Lab 1: Example of the Research Process

Instructor: Jonathan Pipping   *Author: Ryan Brill*

1.1 Getting Started

Suppose you want to start a sports analytics research project, but you don't have any ideas yet. Where do you start? A fantastic way to get started is to simply read about something you're interested in.

1.1.1 Wins Above Replacement

Perhaps you were listening to a podcast or reading an article and came across an interesting fact: Roger Clemens has the most career **Wins Above Replacement (WAR)**, 133.7, in Major League Baseball history. You may also have heard that Pedro Martinez recorded the highest-ever single-season WAR, 11.6, according to FanGraphs.

You may think that WAR is a pretty cool concept, and it makes some intuitive sense as to why it seems like a good way to evaluate pitchers, and more broadly, all players.

Definition 1.1 (Wins Above Replacement (WAR))**.** *Replace a player with a replacement-level player (e.g., the best player you could get off waivers). How many fewer wins would the team have, assuming average teammates and opponents?*

Implementation. Take a player's observed performance, ignoring/adjusting for things they're not responsible for, and map that to wins.

You may not know the math behind WAR, but since you're interested you **read about it**. The most widely-used public WAR implementations are from FanGraphs and Baseball Reference, so we should read up on those.

1.1.2 WAR for Pitchers

When you read about WAR for pitchers, a few things might catch your eye:

1. WAR involves mapping a pitcher's performance to wins. FanGraphs uses Fielding Independent Pitching (FIP), while Baseball Referance uses Expected Runs Allowed (xRA).

   - Fangraphs uses a version of FIP called **ifFIP** to calculate WAR. This variation includes home runs, walks, hit-by-pitches, strikeouts, and infield fly balls. The formula is:

   $$\text{ifFIP} = \frac{13HR + 3(BB + HPB) - 2(K + IFFB)}{IP} + C$$

   where C is a constant that depends on league factors. More specifically,

   $$C = \text{lgERA} - \text{lgifFIP}$$

   where lgERA is the league ERA and lgifFIP is the league ifFIP.

- Baseball Reference calculates WAR using **xRA**, which ignores event ordering (e.g. 1B, out, out, 1B, HR, out vs HR, 1B, 1B, out, out, out) and just uses the events (1 HR, 2 1B, 3 out) to calculate the expected runs allowed of each inning.

2. There are a series of convoluted adjustments on top of the base performance metric (e.g. league adjustment, team defense adjustment).

3. WAR involves mapping a pitcher's performance **averaged** over the entire season into wins.

    - ifFIP divides by IP
    - xRA divides by seasonal xRA

Thoughts: averaging pitcher performance over the course of a season seems odd. We should explore the implications of these modeling assumptions.

1.2   Identifying the Problem

1.2.1   Example 1: Max Scherzer

In Table 1.1, we can see Max Scherzer's performance over six games prior to the 2014 All-Star Break. From everything we know about baseball, having four dominant performances (≤ 1 run allowed) should translate into ≥ 4 wins across the 6 games. However, if we average his performance over the 6 games:

$$\frac{15 \text{ runs}}{41 \text{ IP}} \times 9 \frac{\text{innings}}{\text{game}} = 3.66 \frac{\text{runs}}{\text{complete game}}$$

$$\text{And } 3.66 \frac{\text{runs}}{\text{complete game}} \approx 0.55 \frac{\text{win prob.}}{\text{complete game}}$$

Over the course of 6 games, this would translate into ≈ 3.30 wins, a big discrepancy from the 4+ wins we would expect. Let's look at another example.

game	1	2	3	4	5	6	total
earned runs	0	10	1	2	1	1	15
innings pitched	9	4	6	7	8	7	41

Table 1.1: Max Scherzer's performance over six games prior to the 2014 All-Star Break.

1.2.2   Example 2: Hypothetical Pitchers

Assume we are choosing between two pitchers, A and B, with equal ERAs.

- Pitcher A allows 5 runs per complete game.
- Pitcher B alternates between allowing 10 and 0 runs per complete game.

Which pitcher would you rather have? The answer is obvious: pitcher B. However, existing WAR methodologies would value these 2 pitchers the exact same. This highlights a problem with existing WAR methodology.

1.2.3   Example 3: Inconsistent Pitchers

Once again, we are choosing between two pitchers, except this time they are both inconsistent.

- Pitcher A alternates between allowing 7 and 0 runs per complete game.

- Pitcher B alternates between allowing 14 and 0 runs per complete game.

Existing WAR methodologies would see that pitcher A averages 3.5 RA/CG, pitcher B averages 7 RA/CG, and therefore value pitcher A way more than pitcher B.

However, a knowledge of baseball tells us that both pitcher A and pitcher B win half of their games, so they should be valued roughly evenly. What does this teach us?

- You can only lose a game once!

- Not all runs allowed have the same value!
    - The 8th run allowed in a game is "worth" far less than the 1st
    - The marginal difference in the win probability between allowing the 7th vs 8th run is less than the marginal difference in win probability between allowing the 1st and 2nd run, since in the first scenario you've already lost the game
    - If $R \mapsto \mathrm{WAR}(R)$ is game WAR as a function of runs allowed, then $\mathrm{WAR}(R) - \mathrm{WAR}(R-1)$ should get smaller as R gets bigger. Mathematically, this means that $R \mapsto \mathrm{WAR}(R)$ should be convex.

1.2.4   What's the Problem?

The Problem: Averaging pitcher performance over the course of the season is clearly wrong. Why?

- It ignores the game-by-game variance in pitcher performance.

- $R \mapsto \mathrm{WAR}(R)$ should be convex (i.e. not all runs should have the same value).

- A **win** is the fundamental result of a **game**, not a season.

Goal: Fix this problem, beginning with **one incremental improvement**. That's **RESEARCH**.

We can begin by calculating historical WAR in each invidual game, then setting seasonal WAR as the sum of game WAR. Then the **Task** is to calculate game WAR for starting pitchers.

1.3   Game WAR for Starting Pitchers

How do we do this? Let's break WAR down into its components.

- Wins (W): How many wins did a starter contribute in this game?

- Above Replacement (W_{rep}): How many wins would a replacement-level pitcher have contributed?

We put these together to get the formula for Game WAR:

$$\text{Game WAR} = W - W_{rep}$$

Let's start with the first component: mapping a pitcher's performance to wins. Mathematically, we have some notion of that in the form of **win probability**, but we also need to ensure proper **pitcher valuation**: judging a pitcher only on things he's responsible for. What goes into this?

1.3.1   Evaluating Pitcher Performance

A starting pitcher's in-game performance is a function of the following:

- Runs Allowed (R): The number of runs scored against the pitcher.

- Exit Inning (I): The inning in which a pitcher exits the game. Indicates how long the pitcher pitched.

- Exit Outs (O): The number of outs when a pitcher is removed. Indicates how deep into the inning the pitcher pitched.

- Exit Base-State (S): The number of outs when a pitcher is removed. Indicates if the pitcher left the game with runners on base.

Of these, runs allowed is clearly the most important, because winning a game is defined by Runs. However, there are a number of confounders (or variables outside of his control that affect his performance) that we should be aware of:

- Park: Major League parks are not all the same dimensions, and the climate often differs greatly between cities. Both of these have a significant impact on the number of runs scored.

- Opposing Team's Batting Quality: Facing a lineup with many good hitters (e.g. Yankees, Dodgers) should result in more runs scored than facing a lineup with many bad hitters (e.g. Chicago White Sox).

- His Team's Fielding Quality: The quality of the defense behind the pitcher can have a significant impact on the number of runs scored.

Finally, there are contextual variables that affect win probability at any point in the game:

- League (AL/NL): The quality and different rules of the American and National Leagues can have a significant impact on the number of runs scored.

- Year: Year-to-year differences in the balls/bats used, rule changes, average player quality, and more can have a significant impact on the run-scoring environment.

Note that some variables do **not** affect a pitcher's performance, so we should make sure to explicitly omit them:

- His Team's Batting Quality: The quality of a pitcher's offense is not under his control.

- Opposing Team's Defense: The quality of the defense behind the other pitcher(s) is independent of Pitcher A's performance.

Now that we're aware of how to measure a pitcher's performance and the confounders we need to account for, we can move onto the next step: defining a function that maps performance to wins.

1.3.2   Mapping Performance to Wins

It can be tempting to try to tackle this entire problem at once, but that's a recipe for failure. Instead, we **start simple**, beginning with the easiest version of the task and iterating from there. Let's start with **just** observed performance, then adjust for confounders later.

Task: Given a pitcher's performance in a game, what's his team's win probability when he exits the game?

- Context-neutral: Assume league-average offenses, defenses, and ignore his team's runs scored.

- Start simple: Assume he finishes the inning, so ignore O and S.

We then model the function $f(R, I)$; assuming both teams have league-average offenses, compute the probability a team wins a game after giving up R runs through I complete innings.

Since $f(R, I)$ can be visualized as a 2D grid, we name this metric **Grid WAR**.

This is the simplest version of the question, and it is nontrivial. We now have the foundation of a research project.

1.4   Takeaways

In general, there are two great ways to begin research in applied statistics (especially in sports).

1. "Read First": Read a paper/article/blog post about sports statistics. Replicate it, then check what else has been done and replicate the state of the art. Then, identify one thing you don't like and make an incremental improvement.

    - ex: Grid WAR

2. "Think First": Think of a cool idea, read relevant literature and replicate the state of the art (if any), then solve the problem.

    - ex: Entropy-Based Strategies for Multi-Bracket Pools

References

[BR]   Baseball Reference, *Pitcher WAR Calculations and Details*, Baseball Reference, 2024.

[FG]   Slowinski, P., *Calculating WAR for Pitchers*, FanGraphs, 2012.

[GWAR]   Brill, R. & Wyner, A.J., *Introducing Grid WAR: rethinking WAR for starting pitchers*, JQAS, 2024.

[MMB]   Brill, R., Wyner, A.J. & Barnett, I., *Entropy-Based Strategies for Multi-Bracket Pools*, Entropy, 2024.