3080 Project Part 2

Adam Chow

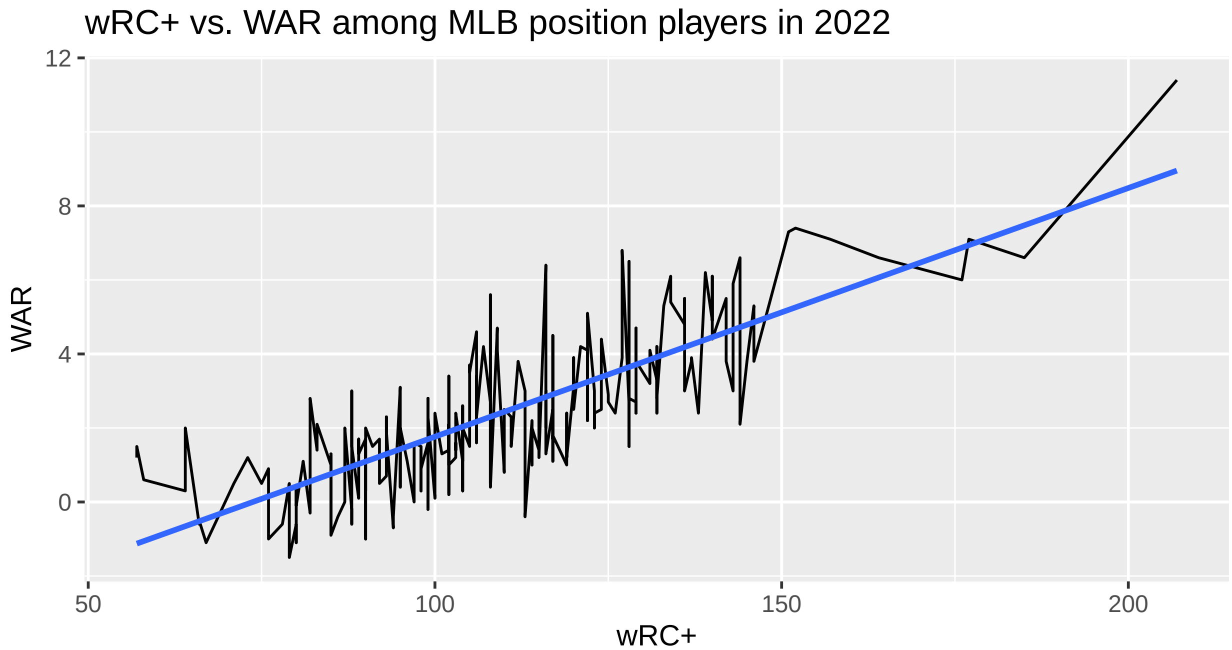
2023-05-03

Introduction

Out of the three research questions I explored in part 1, I have now solidified which question I want to explore in further detail: The Major League average wRC+ is 100. Do players above this mark have an overall higher WAR than hitters lower than this mark? I decided that this question is the most relevant when looking at the overall purpose of my study: which is to provide value and statistical insight to Major League Baseball owners and General Managers who spend hundreds of millions of dollars each year on their teams in hopes of winning a championship. In order to reach this goal, each respective team must win the most amount of games possible in order to eventually qualify for the postseason, only then will each team have the opportunity to compete for a World Series Trophy. In history, it has been a highly congested topic on what batting metric is the best overall indicator of a player's contribution to winning, Win Above Replacement (WAR). Among the best is often Weighted Runs Created Plus (wRC+), which calculates how many runs a player produces while factoring in the effect of each ballpark, since each one is unique and has different advantages (Pate, 2022).

Relationship between wRC+ and WAR among MLB position players in 2022

```
‘geom_smooth()‘ using formula = ’y ~ x’
```



Methods/Analysis

To test whether players with a wRC+ over the league average (100) have a higher overall WAR in a baseball player, I will be conducting a one sided, two sample t-test. Because both wRC+ and WAR are numerical, I must choose a quantitative test. My research question also involves the **mean** wRC+ meaning that I must use a parametric test. I chose to use a two sample test because I will be splitting my data into two independent samples, one for the data with wRC+ over 100 and one lower than 100. These samples are not dependent on one another. With only the population mean and no standard deviation, a t-test is my best test for testing my question. However, to use a t-test, it requires a normal population distribution. This requirement holds true because "wRC+ quantifies run creation and normalizes it" (MLB, 2023). My hypotheses are as follows:

$H_0 : \mu_h = \mu_l$

$H_a : \mu_h > \mu_l$

*where μ_h is the mean WAR of players with wRC+ above 100, and μ_l is the mean WAR of players with wRC+ below or equal to 100

Results

```r
higher_WAR <- mlb_data$WAR[mlb_data$wRC. > 100]
lower_WAR <- mlb_data$WAR[mlb_data$wRC. < 100]
t.test(higher_WAR, lower_WAR, mu=0, alternative="greater")
```

```
    Welch Two Sample t-test

data:  higher_WAR and lower_WAR
t = 11.765, df = 200.91, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.155409      Inf
sample estimates:
mean of x mean of y
3.3224806 0.8148649
```

The test statistic of 11.765 indicates the magnitude of the difference between the two means in terms of the standard error. The degrees of freedom of 200.9 indicate the precision of the estimate. Because of the extremely low p value, we do not have strong evidence to prove that the null hypothesis is true. Therefore, we reject the null hypothesis, there is not enough evidence to say that the mean WAR of the players with a wRC+ above 100 is equal to the mean WAR of players with a wRC+ lower than 100.

Conclusions

```
pwr.t2n.test(d=0.2, n1=129, n2=74, sig.level=0.05, alternative="two.sided")$power
```

```
[1] 0.2763562
```

```
pwr.t2n.test(d=0.5, n1=129, n2=74, sig.level=0.05, alternative="two.sided")$power
```

```
[1] 0.9267956
```

```
pwr.t2n.test(d=0.8, n1=129, n2=74, sig.level=0.05, alternative="two.sided")$power
```

```
[1] 0.999767
```

In the context of my problem, the results of my two-sample one-sided t-test show that there is a significant difference between the mean WAR of players with wRC+ above 100 and players with wRC+ below 100 ($t = 11.765$, $df = 200.91$, $p < 0.00001$). We know from the MLB that wRC+ is a normalized statistic and both of these samples are independent from one another. Therefore, we passed the assumptions of this test. All of this suggests that wRC+ is definitely a good predictor for WAR in a player. As we can see from the power tests above, the power of out test skyrockets with the overall increase in effect size. Even a moderate effect size of around 0.5 would have approximately a 0.92 power: a 92% chance of correctly detecting a false null hypothesis. In non-statistical terms, our p value is so low that it means players with a wRC+ over 100 almost always have a higher contribution to winning (WAR) than players who have a wRC+ lower than 100. These results suggest that wRC+ is a useful indicator of a player's overall performance and value to their team.

In my data, there are a couple of notable limitations. First, WAR for position players is a statistic that includes all facets of the game, not just hitting. WAR factors in hitting, fielding, and base running, all key and specialized skills that contribute to winning. wRC+ on the other hand is a statistic that only accounts for the number of runs that a player produces offensively, completely disregarding for speed or defense. If wRC+ included these other factors, the p value would most likely be 0. Because of the extremely small p value, this magnifies the importance of hitting to MLB teams over every other facet of the game, as these other aspects of the game do not nearly have the impact on winning like hitting does. Another limitation to my data is that I am only checking to see whether one statistic is an accurate predictor of winning. This one test does not help me assess the relative strength of wRC+ compared ot other offensive metrics such as OPS, TB, or wOBA (Pate, 2022; What Is the Best Offensive Stat? - CBSSports.com, 2012). Beyond my data, I think that my test will be slightly weaker than the strength of the p value I received in this test. Because I wanted to ensure that the players I sampled had a minimum of 400 at bats, meaning that they almost never came out of the lineup because these are essentially the players that a manager would stick with in critical games. However, in making this benchmark, I am not including any players that had any deviation from playing an entire season. I think that by including some of these players would weaken my overall p value because most of the players included in the 100-399 at bat range are usually bench players that often pinch hit in dire situations. These players are almost always players that are considered "worse than league average" in many of these statistical categories. I also do not include a lot of injured players, which often regress when inserted back into play.

References

1. Pate, A. (2020). 3 Basic MLB Hitting Stats that Define a Great Hitter. The Hitting Vault. https://thehittingvault.com/3-basic-mlb-hitting-stats-that-define-a-great-hitter/
2. Pate, A. (2022). Advanced Baseball Analytics to Measure a Great Hitter. The Hitting Vault. https://thehittingvault.com/advanced-baseball-analytics-to-measure-a-great-hitter/
3. Weighted Runs Created Plus (wRC+) | Glossary | MLB.com. (n.d.). MLB.com. https://www.mlb.com/glossary/advanced-stats/weighted-runs-created-plus
4. What is the best offensive stat? - CBSSports.com. (2012, November 16). CBSSports.com. https://www.cbssports.com/mlb/news/what-is-the-best-offensive-stat/