

Estimating Aging Curves: Using Multiple Imputation to Examine Career Trajectories of MLB Offensive Players

Quang Nguyen

Department of Statistics and Data Science
Carnegie Mellon University
Pittsburgh, PA 15213
nmquang@cmu.edu

Gregory J. Matthews

Department of Mathematics and Statistics
Loyola University Chicago
Chicago, IL 60660
gmatthews1@luc.edu

2022 CMSAC Reproducible Research Competition

Abstract

In sports, an aging curve depicts the relationship between average performance and age in athletes' careers. This paper investigates the aging curves for offensive players in the Major League Baseball. We study this problem in a missing data context and account for different types of dropouts of baseball players during their careers. In particular, the performance metrics associated with the missing seasons are imputed using a multiple imputation model for multilevel data, and the aging curves are constructed based on the imputed datasets. We first perform a simulation study to evaluate the effects of different dropout mechanisms on the estimation of aging curves. Our method is then illustrated with analyses of MLB player data from past seasons. Results suggest an overestimation of the aging curves constructed without imputing the unobserved seasons, whereas a better estimate is achieved with our approach.

Keywords aging curve · baseball · multiple imputation · statistics in sports

1 Introduction

Supplementary Material

All code for reproducing the analyses in this paper is publicly available at

2 References