Bayesian Analysis of Batting Averages: Player Performance and Career Trajectories

Author: Armin Bazarjani
Presentation format: Paper

Background

This project aims to analyze how Major League Baseball (MLB) batting averages evolve throughout a player's career using Bayesian methods. Specifically, we will investigate whether players show consistent improvement in their early career years before plateauing or declining. Baseball statistics provide an ideal context for Bayesian analysis as we have clear prior information from league-wide performance metrics and can update these beliefs with player-specific data.

Traditional baseball analytics often use simple moving averages or frequentist methods to analyze player performance. However, a Bayesian approach can better account for both league-wide patterns and individual player trajectories, while providing more intuitive probability statements about player ability.

Data Description

The analysis will use the Lahman Baseball Database, a comprehensive open-source database containing MLB statistics from 1871-2022. This is a comprehensive database that contains just about any baseball statistic that one would need for analysis across all of the players and all of the teams

Prior Study Results

A prior study by Bailey et al. (2020) reported PECOTA (Player Empirical Comparison and Optimization Test Algorithm) results for player batting averages.

Method	MAE	MAE CI	ME	Avg	Sd	5th Perct	95th Perct
0.0209	0.0209	(0.021,0.021)	-0.0017	0.261	0.018	0.234	0.291

Statistical Models

I am thinking about implementing three models.
1. Basic Beta-Binomial Model - predict a single players "true" batting ability.
2. Hierarchical Model - predict players batting ability while also allowing for information from other players to be incorporated into the model as well.

Github Repository

The data analytic scripts and supplemental materials for this project will be available at: https://github.com/ArminBaz/573-project

References:

Bailey, S. R., Loeppky, J., & Swartz, T. B. (2020). The prediction of batting averages in major league baseball. *Stats*, *3*(2), 84-93.

Lahman, S. (2022). Lahman's Baseball Database . Available from http://www.seanlahman.com/baseball-archive/statistics/