**Goals for Achievement 6 and Project Premise**

In professional sports, athletes are signed to staggering multi-million-dollar contracts. For better or worse, what was once the exception, has now become the rule. In recent years, the dollar amounts athletes earn becomes greater and greater and doesn’t show a sign of stopping.

For example, in 2023 NFL quarterback, Lamar Jackson of the Baltimore Ravens, signed a five-year contract extension of $260,000,000 with a signing bonus of $72,500,000. This deal has made him the highest paid NFL player in history and came as a shock to many football enthusiasts. One question was if one player can truly be worth that much money to an organization.

However, in the world of Major League Baseball, these types of contracts are nothing new and have been normal for superstar players for years. In baseball, players have been signed to these deals for anywhere from two to ten years.

The **goal** of this project is to determine if Major League Baseball players who have been signed to contracts worth over fifty million dollars produce the results that they are assumed to. In short, does an “all-star paycheck” correlate to an all-star player.

How is a successful player measured? In post-Moneyball baseball, the game has almost run rampant with statistics. Stats like BAIP, LIPS, wRC+, and ISO are just a few advanced statistics that have become recent additions to the game. Baseball has come a long way from simple ERAs and batting averages.

For this project, I’ve decided to look specifically at the WAR, or wins above replacement stat. I’ve determined this statistic to be the most holistic approach to assessing player value.

This is an excerpt from the Major League Baseball official website that explains the function of the WAR statistic in baseball concisely:

**Wins Above Replacement (WAR)**

**Definition**

WAR measures a player's value in all facets of the game by deciphering how many more wins he's worth than a replacement-level player at his same position (e.g., a Minor League replacement or a readily available fill-in free agent).

For example, if a shortstop and a first baseman offer the same overall production (on offense, defense and the basepaths), the shortstop will have a better WAR because his position sees a lower level of production from replacement-level players.

**The formula**

For position players: (The number of runs above average a player is worth in his batting, baserunning and fielding + adjustment for position + adjustment for league + the number of runs provided by a replacement-level player) / runs per win

For pitchers: Different WAR computations use either RA9 or FIP. Those numbers are adjusted for league and ballpark. Then, using league averages, it is determined how many wins a pitcher was worth based on those numbers and his innings pitched total.

Note: fWAR refers to Fangraphs' calculation of WAR. bWAR or rWAR refer to Baseball-Reference's calculation. And WARP refers to Baseball Prospectus' statistic "Wins Above Replacement Player." The calculations differ slightly -- for instance, fWAR uses FIP in determining pitcher WAR, while bWAR uses RA9. But all three stats answer the same question: How valuable is a player in comparison to replacement level?

**Why it's useful**

WAR quantifies each player's value in terms of a specific number of wins. And because WAR factors in a positional adjustment, it is well suited for comparing players who man different defensive positions.

Source: [MLB Advanced Stats, Wins Above Replacement](https://www.mlb.com/glossary/advanced-stats/wins-above-replacement)

**Data Source**

This data is internal data from Baseball Reference, which is a segment of Sports Reference, a group that, “democratize[s] data, so our users enjoy, understand and share the sports they love.” Sports Reference is one of the most reliable sources of sports data in the industry and is recognized by major media organizations and broadcasters as a source for statistics.

While the data I used for this project did not come from a ‘pre-fab’ dataset, all the data is open source with some compiling and collection.

**Data Profile**

*All player statistic data is up to date as of 08/15/2023*

*All contract data is up to date as of 07/31/2023*

Player Statistics Data: Used to find player demographics and player statistics

Player Salary and Contract Data: Used to find the salaries and contract information on players.



All Star Teams Data: Used to find the value of Bambino 2.0, the metric for success in this project.

**Limitations and Ethics**

Content: A limitation is that part of the WAR statistic is dependent on calls made on the field during game play by umpires. Umpires are human and are subject to error and bias. For example, calling a player safe when they’re out, vice versa, calling a ball a strike, etc. There is some difference in the calculation of WAR for pitchers and position players so this could potentially create different “replacement” assessments.

Source: This data comes from reputable sources within the baseball industry. As it is purely statistical data that is collected from game results and contractual information from legal documents that are public records, I don’t believe that there would be any discernable amount of bias.

Collection: As I collected this data, and I have never sourced my own data before, it’s very possible that there is some form of collection bias or oversight.

Bias: I am a baseball fan(atic) so there could be a form of preferential bias.

Ethics: As this data is open-source I don’t think there is an issue of data ethics.

**Questions to Explore**

* Is there a difference in WAR between players in the American League or the National League?
* Do pitchers or position players have a greater WAR value? Is there any discernable difference between these players at all? If there is, explore this topic. (Do pitchers, infielders, outfielders, designated hitters have greater WAR?)
* Does player age affect WAR? Or, does a player’s WAR value change as they get older?
* Does a larger (fiscal) contract correlate to a greater WAR value?
* Is there a tendency for certain teams to have players with greater WAR value? If so, which ones and what are possible explanations for this?
* Does player injury influence WAR? Are there any players who seem to be more prone to injury?
* Is there any correlation between WAR and laterality? (If a player is righ-handed, left-handed, etc.)
* Is there any connection between WAR and country of origin?

**Sources and References:**

Baseball-Reference. (n.d.). *List of all of the baseball players: Major League players*. Baseball-Reference.com. https://www.baseball-reference.com/players/

Baseball-Reference. (n.d.). *Baseball-Reference.com War explained*. Baseball-Reference.com. https://www.baseball-reference.com/about/war\_explained.shtml

*Wins above replacement (WAR): Glossary*. MLB.com. (n.d.). https://www.mlb.com/glossary/advanced-stats/wins-above-replacement

A very special thank you to Baseball Reference for making player data available to the public!!!