Major League Baseball

How Has the Game Evolved Over Time?

Introduction

What metrics most closely correlate to a strong baseball player? How can we find the most relevant gauge to predict which players are best?

Using publicly available data from the MLB and Kaggle, players' statistics can be analyzed to understand which statistics are the most important when drafting new players or trading between teams. Such techniques were used by the Oakland A's in 2002, as portrayed in the movie, Moneyball starring Brad Pitt. Since the team was strapped for cash, they turned to alternate ways of understanding how a player can be developed and the potential each player has based on statistics that may go overlooked. In recent years, statistics in the MLB have drawn more attention from researchers and mathematicians. The MLB stats are much more advanced than that of any other professional sports league because of the massive amount of variables being tracked. These variables allow for composite scores to be computed, such as Fangraph's, fWAR (wins above replacement). One of baseball's most advanced and comprehensive statistics, fWAR (and other WAR formulas by other statistic agencies), takes into account both the offensive and defensive statistics of a player and compares that to a normal player of that position. The composite score is analyzed to calculate how many more wins the player has given their team compared to that of a normal player in that position.

The goal of my project is to analyze the evolution of the league over time. In seeing the growth of the league and the betterment of players becoming more competitive, I would like to research which parts of the game have become more important. I can analyze the correlation between certain metrics, and how they translate to a more efficient player. I can use charts to analyze the average statistics by year and I can also calculate that correlation with newer statistical formulas being used in the league to define which is the most efficient and thus, finding the best way to measure a player.

Data Sources