UNITED STATES MILITARY ACADEMY

PROJECT PROPOSAL: BASEBALL

MA376: APPLIED STATISTICS

SECTION G2

DR. ANNY-CLAUDE JOSEPH

By

CADET AUDREY HAMILTON '21, CO A1

CADET JOSEPH ZUCCARELLI '21, CO F1

WEST POINT, NEW YORK

28 SEPTEMBER 2020

_____OUR DOCUMENTATION IDENTIFIES ALL SOURCES USED AND ASSISTANCE RECEIVED IN COMPLETING THIS ASSIGNMENT.

_____WE DID NOT USE ANY SOURCES OR ASSISTANCE REQUIRING DOCUMENTATION IN COMPLETING THIS ASSIGNMENT.

SIGNATURE: Joseph Zuccarelli Audrey Hamilton

Are Sluggers More Likely to Strikeout?

Audrey Hamilton and Joseph Zuccarelli

September 2020

Research Question

The objective of this study is to determine if there is a relationship between rate of strikeouts and homeruns for Major League baseball players. The response variable is the rate of strikeouts. The explanatory variable is the rate of homeruns.

Data

The observational units in this study are Major League Baseball players who played during the 2010-2019 seasons. There are 5353 complete observational units in the data set. Each unit represents a player's total regular season numbers. Note that the data set was created using the "Sean Lahman Baseball Database" in R.[1] The variables included in the study are as follows: *StrikeoutRate*, *HomerunRate*, *BB*, *H*. *StrikeoutRate* is a quantitative variable that is calculated by dividing the number of total strikeouts by the total number of at-bats. In this study, *StrikeoutRate* is the response variable. *HomerunRate* is a quantitative variable that is calculated by dividing the number of total homeruns by the total number of at-bats. In this study, *HomerunRate* is the explanatory variable. *BB* is a quantitative variable that represents the number of times that a player reached base on balls. In this study, *BB* is a confounding variable. *H* is a quantitative variable that represents the number of times that a player reached base on hits. In this study, *H* is a confounding variable. Hits and walks are the two confounding variables that have been identified that could affect ~~effect~~ the response variable. The following table summarizes the four variables that are included in the study, as well as their classifications.

Variable	Type	Classification	Units
StrikeoutRate	Response	Qualitative	$\frac{\text{Number of Strikeouts}}{\text{Number of At-Bats}}$
HomerunRate	Explanatory	Qualitative	$\frac{\text{Number of Homeruns}}{\text{Number of At-Bats}}$
H	Confounding	Qualitative	Number of Hits
BB	Confounding	Qualitative	Number of Walks

The following figure includes the first six observations in our data set.

	StrikeoutRate	HomerunRate	H	BB
1	0.2303665	0.034904014	146	87
2	0.2435233	0.005181347	45	4
3	0.4444444	0.022222222	12	10
4	0.3429395	0.046109510	89	37
5	0.2193548	0.012903226	28	5
6	0.2131148	0.032786885	18	3

Figure 1: 2010-2019 Batting Data

Author Contributions

1. CDT Zuccarelli proposed the research question, wrote the description on the response and explanatory variables, and created Figure 1 in R.
2. CDT Hamilton wrote the descriptions of the confounding variables and created the table that explains the variables.
3. Both CDTs worked on the proposal document, discussed the data set, created the works cited page, and established goals for the research project.

Works Cited

[1] Lahman, Sean. *Batting*. Lahman's Baseball Database, http://www.seanlahman.com/baseball-archive/statistics/. Accessed 25 September 2020.