

Data analytics effects in Baseball

Chenyu Mu
Department of Statistics
University of Connecticut

October 30, 2023

Abstract

Over the past two decades, there has been a notable resurgence in the utilization of data analytics across professional sports, businesses, and governmental sectors. This article will delve into an in-depth analysis of the myriad factors that potentially impact baseball players while they are actively engaged in the game. The study aims to explore and elucidate the diverse elements that might influence and shape the performance, strategies, and outcomes within the realm of baseball.

1 Introduction

Sport psychology, an interdisciplinary field integrating concepts from both psychology and sports science, has become a fundamental element in enhancing athletic performance across a spectrum of sports, baseball included. Its primary focus is to comprehend and optimize the psychological components that affect an athlete's mindset, emotions, and conduct, ultimately aiming to elevate their performance. Now, let's explore specific key domains within sport psychology that exert significant influence on baseball players. Baseball was selected as the focal sport due to its comprehensive nature, allowing for an analysis of nearly every facet of the game, except for the mental readiness of the player upon entering the field. Previous studies in the field [Dalmass \(2018\)](#) have provided valuable insights, and we seek to build upon their findings.

We measure performance in these two dimensions:

1. the player's subjective assessment of their performance.
2. the statistical analysis of the player's actual performance. Consequently, the research inquiries encompass the following:
 - Does sports psychology have an effect on a player's physical performance?
 - Does sports psychology have an effect on the way a player feels about his performance?
 - Can sports psychology be used to ensure that a player is almost always giving his best performance?

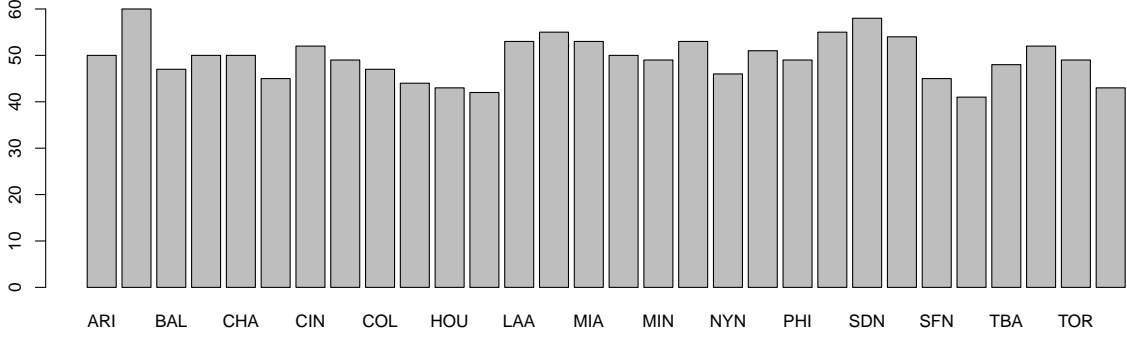


Figure 1: The number of players in each team in 2016 batting dataset

The rest of the paper is organized as follows.

The data will be presented in Section 2.

The methods are described in Section 3.

The results are reported in Section 4.

2 Data

Following are definitions we may use:

AB: At bats

AVG: Batting average

H: Hits

K: Killed, Strikeout

HR: Home Run

BABIP: Batting Average on Balls in Play

SF: sacrifice flies

$$AVG = H/AB, \quad (1)$$

which means the total number of hits divided by the total number of at-bats

$$BABIP = (H - HR)/(AB - K - HR - SF), \quad (2)$$

which states that a batter's average for balls that are put in play (excludes strikeouts, home runs, and sacrifice flies)

We can see the number of players in each team in Figure 1.

Table 1: The distribution of censored participants

Position	Assigned Group	Did Not Meet AB or IP	Did Not Play
Pitcher	Control	3	1
Pitcher	Experiment	2	1
Infield	Control	1	0
Infield	Experiment	2	2
Outfield	Control	1	0
Outfield	Experiment	1	0

3 Methods

The research was structured as a two-sample experiment utilizing a differences-in-differences analysis method. The participants were classified into three positional groups and then subdivided into either control or experimental groups. The outcomes were assessed using t-tests to analyze the differences in means between the groups. Method way [Dalmass \(2018\)](#).

4 Results

These limitations encompassed various factors, including but not restricted to the sample size, the participants' in-season playing time, and the number of at-bats and innings pitched. The relatively small sample size utilized in the study, which significantly impacted the practicality of the results [1](#).

The study involved recruiting 26 members from the baseball team. Among these participants, 0.385, were pitchers, 0.423 were infielders, and 0.192 were outfielders. Splitting these groups, half of the pitchers were in the control group, while the remaining half were in the experimental group. For the infielders, 0.455 were in the control group, and 0.545 were in the experimental group. As for the outfielders, 0.4 were in the control group, and 0.6 were in the experimental group. Due to limitations in the sample size—considered small—the study's reliability may have been affected since it was conducted as a convenience sample. While the recruitment method was straightforward as it targeted the university's varsity team, the size of the sample could potentially impact the study's outcomes. Notably, the study would have ideally required a minimum sample size of 128, calculated using a specific formula considering an alpha level of .05, t-tests for independent groups, an effect size of 0.5, and a power of 0.8. Implementing this study on a larger scale, such as within a Minor League Baseball organization, would have made achieving the required sample size more feasible.

References

Dalmass, K. (2018). The psychology of baseball: How the mental game impacts the physical game. *Honors Scholar Theses* 578, 99.