STATS 50: Mathematics of Sport **Spring 2019**

Week 3 – Derivation of the wOBA

Lecturer: Maxime Cauchois

Warning: *these notes may contain factual errors*

1 The wOBA in baseball - See the Wikipedia page

In baseball, wOBA or weighted on-base average, is a statistic, based on linear weights, designed to measure a player's overall offensive contributions per plate appearance. It is formed from taking the observed run values of various offensive events, dividing by a player's plate appearances, and scaling the result to be on the same scale as on-base percentage. Unlike statistics like OPS, wOBA attempts to assign the proper value for each type of hitting event. It was created by Tom Tango and his coauthors for The Book: Playing the Percentages in Baseball.

The formula for the 2018 season was, according to the sabermetrics website FanGraphs:

$$\frac{0.69 \times NIBB + 0.72 \times HBPP + 0.88 \times 1B + 1.247 \times 2B + 1.578 \times 3B + 2.031 \times HR}{AB + BB - IBB + SF + HBP}$$

where:

- AB = At Bat
- BB = Base on Balls
- IBB = Intentional base on balls
- NIBB = Non-intentional base on balls (also called walks)
- SF = Sacrifice Flies
- HBP = Hit by Pitch
- 1B = Single
- 2B = Double
- 3B = Triple
- HR = Home-Run
- PA = Plate Appearance

In this lecture, we will explain how to compute the weights in front of every offensive statistics.

2 The computation of the wOBA

The wOBA creation comes from the simple observation that not all hits are equal, meaning that the batting average is somehow a poor estimator of the real value of a player. We therefore want to weight every hit according their "real" value, but it is not obvious what should be this value. In particular, is a double really worth twice a single? Not really actually.

Weighted On-Base Average combines all the different aspects of hitting into one metric, weighting each of them in proportion to their actual run value. While batting average, on-base percentage, and slugging percentage fall short in accuracy and scope, wOBA measures and captures offensive value more accurately and comprehensively.

We have several requirements for our metrics, based on the fact it has to relate with the OBP, and be on a same "scale":

1. In OBP, an out is worth zero, so we will have to adjust the wOBA to reflect this specificity
2. We want it to average to the OBP, because we want to be able to compare each player's wOBA to the same average as the OBP, ie:

$$\mathbb{E}_{\text{league}}(wOBA) = \mathbb{E}_{\text{league}}(OBP)$$

2.1 Computing Linear Weights

What we want to determine is the average run value of a walk, HBP, single, double, triple, and home run. At the beginning of any play, the bases are occupied in one of 8 different possible states, encoding whether or not a player is present in the first base, second base and so on. In the same way, after any batting event, the bases have possibly moved to a different state. One also has to take into account the current number of Outs when considering the state, which increases to exactly 25 the number of possible different states (counting the final state with 3 Outs).

Suppose that we can award a certain value to each possible state, and for each state s, we call this value function $Q(s)$. The fair value V_i that one would grant to a given play p is the following:

$$V_i = Q(s_p^{fin}) - Q(s_p^{ini}) + R_p \tag{1}$$

where s_p^{ini} and s_p^{fin} are the respective initial and final states of the play and R_p, which can be understood as a "reward", is nothing else than the numbers of runs on play.

The question remains how to attribute a run value to each possible state. The best way to perform this task is to count the average numbers of runs scored after the state appeared (in the same half inning). In other terms:

$$Q(s) = \frac{1}{\sum_p \mathbf{1}_{s_p^{ini}=s}} \sum_{s_p^{ini}=s} TR_p$$

where TR_p stands for the total number of runs scored not only during the play but also after it in the same half inning. In short $Q(s)$ is the expected number of runs scored when in a given state: for instance, a state 111 (all bases occupied) will have more value than a state 000, because one expects to score more runs when all bases are occupied than when none is.

Now that each play has been awarded a given value (thanks to (1)), we can divide them according to the event that occurred during each play. By taking the average value of the runs for each of the different events, we get the runs above average produced by each of these kinds of events, also known as linear weights. We will call those first weights w_e^{ave}, where e represents the event in question (NIBB, HBP, 1B, 2B, 3B or HR). They are said to be above average because the average of those weights w_e^{ave} is exactly the average number of runs scored per play.

2.2 Scaling

For wOBA, we have the runs above average for walks (0.29), HBP (0.31), singles (0.44), doubles (0.74), triples (1.01), and home runs (1.39), but what we want to do now is put wOBA on a scale that will look like OBP in order to make it easier to understand. In OBP, an out is worth zero, so the first thing we want to do is adjust the run value scale so that an out is equal to zero.

There is an easy way to do this. First, we need to find the linear weight for all outs using the same method we used to find the value for the other events: we get a similar weight w_{out}^{ave}, which we will suppose is equal to -0.26 for 2015. This means that an out is worth -0.26 runs less than the average PA when it comes to run expectancy. What we want to do now is add 0.26 to each of our run values so that outs are equal to zero. So for walks, which we said are worth 0.29 runs above average, we bump those up to 0.55 runs relative to an out. Using linear weights, walks are worth 0.55 runs more than outs. We repeat this for each of the other positive offensive outcomes to get our weights w_e^{out}:

$$w_e^{out} = w_e^{ave} - w_{out}^{ave}$$

Now, compute the wOBA for each player in the league using the current linear weights w_e^{out}:

$$wOBA_j = \sum_i w_e^{out} C_{ej} / PA_j$$

where each C_{ej} is the count of events e performed by player j, and PA_j is his number of appearances.

We can then compute a first average wOBA over the league: say it is 0.25 and the OBP average is 0.37. Since we want them to be equal, we are simply going to multiply our current weights to get our final weights w_i^f:

$$w_e^f = w_e^{out} \frac{\mathbb{E}_{\text{league}}(OBP)}{\mathbb{E}_{\text{league}}(wOBA)}$$

This will ensure that both wOBA and OBP have indeed the same average over the league, because our wOBA formula is linear in the weights we have just computed.

References

[1] Tango T., Dolphin A. and Lichtman M. (2006). Drafting Errors and Decision Making Theory in the NBA Draft

[2] FanGraphs https://library.fangraphs.com