**Batting Order is Over Rated**

In today’s article, we introduce a technique which will allow us to analyze baseball events, players, and strategies from the view of **expected average runs created.** We’ll be able to assign values like “hitting a double is worth \_\_\_\_ runs, on average” or “replacing Billy Hamilton with Mike Trout in your batting order is worth \_\_\_\_ runs, on average”. Today we actually talk about a topic which is discussed all the time and is oft debated between baseball fans: batting order. My conclusions are actually a bit surprising. We are going to talk about batting order from the perspective of the 2016 Champions, the Chicago Cubs. We will show in today’s article that using the Cubs Opening Day lineup versus a **random** batting order is only worth about 20 total runs on the season. From a Pythagorean perspective, this difference for the 2016 Chicago Cubs is only about 1.5 wins on the season.

**The Method**

In order to talk about batting order optimization, we will simulate hundreds of thousands of games played by the 2016 Cubs. How do we do it? Well, only two things need to be simulated: the outcome of an at bat (BB, Out, Double, etc.) and base-runner advancement. For simplicity we ignore minutiae such as base stealing, fielder’s choices, and sacrifices (which we will return to at a later date!).

First, how can we simulate the outcome of an at-bat? For an example, let’s look at the outcomes of all of Anthony Rizzo’s 2016 stats. Here is the total count of Rizzo’s plate appearances (PA), Walks (BB), singles (1B), doubles (2B), triple (3B), home runs (HR), intentional walks (IBB), and times hit by pitch (HBP).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PA | BB | 1B | 2B | 3B | HR | IBB | HBP |
| 679 | 74 | 91 | 43 | 4 | 32 | 8 | 16 |

Now, we make one observation: BB, IBB, and HBP have the exact same effect for any runners currently on base and always result in a player on first. Therefore, for all intents and purposes, these outcomes are identical and can be combined into a single “BB %” or, probability of walking. Doing this for Rizzo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BB Prob. | 1B Prob. | 2B Prob. | 3B Prob. | HR Prob. |
| 14.4% | 13.4% | 6.3% | 0.6% | 4.7% |

So, to simulate an at-bat for Anthony Rizzo, we ‘roll a dice’ that has a 14.4% probability of him walking, a 13.4% probability of hitting a single, etc. We can do the same thing for every player on the Cubs’ roster so we can determine the outcome of their at-bat.

Now, after an at-bat has a defined outcome, we need to define what happens with the players already on base. For instance, a player on 3rd will score on any hit. Every player on base will score for a home run. What about the other cases that aren’t so clear cut? We use the following decision matrix to determine what happens to the runners. We omit walks and outs because there is no randomness in what happens to the runners on base.

|  |  |  |  |
| --- | --- | --- | --- |
| At-Bat Outcome | Runner on 1st | Runner on 2nd | Runner on 3rd |
| Single | 2nd Base: 60%  3rd Base: 40% | 3rd Base: 40%  Scores: 60% | Scores |
| Double | 3rd Base: 60%  Score: 40% | Scores | Scores |
| Triple | Scores | Scores | Scores |
| HR | Scores | Scores | Scores |

Note: We chose these probabilities to agree with what has been historically true. That is, runners have scored from 1st on a double roughly 40% of the time in the past.

With being able to simulate what happens with an at-bat and what happens with the runners already on base, we can in-fact simulate entire innings and entire games to see how many runs are scored. Most importantly, because everything is simulated in a computer, we are free to change things like batting order, player average, etc. to see how the average number of runs scored changes.

**How well does this simulate real life?**

Well, if this technique works and our analysis is on sound statistical footing, then it should represent real life quite well. To verify our technique, we simulated thousands of games. We used the Cubs Opening Day lineup with a mild change at pitcher. Since there was a huge difference between the offensive output of Jon Lester (.102 BA) and Jake Arrieta (.262 BA), using one or the other doesn’t accurately represent the season. So, for the 9th spot in the lineup, we totaled the pitcher at-bats for the year and used the aggregate average.

With the above described technique we computed the Cubs average runs per game at 5.07. For a 162 game season this comes out to a total of 821 runs. Now, the Cubs *actually* scored 808 runs on the season. Now we need to take into account the fact that the best players routinely take games off and are replaced with players who have substantially lower averages. We can see our technique works very well.

**Batting Order Optimization**

We simulated a few different types of batting orders for the Opening Day roster. In particular we used the following:

1. The Cubs Opening Day order
2. The Cubs Opening Day order backwards
3. The roster in decreasing order of OPS (On Base Plus Slugging)
4. Random batting orders

We computed the average runs scored for the season using these different techniques and we saw the following:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lineup | Opening Day | Backwards Opening Day | Decreasing OPS | Random |
| Average Runs | 5.07 | 5.05 | 5.08 | 4.93 |
| Season Runs | 821.3 | 818.1 | 823.0 | 798.6 |

There are two conclusions we can make from this. First, the difference between ‘the best’ orders and a truly random batting order is not that much. As stated in the intro, plugging the difference between random and decreasing OPS into the Pythagorean Expectation, you only get a difference of about 1.5 games. One and a half games is nice, but truly is not worth sweating over as a manager.

However we shouldn’t actually be talking about the difference between normal orders and a random order. A bench coach doesn’t use random orders; they try to decide between small things like having your best guy hit third or fourth. The extremely small difference between the first three orders suggest a similarly small difference between these two alternatives. Realistically, using any reasonable order is good enough. Realistically, the batting order doesn’t really matter.