Quinnipiac UNIVERSITY

Baseball Stats and WAR

Chris Callender

2023 Mathematics Senior Project

Problem Statement

Many people question who is the greatest baseball player of all time. We will answer this question by looking at some of the greatest baseball players since 1900 and seeing who is the best. Also, since many people say WAR is the most valuable statistic, we will use Principle Component Analysis (PCA) to determine whether that is true. WAR is wins above replacement and it calculates the number of wins a player contributes to his team. It compares a player to a replacement level player by considering the player's number of runs above average, the number of runs provided by a replacement player, as well as adjustments for league and position.

1 Results

This table shows the top 10 players in points received, based on my test. I sorted each statistic from highest to lowest and the leader received 10 points, second place received 9 points, and so on.

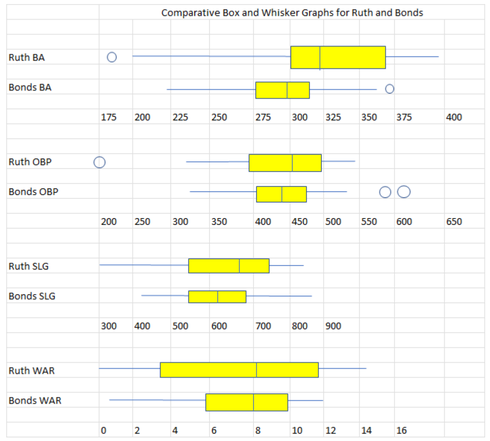
Player	Points Received
Babe Ruth (RF)	118
Josh Gibson (C)	89
Barry Bonds (LF)	84
Ty Cobb (CF)	72
Ted Williams (LF)	57
Oscar Charleston (CF)	55
Hank Aaron (RF)	48
Lou Gehrig (1B)	45
Rickey Henderson (LF)	43
Cap Anson (1B)	26

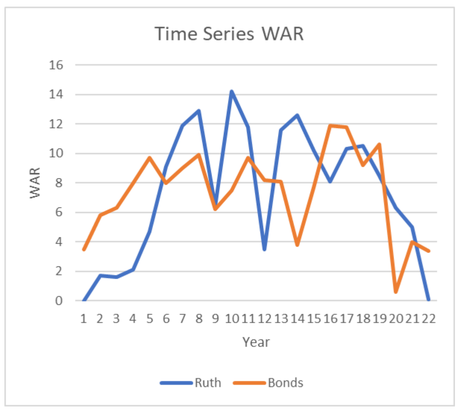
The following table shows the top 10 players in career WAR.

Player	Career WAR
Barry Bonds	162.8
Babe Ruth	162.7
Willie Mays	156.1
Ty Cobb	151.4
Hank Aaron	143.0
Tris Speaker	134.8
Honus Wagner	130.8
Stan Musial	128.6
Rogers Hornsby	127.3
Eddie Collins	124.4

2 Ruth and Bonds Comparison

It seems clear that Ruth and Bonds are the greatest players of all time, since they are in the top 3 in points received and have the two highest career WARs. So, I decided to compare the two players to see who is better.





I believe that Babe Ruth is the greatest player of all time. Ruth and Bonds are the two greatest players of all time, but I think Ruth is slightly better than Bonds.

3 Principle Component Analysis

Now we will use PCA to determine which of the criteria primarily differentiates the players.

The first 6 PCs explain $\frac{20.9}{22} = .95 = 95\%$ of the variation. So, we will retain the first 6 PCs. That makes k, the number of PCs retained, relatively small while also explaining a large amount of the variation. We want k to be as small as possible, but we want the variance explained by the first k PCs to be large.

Here is the eigenvector for the first PC. It explains 36.9% of the total variance.

$$e_1 = \begin{bmatrix} .250 \\ .243 \\ .286 \\ .157 \\ .313 \\ .203 \\ .272 \\ .177 \\ .236 \\ .260 \\ .172 \\ .224 \\ .193 \\ 0 \\ .298 \\ .154 \\ 0 \\ .316 \\ .217 \\ .129 \\ 0 \\ 0 \end{bmatrix}$$

The first PC explains the most variation and most heavily weights the variable WAR since .316 is the greatest component of e_1. That means that WAR primarily differentiates the players. This supports my previous belief that WAR is the most valuable statistic for baseball players. WAR encompasses every aspect of the game, such as batting, running, and fielding, so that may be why it is so useful.

4 Conclusion

I believe that Babe Ruth is the greatest player of all time. However, there are nine players in a lineup, so I will create what I believe is the all time all star baseball lineup.

- **C** Gibson
- **1B** Gehrig
- **2B** Hornsby
- **SS** Rodriguez
- **3B** Beltre
- **LF** Bonds
- **CF** Cobb
- **RF** Ruth
- **DH** Williams

Bench

- Aaron
- Charleston
- Trout
- Judge

The bench consists of players that were not the best at their position, but were very close. I feel that although they are not the best, they still deserve recognition, so that is why I put them on the bench.

5 References

- Rothman, Stanley. *Sandlot Stats: Learning Statistics with Baseball.* Baltimore, The Johns Hopkins University Press, 2012.
- "MLB Stats, Scores, History, & Records." *Baseball*, `https://www.baseball-reference.com/`.
- "Career Leaders & Records for WAR Position Players." *Baseball*, `https://www.baseball-reference.com/leaders/WAR_bat_career.shtml`.
- Penn State University. "Lesson 11: Principle Component Analysis (PCA)." `https://online.stat.psu.edu/stat505/lesson/11`.