# Does the Universal DH Have A Significant Impact on Offensive Productivity?

Carson Slater and Tyler Weiner

4/27/2022


## Abstract

Given the new collective bargaining agreement between the players and the owners, the MLB has added a designated hitter to the National League for the forseeable future. The goal of this agreement was to increase hitting and make the game more interesting for the fan. With the added hitter, the league is hoping for more offensive production during the games. In this report we look at the advanced hitting metrics from the 2020 season, when there was an extra hitter, and compare it to 2021, when there was not an extra hitter. According to the ANOVA tables, each hitting statistic was not changed by the extra hitter. The p-values for each significant statistic was less than .05, which means we have to reject the claim that the hitting improved. The MLB will need to find a new way to increase offensive productivity.


## Introduction

##Background

This past offseason (Winter 2021-22), Major League Baseball underwent a player strike over the new Collective Bargaining Agreement between the League and the Player's Union. After finally coming to an agreement, a number of new implementations for baseball have been put into place, including the instatement of a universal designated hitter (DH). Traditionally in the National League, the pitcher has also been required to hit in the starting lineup, whereas the American league has allowed for an additional position, the DH, who can hit in the place of the pitcher, or another other one position player. Because pitchers are notoriously bad hitters, they traditionally have been the hitter in the lineup for whom the DH hits for. We want to analyze the effect of the Universal DH for offensive statistics by looking at team offensive statistics. By performing analysis of variance, and performing confidence intervals for the difference of population means, we can see if overall team offensive productivity has decreased between the years 2020, where a universal DH was implemented for a shortened season, and 2021, where the National League took away the universal DH for one last season.

### Methods

We will be using Major League batting and advanced batting data from baseball reference. This data contains the average statistics for each team. Supposing there is not a statistically significant difference in the distributions, we will use Analysis of Variance (ANOVA) for each of the following batting statistics:

| Variable | Statistic | Variable | Statistic |
|---|---|---|---|
| `R/G` | average runs per game | `HR%` | home run % |

| Variable | Statistic | Variable | Statistic |
|---|---|---|---|
| `BA` | batting avg. | `SO%` | strike out % |
| `OBP` | on base % | `HardH%` | hard hit % (greater than 95 mph) |
| `SLG` | slugging % | `GB/FB` | ground ball to fly ball ratio |
| `OPS` | on base plus slugging % | `RS%` | runs scored % |
| `OPS+` | on base plus slugging (ballpark adjusted) | `SB%` | stolen base % |
| `BAbip` | batting avg. for balls in play | `XBT%` | extra bases taken % |

Caption: Different statistics to measure offensive productivity.

Because the 2020 MLB season was only 60 games per team, and the 2021 season was a full-length season with 162 games per team, we are unable to perform ANOVA on certain cumulative statistics, such as total home runs per team, total hits, or total stolen bases. We navigate this by comparing other non-cumulative statistics based on averages or scales, like `BAbip` and `OPS+`. To further investigate and confirm our ANOVA results, we will be forming confidence intervals for the difference of means for each statistics.

## Obstacles and Weaknesses of This Analysis

While we can assume that each year is an independent random variable, there are some assumptions that have to be made. Even though there is an traditionally better hitter added to each lineup for each game, changes continue to be made to the game. For instance, batting average have gone down throughout the years and homeruns have gone up every year. Also, the game has continued to evolve over time. Today's game relies more on the homerun and launch angle, while the past has relied more on baserunning and speed. To try and eliminate this issue, we took the data set from 2020 when there was a DH, and 2021 when there was not a DH, and compared these two statistics.

## Exploratory Data Analysis

To perform ANOVA, three assumptions must be made, and it is helpful to provide evidence that the assumptions being made are in fact viable. The three assumpltions are:

- **The samples drawn are independent from *k* populations.** This is the only assumption we will not be checking. In baseball, each season is always assumed to be independent of one another.
- **Each of the populations are normally distributed with** $\mu_i$.
- **The *k* populations have a common variance** $\sigma^2$.

### Checking Normality of Statistics

There are a few ways to assert whether or not a variable is normally distributed. Some are more robust than others. Traditionally, given our statistical knowledge, we have observed the shapes of histograms to check for a Gaussian shape to any given data set. For this particular data set, we will be performing a Shapiro-Wilks normality test on all of the variables for both years. A Shapiro-Wilks test is based on the null hypothesis that the population of the variable under consideration is normally distributed. The test statistic can be calculated as follows:

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

where $x_{(i)}$ (not to be confused with $x_i$) is the $i^{\text{th}}$ order statistic, $\bar{x}$ is the sample mean, and the coefficient$(a_1, ..., a_n)$ are given by

$$(a_1, ..., a_n) = \frac{m^\intercal V^{-1}}{\|V^{-1}m\|}.$$

In this particular formula for the coefficients, $m$ is a vector made of expected values for order statistics, $V$ is covariance matrix of those order statistics. Monte Carlo simluations are used to calculate the test statistic, because it has no known distribution. The Shapiro-Wilks test returns a p-value with the statistical significance of the null hypothesis. Here we see the p-values:

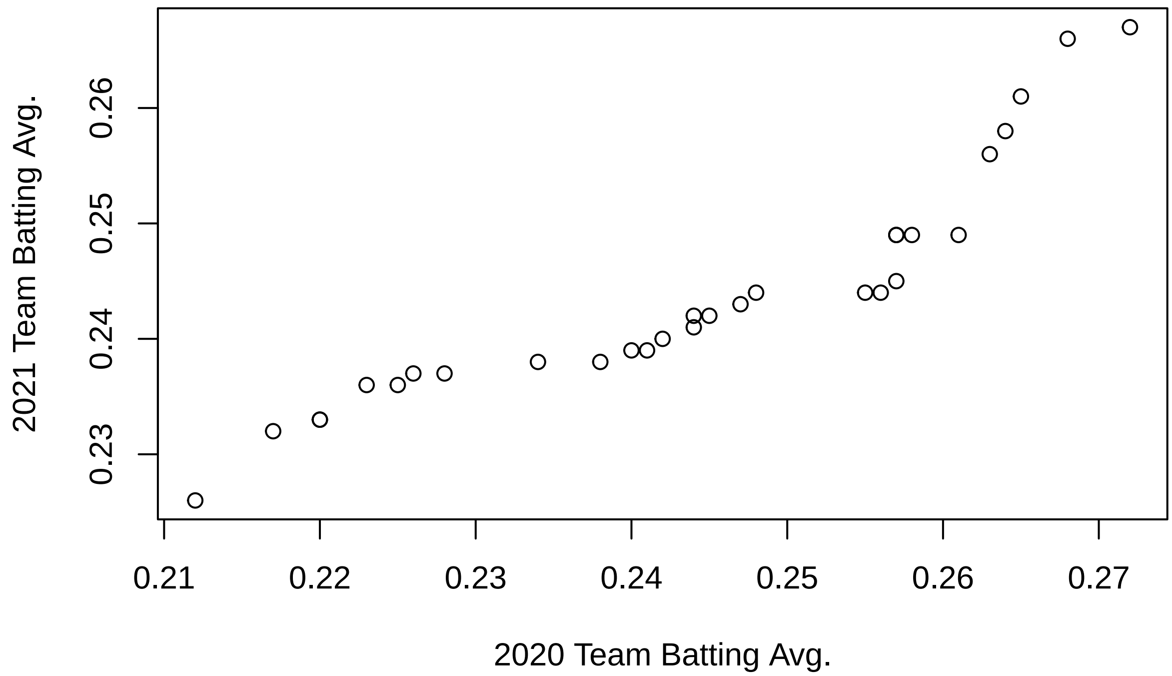
```
shapiro.table %>% kable(col.names = c("Statistic", "2020 P-Value", "2021 P-Value"), caption = "Table of
```

Table 2: Table of P-Values for the Shaprio-Wilks tests on all of the yearly offensive productivity statistics.

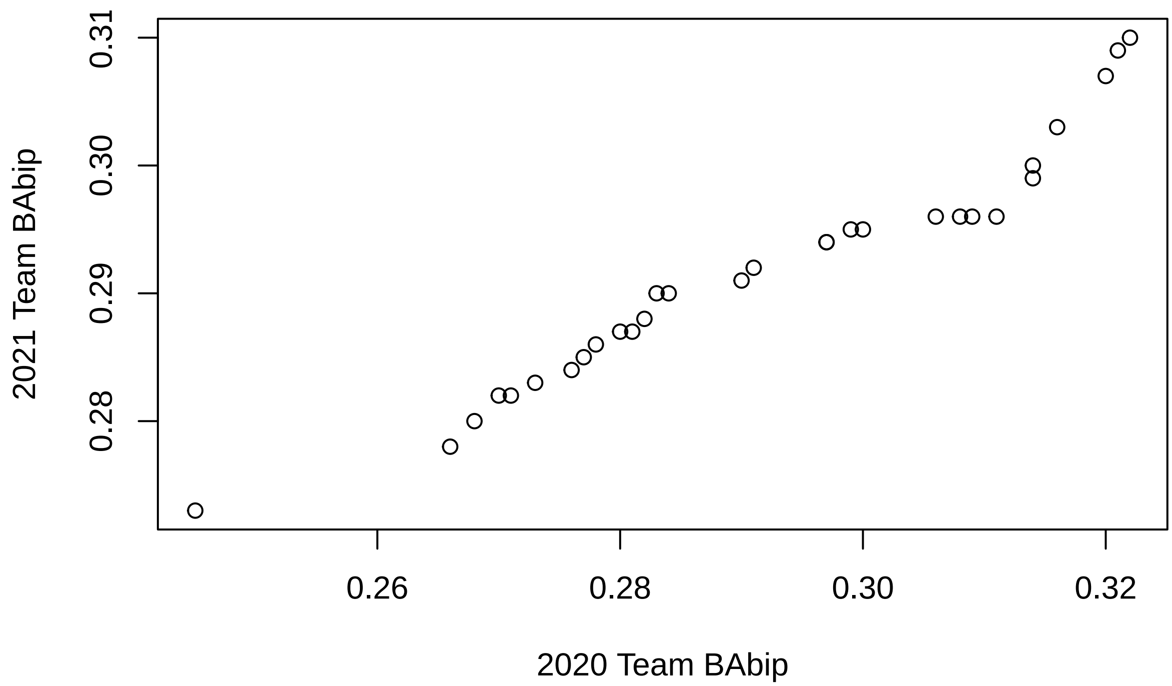
| Statistic | 2020 P-Value | 2021 P-Value |
|---|---:|---:|
| Runs per Game | 0.6898 | 0.3175 |
| Batting Avg. | 0.2171 | 0.0572 |
| On Base % | 0.3490 | 0.8154 |
| Slugging % | 0.4271 | 0.9758 |
| On Base plus Slugging | 0.5269 | 0.7030 |
| On Base plus Slugging + | 0.1692 | 0.4393 |
| Batting Avg. for Balls in Play | 0.2659 | 0.8555 |
| Home Run % | 0.1773 | 0.9268 |
| Strike Out % | 0.6300 | 0.9424 |
| Hard Hit Ball % | 0.7276 | 0.8788 |
| Ground Ball to Fly Ball Ratio | 0.5450 | 0.1689 |
| Runs Scored % | 0.4309 | 0.1021 |
| Stolen Base % | 0.4739 | 0.1638 |
| Extra Bases Taken % | 0.4142 | 0.4455 |

We observe that all of the p-values are all above the standard $\alpha = 0.05$ significance level. Therefore, we fail to reject the null hypotheses that any of these variables does come from a normally distributed population. An alternative approach to testing normality is a Quantile-Quantile plot (Q-Q plot). It compares two distributions by plotting their quantiles against each other. If two of the distributions are similar, the plot will lie on the line $y = x$. This is traditionally a non-parametric approach to comparing underlying distributions. Below are some of the Q-Q plots from these data:

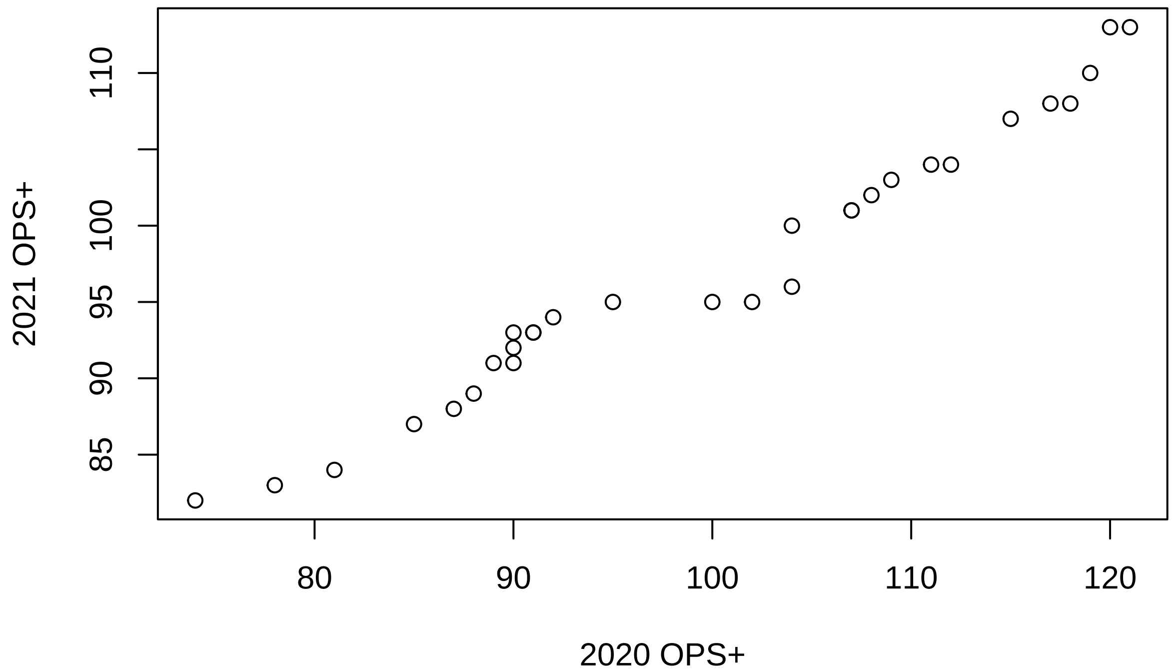
```
qqplot(twenty.team.batting$BA, twentyone.team.batting$BA, xlab = "2020 Team Batting Avg.", ylab = "2021
```



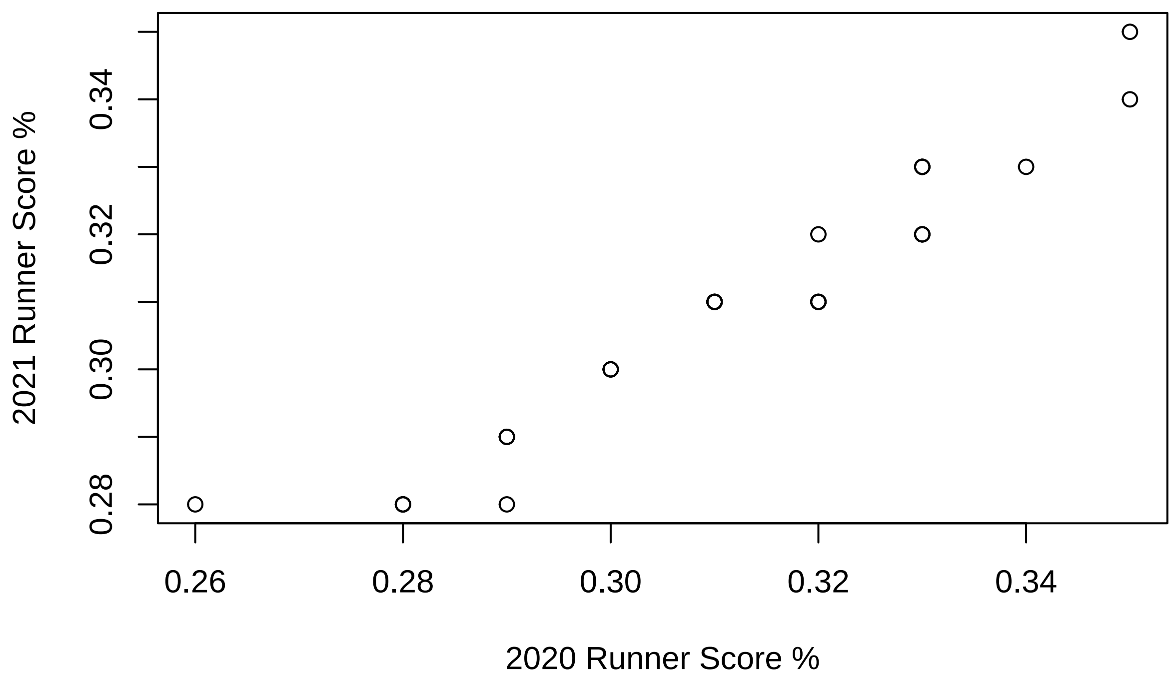

```
qqplot(twenty.team.batting$BAbip, twentyone.team.batting$BAbip, xlab = "2020 Team BAbip", ylab = "2021 '
```




```
qqplot(twenty.team.batting$`OPS+`, twentyone.team.batting$`OPS+`, xlab = "2020 OPS+", ylab = "2021 OPS+
```




```
qqplot(twenty.team.batting$`RS%`, twentyone.team.batting$`RS%`, xlab = "2020 Runner Score %", ylab = "20
```




The Q-Q plots that we observe, not all of which were included, generally followed a linear pattern in their plots, further validating our assumption for ANOVA that all of the samples are drawn from similar populations.

#### Checking Common Variance

One of the assumptions made in ANOVA is the fact that each population that is being compared possesses the same variance. Although the true population variance is unknown, the sample variance can be used as an estimator for the population variance. To check if there are similar population variances, we will use the rule of thumb that if the ratio of sample variances is greater than 3, then our assumption of equal population variance is violated. Hence, if

$$\frac{\sigma^2_{2020}}{\sigma^2_{2021}} \leq 3,$$

then our assumption is not violated. Taking a look at sample variances, we observed the sample variances and thier ratios.

```
var.check %>% kable(col.names = c("Statistic", "2020 Variance", "2021 Variance", "Ratio (2020/2021)"))
```

| Statistic | 2020 Variance | 2021 Variance | Ratio (2020/2021) |
|---|---|---|---|
| Runs per Game | 0.286357 | 0.188361 | 1.5203 |
| Batting Avg. | 0.000286 | 0.000098 | 2.9277 |
| On Base % | 0.000251 | 0.000135 | 1.8604 |
| Slugging % | 0.001259 | 0.000557 | 2.2610 |
| On Base plus Slugging | 0.002437 | 0.001109 | 2.1973 |
| On Base plus Slugging + | 183.247126 | 76.695402 | 2.3893 |
| Batting Avg. for Balls in Play | 0.000389 | 0.000081 | 4.7685 |
| Home Run % | 0.000042 | 0.000024 | 1.8006 |
| Strike Out % | 0.000425 | 0.000259 | 1.6390 |
| Hard Hit Ball % | 0.000853 | 0.000494 | 1.7276 |
| Ground Ball to Fly Ball Ratio | 0.006040 | 0.004550 | 1.3274 |
| Runs Scored % | 0.000485 | 0.000359 | 1.3507 |
| Stolen Base % | 0.006589 | 0.003515 | 1.8744 |
| Extra Bases Taken % | 0.002902 | 0.001136 | 2.5548 |

From checking the ratios of variance, we can see that `BAbip` is substantially greater than three, hence violating the assumption for equal population variance. Because each of the two populations possesses the same amount of observations ($n_i = 30, i = 1, 2$), we will proceed with ANOVA and lean on the robustness of ANOVA to its assumptions.

# Statistical Analysis

## ANOVA

Now that all of the p-values for the ANOVA tests have been calculated, we evaluate them to make judgement calls about the equality of the means for each distribution.

```
aov.table %>% kable(col.names = c("Statistic", "P-Value"), caption = "Analysis of Variance Between 2020
```

Table 4: Analysis of Variance Between 2020 and 2021 MLB Offensive Productivity Statistics

| Statistic | P-Value |
|---|---|
| Runs per Game | 0.0016 |
| Batting Avg. | 0.0131 |
| On Base % | 0.0078 |
| Slugging % | 0.0010 |

| Statistic | P-Value |
|---|---|
| On Base plus Slugging | 0.0005 |
| On Base plus Slugging + | 0.0002 |
| Batting Avg. for Balls in Play | 0.0127 |
| Home Run % | 0.0012 |
| Strike Out % | 0.0036 |
| Hard Hit Ball % | 0.0028 |
| Ground Ball to Fly Ball Ratio | 0.0005 |
| Runs Scored % | 0.1269 |
| Stolen Base % | 0.5532 |
| Extra Bases Taken % | 0.0186 |

Testing at an $\alpha = 0.05$ significance level, it is interesting because the only offensive productivity statistics where we can reject the null hypotheses is strictly the baserunning statistics, `RS%`, `SB%`, and `XBT%`.

## T-tests

To further investigate this, we conducted paired t-tests for the difference of means of each offensive productivity statistic and formed confidence intervals for each them.

```
ttest.table %>% kable(col.names = c("Statistic", "Observed T", "P-Value", "CI Lower Bound", "CI Upper Bo
```

| Statistic | Observed T | P-Value | CI Lower Bound | CI Upper Bound | Estimate |
|---|---|---|---|---|---|
| Runs per Game | 1.3328 | 0.1930 | -0.0608 | 0.2881 | 0.1137 |
| Batting Avg. | 0.1671 | 0.8685 | -0.0052 | 0.0062 | 0.0005 |
| On Base % | 1.7861 | 0.0845 | -0.0007 | 0.0102 | 0.0047 |
| Slugging % | 1.1887 | 0.2442 | -0.0046 | 0.0173 | 0.0064 |
| On Base plus Slugging | 1.5087 | 0.1422 | -0.0039 | 0.0258 | 0.0110 |
| On Base plus Slugging + | 1.3860 | 0.1763 | -1.2684 | 6.6018 | 2.6667 |
| Batting Avg. for Balls in Play | 0.0104 | 0.9918 | -0.0065 | 0.0066 | 0.0000 |
| Home Run % | 1.8227 | 0.0787 | -0.0002 | 0.0039 | 0.0018 |
| Strike Out % | 0.8386 | 0.4085 | -0.0041 | 0.0097 | 0.0028 |
| Hard Hit Ball % | -2.0619 | 0.0483 | -0.0193 | -0.0001 | -0.0097 |
| Ground Ball to Fly Ball Ratio | 0.0000 | 1.0000 | -0.0246 | 0.0246 | 0.0000 |
| Runs Scored % | 0.5931 | 0.5577 | -0.0065 | 0.0119 | 0.0027 |
| Stolen Base % | -0.6536 | 0.5185 | -0.0468 | 0.0241 | -0.0113 |
| Extra Bases Taken % | 2.4526 | 0.0204 | 0.0037 | 0.0410 | 0.0223 |

From observing the confidence intervals of the three baserunning statistics, the only decisive difference of means was in the `XBT%` variable, or "Extra Bases Taken %." This metric measures the percentage of times a baserunner successfully advances an extra base when he traditionally only advances one base (i.e. A runner on first advances to third on a base hit) The only intuitive explanation would be an increase in the percentage of balls in play, and the only way a DH could be a cause for that is if an additional good hitter in the lineup. For the year 2021, both the National League and American League `XBT%` average is 40%. It does not appear that the cause of the slight increase for `XBT%` is the presence of the designated hitter in both leagues.

Further examination of the confidence intervals shows that we have 95% confidence that the difference in means for Hard Hit Ball % (`HardH%`) in 2020 and 2021 respectively is between -0.0193 and -0.0001. In this particular analysis, this stands out because it is the only definitive difference in means within this analysis that can be intuitively explained by the implementation of a universal DH. It is intuitively expected that a season with a universal DH would have a higher Hard Hit Ball % in than a season without a universal DH. Pitchers, who are worse hitters, do not have as many "barreled," or hard hit balls as a DH in their place would. However, we must consider that this was the t-test with the lowest p-value, and does not meet our $\alpha = 0.05$ significance level. It would be a stretch to draw any conclusions from these data.

## Conclusion

According to these data, we learned that the change in the rule of adding an extra hitter to the lineup instead of allowing the pitcher to bat did not have neither statistically nor practically significant impact on the hitting statistics between each year. This goes against our intuition, but our statistical anaylsis has proven otherwise. On average a pitcher will only get two at bats per game, before they are pulled out of the game and a pinch hitter will take their place. Getting a hit is already a low probability event; the low percentage of hits a pitcher will get does not affect the total averages as much as we originally assumed. When pitchers hit, they are also tasked with bunting the ball. If the pitcher successfully bunts the ball, the plate appearance does not result as an at bat therefore not hurting the team statistics. This will also lower the total at bats for pitchers resulting in no change. So while the commissioner and owners believe there is a change with adding a universal DH, the statistics behind the change does not change. All the rule change has done has made the traditional fans upset about the change in the rules.