How Does Travel Effect Baseball Performance

Brandon Zink, Cameron Connor, Abiel Fattore

Description

An MLB Baseball season consists of 162 games, 81 of those played at away stadiums in series, making for ~26 total series at away stadiums.

We're going to look at the effect that distance of travel and number of time zones changed has on both batting and pitching performance, across a number of basic (BA, OBP, ERA, etc.) and advanced (wRC, WAR, RAR, RA9, OPS+, etc.) statistics.

Prior Work

There have been two found studies done in this area, looking solely at the number of time zones crossed, but only looked at basis statistics (BA, ERA).

Both studies concluded that eastward travel had a greater effect than westward , but nothing based on total distance/time zones crossed.

Datasets

We will be using Retrosheet Events data, which has the outcome of every AB in the last 20 years (including number of pitches, situation, score, where the ball was hit, what kind of hit, etc.)

http://www.retrosheet.org/

Proposed Work

Cleaning - removing unnecessary columns

Processing - calculating distance traveled/time zones crossed, aggregating statistics based on travel information, calculating advanced statistics including ballpark factor (the variation accounted for by difference in ballpark altitude/size)

Integration - Using probability models to see if the differences we find are relevant, interpreting in game effects by team (do the Yankees have a noticeable advantage over Mariners since they travel less), calculating expected drop in player performance based on new team.

Tools

SQL/MySQL Workbench for data gathering, Python/Jupyter Notebooks for math analysis, potentially Google Maps API for distance calculations.

Evaluation

We will use Hypothesis Testing in order to determine if there is a significant change in performance by travel distance, as well as create a model to predict a change in player performance based on the distance traveled before a game or over a season.