MLB Team Statistics between 2000 - 2023 on Predicting Games Won per Season[1]

Tracy Yang

April 12, 2024

Abstract   Using MLB key team statistics between 2000 and 2023, this paper investigates the outcome of the 2024 season using each team's current statistics on predicting their total wins out of 162 games. We also investigate the common conception that baseball is a numbers game by evaluating the predictability of wins that a team would obtain during a season from looking at their batting average (AVG), defensive efficiency ratio (DefEff), and earned run average (ERA).

Table of contents

---

[1]Code and data are available at: https://github.com/YcartXin/MLB_Team_Stats

1 Introduction

Behind the entertaining personalities and eye-catching traditions of baseball is a game where teams and are meticulously crafted with calculation. Baseball's fascination with numbers persisted since its creation (Swhwarz, 2004). Runs scored have always been recorded to evaluate performance, while other statistics developed overtime into the complicated net of data that now reside upon the MLB official site (XX). The term "sabermetrics" even developed to describe in specific baseball statistics (XX) However, a team's performance can differ greatly year to year. A team's last year wins are not often indicative of the coming season. Many factors play into the seasonal performance of a team including more unpredictable variables such as injuries, and luck. However, with the common narrative of baseball being a numbers game, we aim to estimate the impact of important statistics of a team on its games won per season, more specifically, the progressing 2024 season.

Important statistics for a team in this paper is defined as the team's batting average (AVG), defensive efficiency ratio (DefEff), and earned run average (ERA). These three statistics each refer to an important aspect of a strong team: hitting, pitching, and fielding prospectively. A baseball team cannot succeed with a weak end in any of the three categories. Therefore, these are important indicators on the wins a team is predicted to earn in a season. The specifics of these indicators will be elaborated on in Section 2. Among present research, 2024 season predictions exist, however, they mainly focus on the win for a game or series for betting purposes. This paper is unique in its approach to predict and investigates the predictability of overall seasonal wins for baseball teams.

The rest of the paper is structured as follows. Section 2 (Data) will explain the source of the data used, its collection process, and the process of cleaning that led to the final data used. The data section will also explore the potential biases the collected and final data may possess. Section 3...

The overall analysis shows...

2 Data

Talk more about it.

And also planes. (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [Appendix].

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

3.1.a Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in .

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Weaknesses and next steps

Weaknesses and next steps should also be included.

6 Additional data details

7 Model details

7.1 Posterior predictive check

In 1 we implement a posterior predictive check. This shows...

In 2 we compare the posterior with the prior. This shows...

7.2 Diagnostics

is a trace plot. It shows... This suggests...

is a Rhat plot. It shows... This suggests...

8 References

Bibliography