plot below, we observe that right-handed batters have significantly better on-base percentages against left-handed pitchers than against right-handed pitchers, while left-handed batters have slightly better on-base percentages against right-handed pitchers than against left-handed pitchers. Overall, left-handed pitchers tend to do better than right-handed pitchers, and left-handed batters tend to do better than right-handed batters.

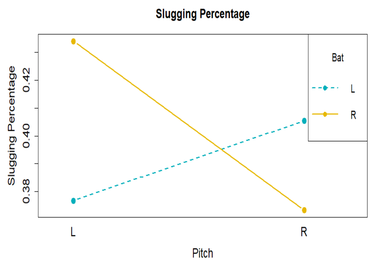
	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Pitch	1	0.388	0.3879	98.34	< 2e-16
Bat	1	0.090	0.0904	22.92	1.77e-06
Pitch:Bat	1	0.343	0.3426	86.86	< 2e-16
Residuals	2929	11.553	0.0039		



Slugging Percentage Results

From the interaction plot below, we observe that right-handed batters have significantly better slugging percentages against left-handed pitchers than against right-handed pitchers, while left-handed batters have slightly better slugging percentages against right-handed pitchers than against left-handed pitchers. Overall, left-handed pitchers tend to do better than right-handed pitchers, while there is not a significant overall difference between left- and right-handed batters.

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Pitch	1	0.68	0.6766	48.407	4.26e-12
Bat	1	0.00	0.0019	0.136	0.712
Pitch:Bat	1	1.25	1.2505	89.466	< 2e-16
Residuals	2929	40.94	0.0140		



Conclusion

There is a significant difference in performance between batters facing a same-handed pitcher versus an opposite-handed pitcher. Batting on the opposite side leads to better batting averages, on-base percentages, and slugging percentages. Interestingly, performance appears to be better for right-handed batters against left-handed pitchers than for left-handed batters against right-handed pitchers.

For future work, we could use a fuller data set of the breakdown of each batter against each pitcher. Even though this data set would still violate the independence assumption of Two-Way ANOVA, I could have performed a Two-Way ANOVA with repeated measures instead, which would account for the fact that the same players appear in the data set multiple times. We could also test the effect of handedness on other baseball statistics and use more seasons of MLB data.

References:
Chu, C. Y. C., Chang, T., & Chu, J. (2016). Opposite hand advantage and the overrepresentation of left-handed players in major league baseball. *Academia Economic Papers, 44*(2), 171-205.

Faculty Advisor: Professor Tiffany Kolba, Department of Mathematics and Statistics

Valparaiso University