The Effect of the Designated Hitter in Major League Baseball

Nick Norton
Math Department
Seattle University
Senior Synthesis

Last Updated: June 8, 2017

1 Abstract

In Major League Baseball (MLB), the distinct rule change between its two leagues is the use of the designated hitter (DH) in the American League (AL) and the absence of one in the National League (NL). We explore the effect this has on the total runs scored in games in each league as well as match ups between NL teams and AL teams. Through sabermetrics and statistics, we can better understand the impact of the designated hitter in MLB. Today, this is the only rule change between the MLB's two leagues which offers an established control and experimental group to analyze the effect of the DH in MLB.

2 Introduction

2.1 History

In 1903, the AL and the NL merged to become MLB. The popularity of baseball in America during this time was unmatched by any other sport. It was not until the 1960s, that baseball had begun to see its title of America's most popular sport threatened. Pitchers had begun to dominate hitters causing a decline in offensive run production.[4] And the rise in popularity of the National Basketball League and the National Football League also diminished MLB's attendance. With the AL suffering lower attendance than the NL, they were desperate to bring more attention to the sport. One way to do this was inflate the scoring to create more excitement to the game. In 1973, the AL chose to adopt the DH rule with it stating:

> "A hitter may be designated to bat for the starting pitcher and all subsequent pitchers in any game without otherwise affecting the status of the pitcher(s) in the game. A Designated Hitter for the pitcher must be selected prior to the game and must be included in the lineup cards presented to the Umpire in Chief."[1]

The rules of the home team determine how the game is played. This rule change affects how teams from the NL and AL play each other in the World Series and in interleague play. Interleague games are regular games during the season when teams from both leagues play each other and were first introduced in 1997 to MLB. We will investigate how much this affects the game of baseball.

However, the adoption of the DH rule was met with controversy and has continued to be a topic of conversation throughout the league.

2.2 Proponents of the DH

- The DH ideally increases offensive run production.
- Allows for older players who can still hit but have lost their ability to play in the field to prolong their careers. (However, this could be seen a disadvantage to younger players who do not end up making teams because a skilled veteran player takes his position).
- A pitcher in the AL that is forced to bat and run the bases is at risk to injury since they do not normally practice that aspect of the game.[2]

2.3 Opponents of the DH

- Traditionalists have argued that the pitcher like all other field positions should be responsible for their own offensive output.
- Some strategy is lost in the AL when a power hitter moves into the batting order. Pitchers at bat will often make sacrifice bunts and flies to allow a player on base to advance further down the baseline. Another situation that managers in the NL have to face is the decision late in the game to remove the pitcher from the batting order for a pinch hitter. For example, at the top of the 9^{th} inning and the batting team is losing, they would pull the pitcher to put in a better hitter off the bench because if they do not score that inning, then the game is over. But, once a player is taken out they cannot go back in for the rest of the game. This scenario also happens earlier in games, after the pitcher has thrown a lot of pitches and can not finish the rest of the game.[4]

2.4 Batting Averages (BA)

The BA is the number of hits divided by the number of at bats. This stat is an effective way to observe a players hitting ability. Let's consider the overall BA for the pitcher, non-pitchers, and DH from the 2016 MLB season. NL pitchers had a BA of 0.138, while AL pitchers only had a BA of 0.106. This makes sense because a pitcher that can hit is more valuable to an NL team then to an AL team and thus an NL team is more likely to pay more to sign them. Since there are much more at bats for NL pitchers than AL pitchers in a season, the BA of all pitchers last season was closer to the BA of NL pitchers at 0.135. For all players (except the pitcher), the BA was 0.254. While the DH had the highest BA for any position with 0.260. Comparatively, the BA for the average DH is 0.125 better than for the average pitcher.[3]

2.5 Why are pitchers bad hitters?

There are 25 players on the active roster in MLB. With 30 teams in the MLB, that means there are 750 players to start the season on the active roster. About half of the roster is composed of pitchers so there are about 375 pitchers in MLB to start the season and 375 players in other positions to start the season. Since there are about 455,300 high school baseball players in America and about 25,700 college baseball players, there is a lot of competition to reach MLB.[7] In an intensely competitive and strategic sport, only the 375 best pitchers and the 375 best hitters and fielders make it into MLB. The odds of finding a player that can do it all (hit, field, and pitch at the professional level) are even more reduced. Plus, there is a DH rule in the NCAA which allows the pitcher to be replaced by a DH. So, pitchers in college do not necessarily have to practice their hitting before they reach MLB. These are just a few reasons why pitchers are notoriously known for being atrocious hitters. Implementing the DH allows the pitcher, generally an ineffective hitter, to be replaced in the batting order by a specialty hitter. However, this is not a mandatory rule, which means that a team can use their pitcher instead. This is a very rare occurrence though.

3 Background

3.1 Sabermetrics

Players were once overlooked for a variety of biased reasons and perceived flaws: age, height, weight, and personality. Bill James and mathematics tossed this notion out the window with comprehensive statistical analysis of baseball data which he called, sabermetrics. Decades before, the use of statistics is baseball had been measured, most notably by Henry Chadwick. This analysis paved the way for modern big data analysis. It was not until the 1990s when the Oakland Athletics first truly implemented this big data analysis philosophy into running major league teams. The Athletics began to find tremendous success despite having one of the lowest payrolls in baseball. From then on, sabermetrics use has grown in MLB and has allowed people who run MLB clubs to better understand players and their teams. Players that were once undervalued for superficial reasons, are now looked at through a lense of statistical analysis. Since baseball records all kinds of data and the information is readily available, with the help of the Retrosheet and Lahmen databases, we can now build off that intelligence to analyze all aspects of the game. Statistics may not always produce the expected result or fully explain occurrences, but it does allow us to better understand the game. In this paper, instead of studying individual player analysis, the focus remains on run production and wins between between the NL and AL.[6]

3.2 Box plot

A box plot is a representation of numerical data through their quartiles.

- Q1: median of the lower half of the data set
- Q2: median of the entire data set
- Q3: median of the upper half of the data set
- IQR: interquartile range from Q1 to Q3
- Extreme values (whiskers): outliers in the data set
  - Larger than Q3 by at least 1.5 IQR
  - Smaller than Q1 by at least 1.5 IQR

3.3 Statistics

Standard deviation (σ) is the amount of variation of a set of data values

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \qquad (1)$$

- x_i each value in the data set
- $\bar{x}$ mean value of the data set
- n total number of data values in the data set

Standard error of the mean ($\sigma_{\bar{x}}$) is the standard deviation of the sample mean's estimate of the population mean. The smaller the standard deviation, the more representative the sample will be of the entire population. The formula is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

3.4 Confidence intervals

A confidence interval is a range of values that we think has a given probability of containing our mean. We can use the mean and standard deviation from a sample population to construct an interval estimate of potential values for the population parameter. We want to determine how good the estimate is from a given sample mean. The confidence level of an interval estimate is the probability that the interval estimate will contain the population mean. Since we do not know the true distribution of runs per game we only have samples of the runs per game from the data we have seen. A 95% level of confidence would mean that if we took 100 repeated samples, we would expect 95 of the means to fall between the calculated interval estimate. The upper bound of a 95% confidence interval is $\bar{x} + (2 \times \sigma_{\bar{x}})$ and the lower bound is $\bar{x} - (2 \times \sigma_{\bar{x}})$.The sample mean $(\bar{x})$ is the average total runs per game per season of the population mean which is the entire MLB population. Each data set is a given year with n being the number of games in that data set. The terms in the standard deviation (σ) where x_i is the total runs scored in each game in the data set, $(\bar{x})$ is the average total runs per game in the data set.

3.5 ANOVA: P values

We use a two-way ANOVA test to compare the mean differences between two groups that have been divided based on two independent variables. The objective of a two-way ANOVA is to determine if there is an interaction between the two independent variables on the dependent variable. In our case, the two variables are games that used a DH and games that did not use a DH. This classification divides our MLB population into two independent populations. This will allow us to understand if the DH has an effect in MLB games. In an ANOVA test, we conduct a hypothesis test which gives us a P value to determine whether our results are statistically significant. From the P values, we can determine whether there is enough evidence from a sample of data to support that there is a significant effect of using a DH in a MLB game. This test is preformed to consider that a random sampling error might have occurred in that test and that the data gathered might not be an accurate representation. A two-tailed test is used to consider if the DH has an added effect or a lessened effect. This is done because we want to be unbiased about the results.

To begin the test, we present two hypotheses:

1. The null hypothesis is $(H_0) : \mu_1 - \mu_2 = 0$. This means that the designated hitter does not have an effect on the total runs scored in a game. We want to show that there is enough evidence to show that this is false.

2. The alternative hypothesis is $(H_A) : \mu_1 - \mu_2 \neq 0$. This means that the DH does have an effect on the total runs scored in a game. We to show that there is enough evidence to show that this is true.

The formula below shows us how to get a t value which we can then use to get a P value:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{2}$$

- $\bar{x}_1$: sample mean of 1^{st} data set
- $\bar{\mu}_1$: true mean of 1^{st} population
- σ_1: standard deviation of 1^{st} data set
- n_1: size of 1^{st} data set
- $\bar{x}_2$: sample mean of 2^{nd} data set
- $\bar{\mu}_2$: true mean of 2^{nd} population
- σ_2: standard deviation of 2^{nd} data set
- n_2: size of 2^{nd} data set

Once we have calculated our t value and our degrees of freedom (which is just the number of games in a season $-$ 1), we use a t-distribution (two-tailed) table to determine the P value.

We want to show that there is enough evidence support the alternative hypothesis. A P value is between 0 and 1 and a low P value means you can reject the null hypothesis. A significance level of less than 0.05 indicates strong evidence that we reject the null hypothesis. A significance level of greater than 0.05 indicates weak evidence that we reject the null hypothesis.

We must consider two possible errors. A type 1 error would be that we reject our null hypothesis but in fact it was actually true. This would be if I tell everyone that the DH does affect the average total runs per game even though it does not. A type 2 error would be that we fail to reject our null hypothesis but in fact it was actually untrue. This would be if I tell everyone that the DH does NOT effect the average total runs per game even though it actually does. A significance level of 0.05 indicates a 5 percent risk of concluding that a difference exists when there actually is no difference.

	Reject H_0	Retain H_0
H_0 true	Type I error	OK
H_0 false	OK	Type II error

Therefore, there are 4 possible outcomes that can occur. There could be a type I or type II error. Or we could fail to reject our null hypothesis or reject our null hypotheis.

3.6 Confidence intervals vs P values

P values only focus on whether or not there is evidence to support a significant effect. Confidence intervals tell us about the range of values of an estimated effect this provides more information about how big or small the true effect might possibly be.

3.7 Quadratic Regression

Quadratic regression is the procedure of solving the equation of the parabola that best fits a set of data. As a result, we get an equation of the form:

$$y = ax^2 + bx + c \qquad \text{where } a \neq 0 \tag{3}$$

Using the least squares method:

$$S = \sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]^2 \tag{4}$$

We can find the values of a, b, c such that the squared vertical distance between each point (x_i, y_i) and the quadratic curve $y = ax^2 + bx + c$ is minimal.

4 Main Results

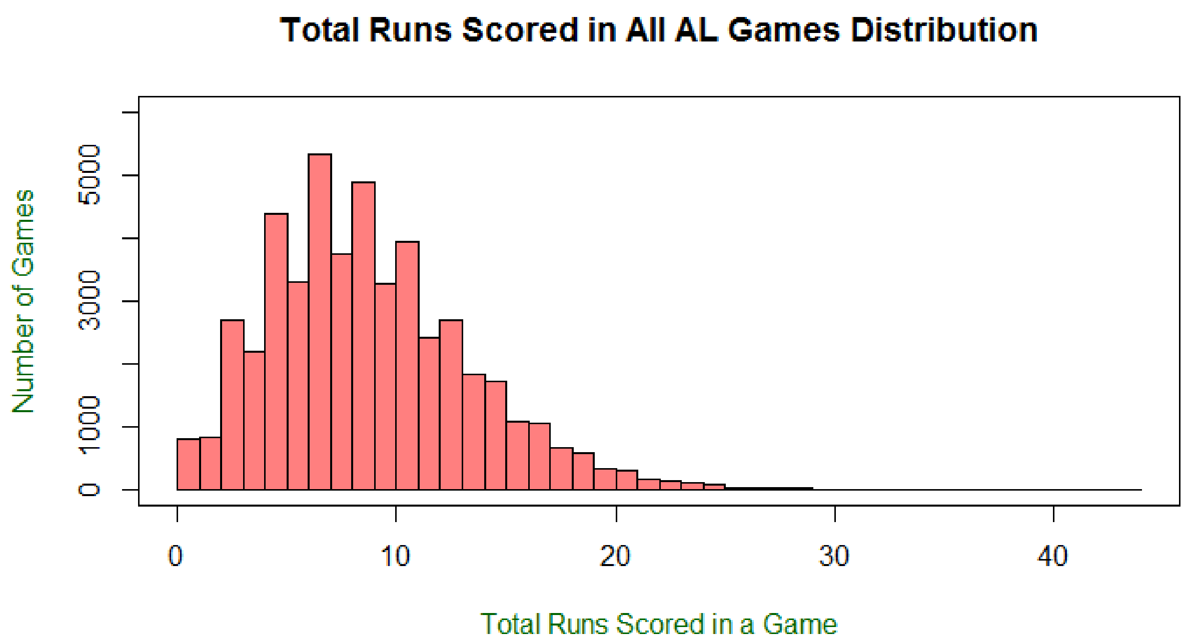
4.1 Runs Per Game

Instead of looking into player analysis we look at what kind of effect does the DH have in the scores of games. The Retrosheet database offers baseball statistics dating back to 1871. But for this paper we only needed the data dating back to 1973. With the help of the programming language R, we can gather box score records for every game since the start of the 1973 season.[5]

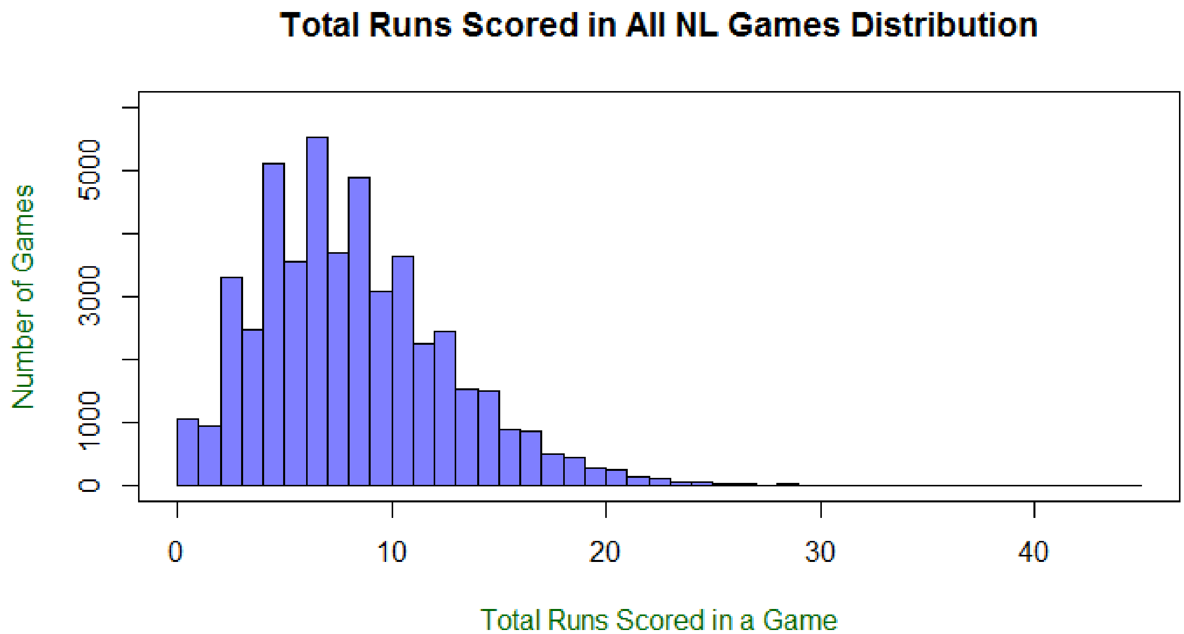
From the data collect, the games were divided into two groups whether a DH was used or not. Examining the total runs per game between these two groups allows us to better understand the effect of how the rule has changed the game. The goal is to show whether we can statistically say whether the designated hitter significantly affects the offensive production in a game.

4.2 Histograms

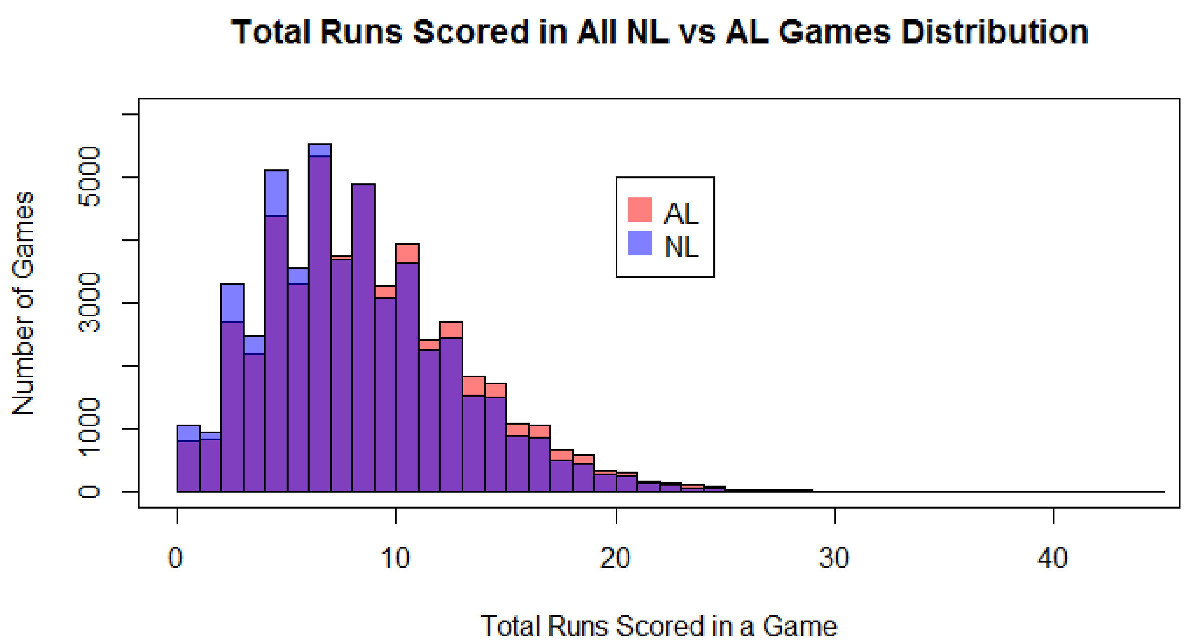
After gathering information about MLB games, below is a histogram of the frequency of the total runs scored in all AL games since 1973. The total runs scored per game is the sum of the runs scored by both teams in each game. The average total runs scored per game for this data is 9.24 runs scored per game.



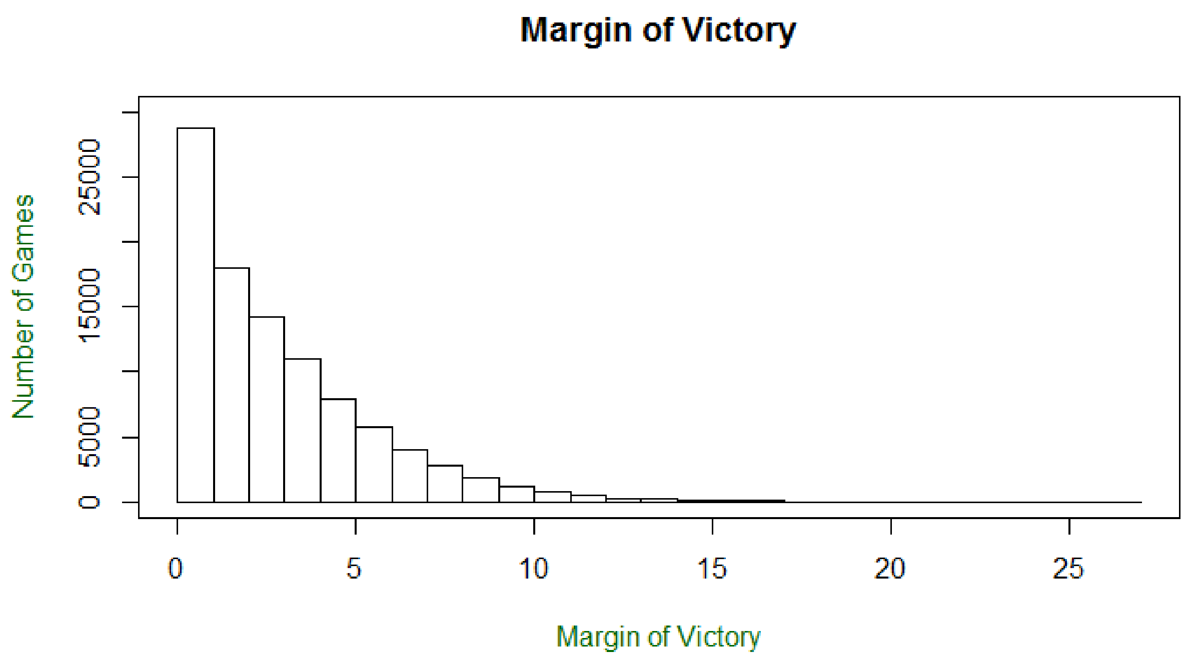
Below is a histogram of the frequency of the total runs scored in all NL games since 1973. The average total runs scored per game for this data is 8.69 total runs scored per game.



Below is comparison of both histogram models from before displayed in the same diagram. As you can see the data for the NL is skewed slightly more to the right than for the AL. Since games cannot end in a tie, even total run scored games are less common than games with odd total run scored games.

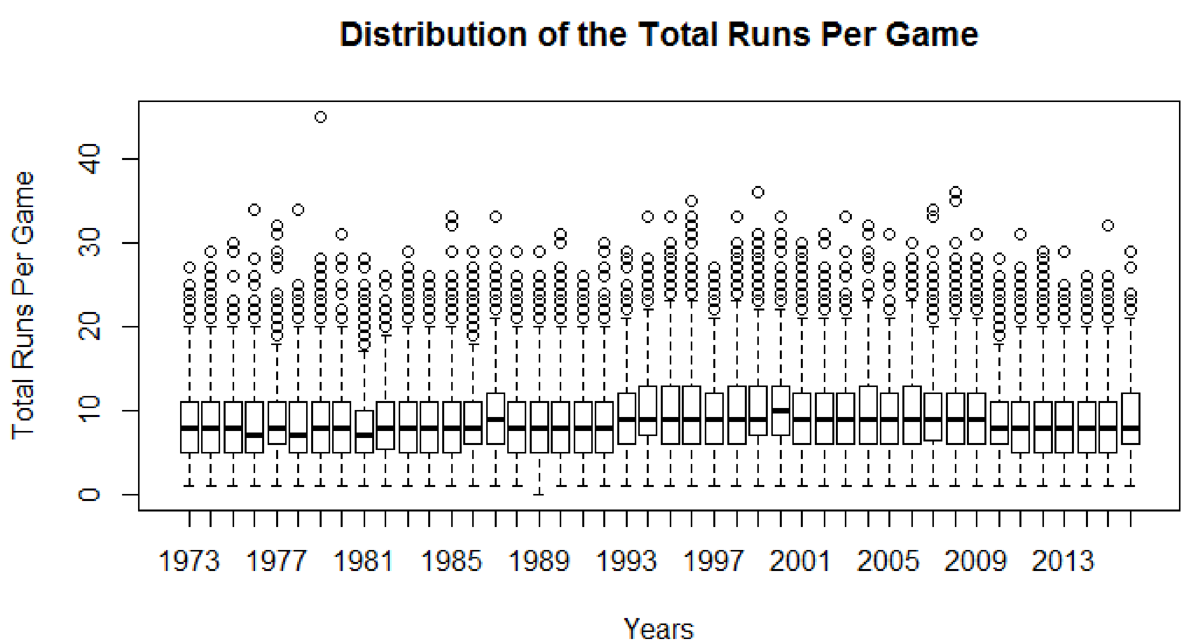


Below is the margin of victory in all games Since 1973. In all 97,294 games since 1973, nearly 30% of games were determined by one run.



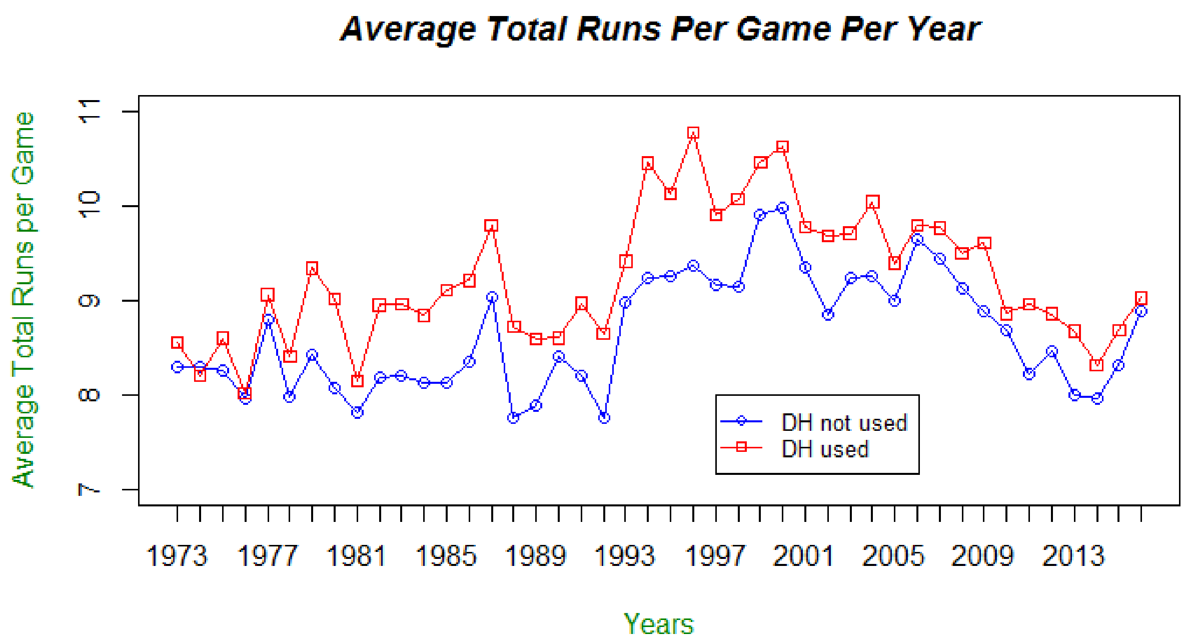
4.2.1 Box Plot

Using the information in 3.2 we can constuct a box plot of all games in every season from 1973 to 2016. This diagram shows how the distributions changed from year to year.



4.2.2 Average total runs scored per game

With the same total runs scored in games data from before, we can also calculated the mean total runs scored per game in all games when a DH was used and all games when a DH was not used. This allows us to see the effect of the DH from a broader approach. Addressing each season as an independent sample is critical ensure that any rule changes will be consistent with the respective season. Below is a graph of the mean total runs scored in each game in each season since 1973.

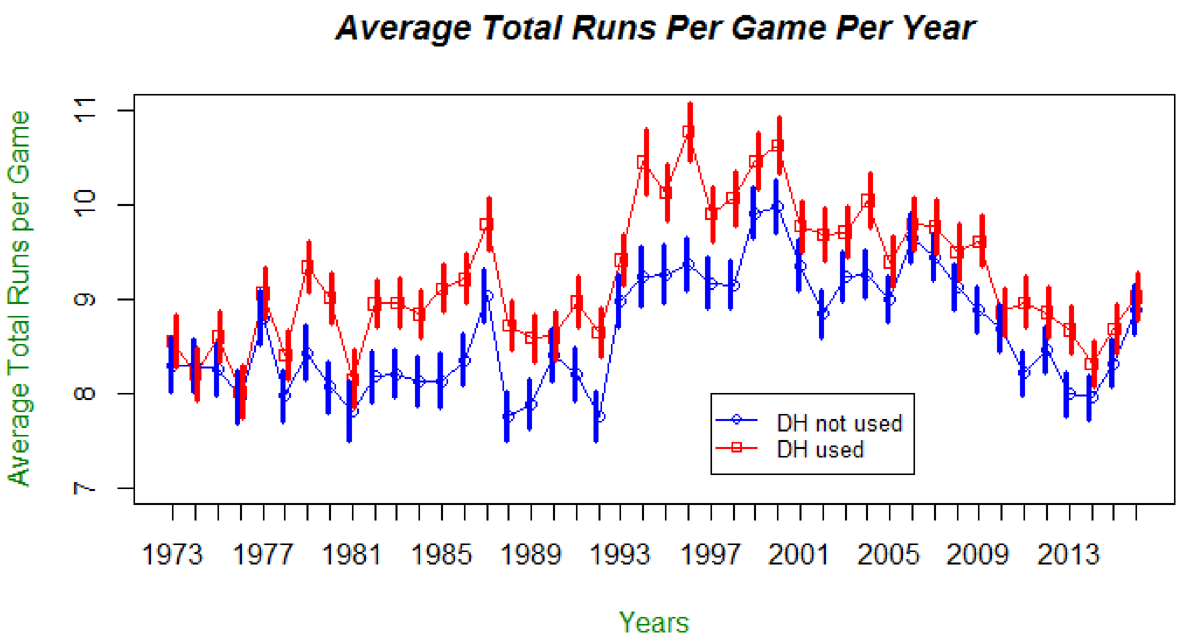


In almost every year the AL holds the advantage over the NL in terms of average total runs per game. Only in 1974 did the NL outscore the AL. One could assume that that there was an adjustment period for the AL to maximize the potential of the DH rule. We can also see a rise in overall offensive performance during the steroid era, from the early 1990s to the late 2000s. Players in both leagues benefited from performance enhancing drugs during this time when steroid use was abundant and a meticulous process for identify performance enhancing drugs was not available. With all the records that were broken during that era and all the players that were revealed to have doped, this rise in offensive production is not surprising.

We do not know the true population mean (average total runs per game per season). But we do have sample means for each season.

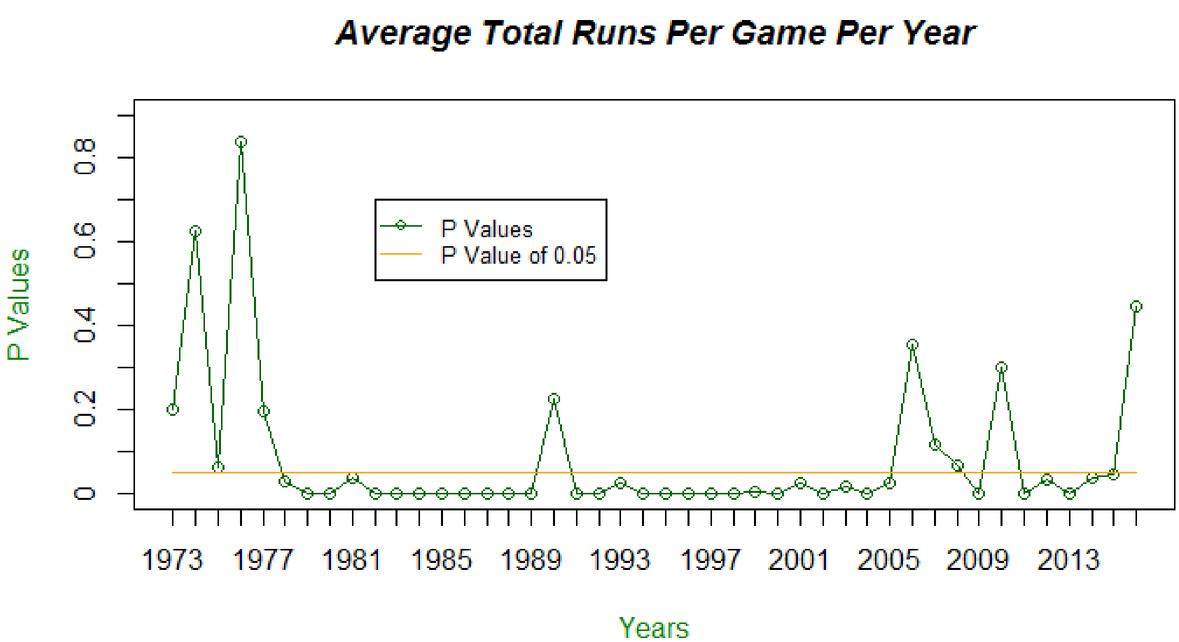
4.2.3 Confidence intervals

Using the equations from sections 3.2 and 3.3 equations we can calculate the upper and lower bounds for this data. On the next page is the graph of the mean total runs scored in each game in each season since 1973 with the confidence intervals added. Of the 44 seasons, 22 seasons had overlapping confidence intervals between the AL and the NL.



4.2.4 P Values

Using the equation and information from section 3.4, we can calculate the P values for every season from 1973 to 2016, to see if there is statistical significance that the DH influence the total runs in all games. Of the 44 seasons, 33 seasons had P values less than 0.05.



4.2.5 Quadratic Regression

Using our total runs per game data, we can determine a quadratic equation model that approximates that data as close as possible. Quadratic regression models the expected value based on all data points in the data set. We can calculate the coefficients and accuracy of the quadratic regression model with the equations in 3.6 and the use of the programming in R.

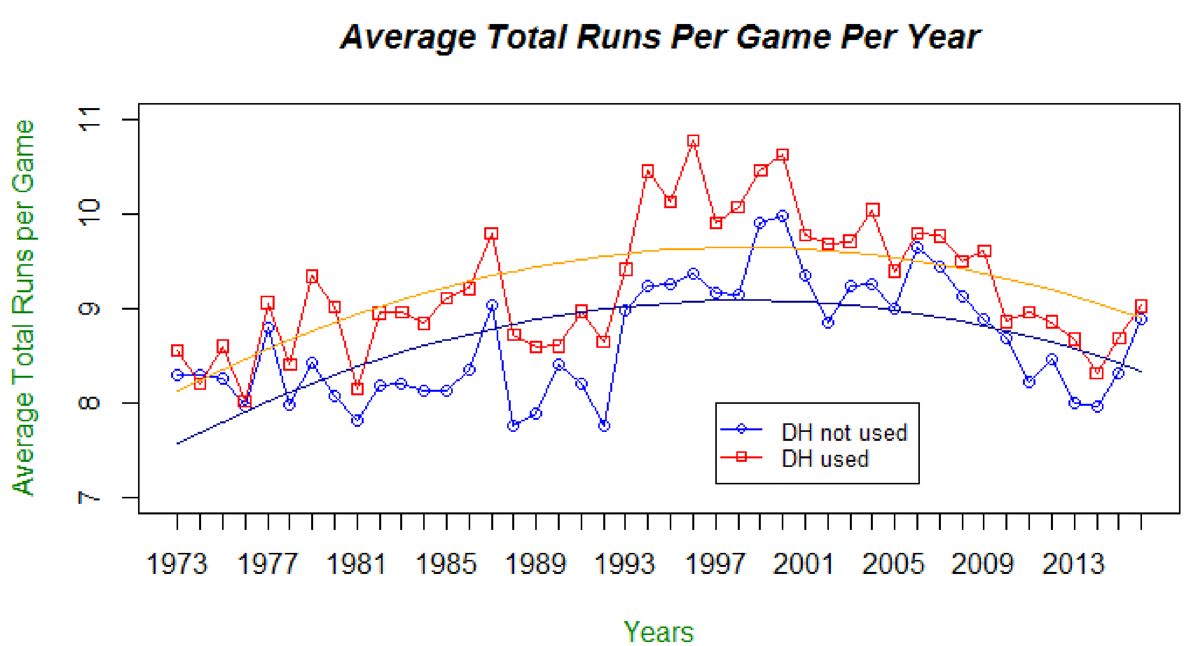
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.4477116  0.0496917  149.88   <2e-16 ***
Year         0.1247557  0.0047209   26.43   <2e-16 ***
I(Year^2)   -0.0023745  0.0001004  -23.65   <2e-16 ***
DH_Used      0.5587941  0.0288030   19.40   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.488 on 97289 degrees of freedom
Multiple R-squared:  0.01139,	Adjusted R-squared:  0.01136 
F-statistic: 373.8 on 3 and 97289 DF,  p-value: < 2.2e-16
```

A low P value of $2.2e^{-16}$ for the DH being used, means that there is enough evidence to support a significant effect of having a DH. We calculate two regression models based on all the data. A quadratic regression model without the designated hitter, $y = A + Bx + Cx^2$. And a quadratic regression model with the designated hitter, DH quadratic regression: $y = A + Bx + Cx^2 + D$. From R, we can determine these coefficients:

- A= 7.447712
- B= 0.1247557
- C= -0.00237452
- D= 0.5587941
- n=97294 games
- i=each game

A difference of 0.5587941 total runs added per game with DH greatly influences the game. This means that a team playing with a designated hitter is likely to score 0.27939705 more runs per team. Since 1973, 28760 games have been determined by one run. This means that 12 games a year are likely to be changed because of a designated hitter. A few more wins a season will influence whether a team will make the playoffs.

The graph below shows the average total runs per game with the added quadratic equations for both populations.



4.3 Interleague Play

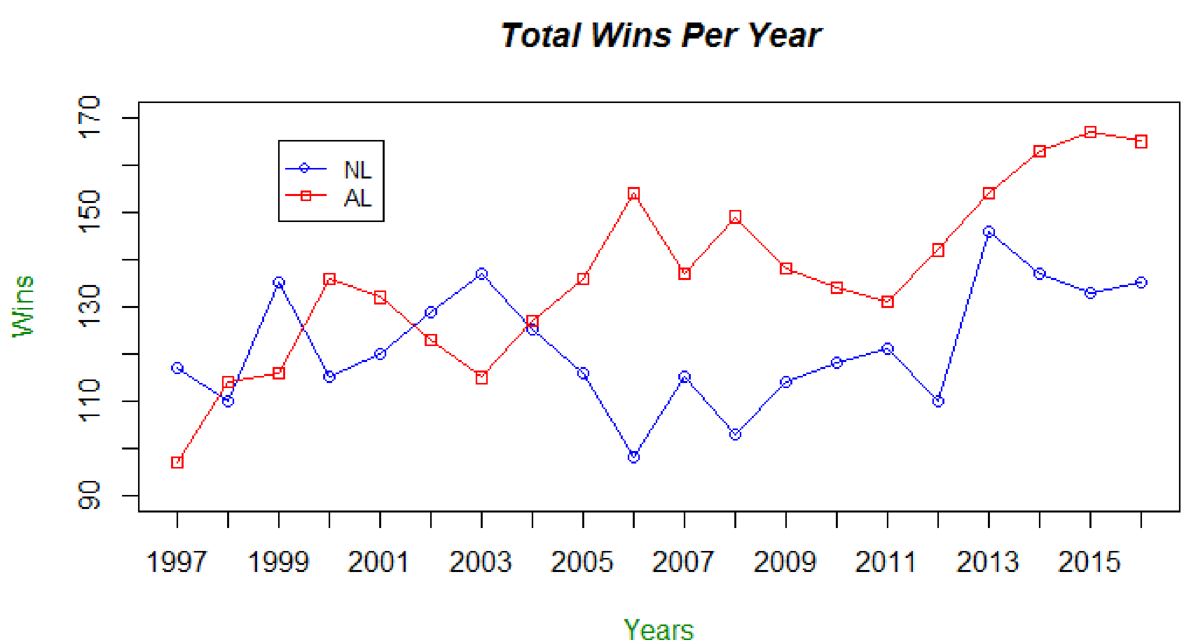
In the 1997 season, interleague play was introduced to MLB. Previously, teams in the AL would only play teams from the NL in the World Series. But for the last 20 years teams in both leagues have played each other in the regular season. The home ballpark determines the use of the designated hitter. In an AL ballpark, the use of a DH is permitted for both teams. This means that although a NL team may not hold a roster spot for a designated hitter, they are welcome to use a designated hitter for that game. Since 2013, teams only play 20 interleague games out of 162 games a season.

Since an NL team only plays 10 games in an AL ballpark, therefore, NL teams do not hold a roster spot for designated hitters throughout the season. When NL teams get to use a designated hitter, the best hitter on the bench will play as the DH since they can still hit better than pitchers. When an interleague game is played in a NL ballpark, both teams pitchers are responsible for their own offensive output. Occasionally, when a team possess a powerful designated hitter, the team will move the designated hitter to a field position, to retain that powerful hitting but sacrifice fielding in the process.

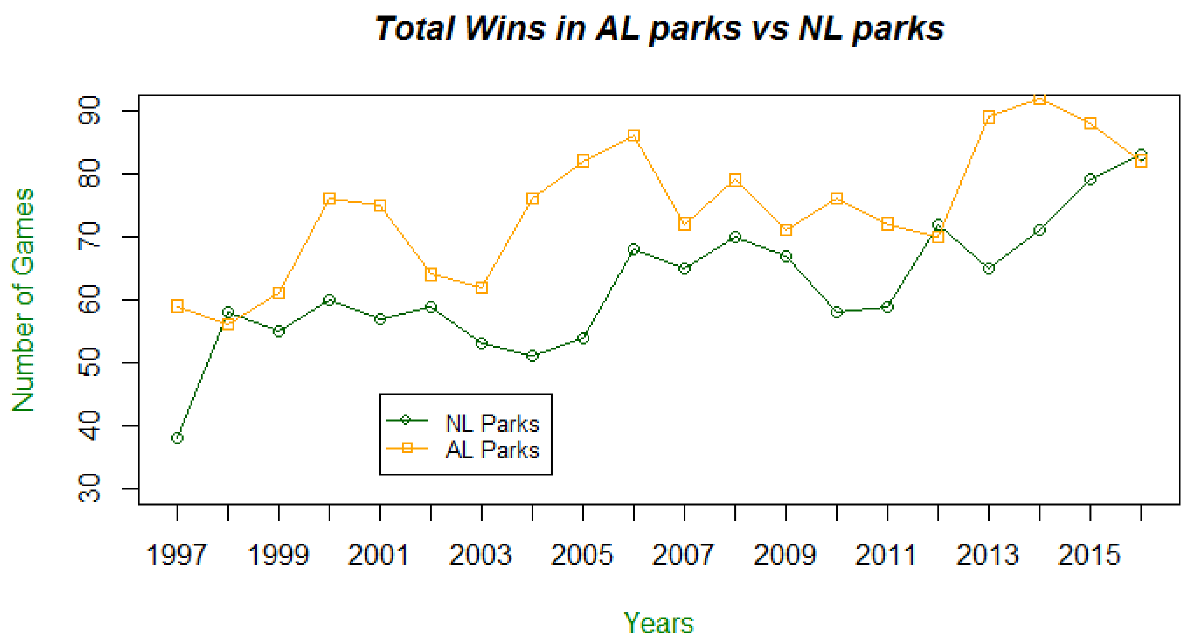
Due to pitchers having relatively similar low batting averages in both leagues little advantage is held by the NL whose pitchers will have the benefit of seeing more at bats than their counterparts in the AL. Pitchers in the AL do not have to hit and do not prepare and practice hitting. A DH is also used in the NCAA, so pitchers in MLB are not given the opportunity to practice hitting before they turn pro. Some claim that this is responsible to a number of pitchers injuring themselves. Most notably, Chien-Ming Wang suffered a foot injury while running the bases in 2008 that missed the rest of the season. [2]

4.3.1 Wins by League

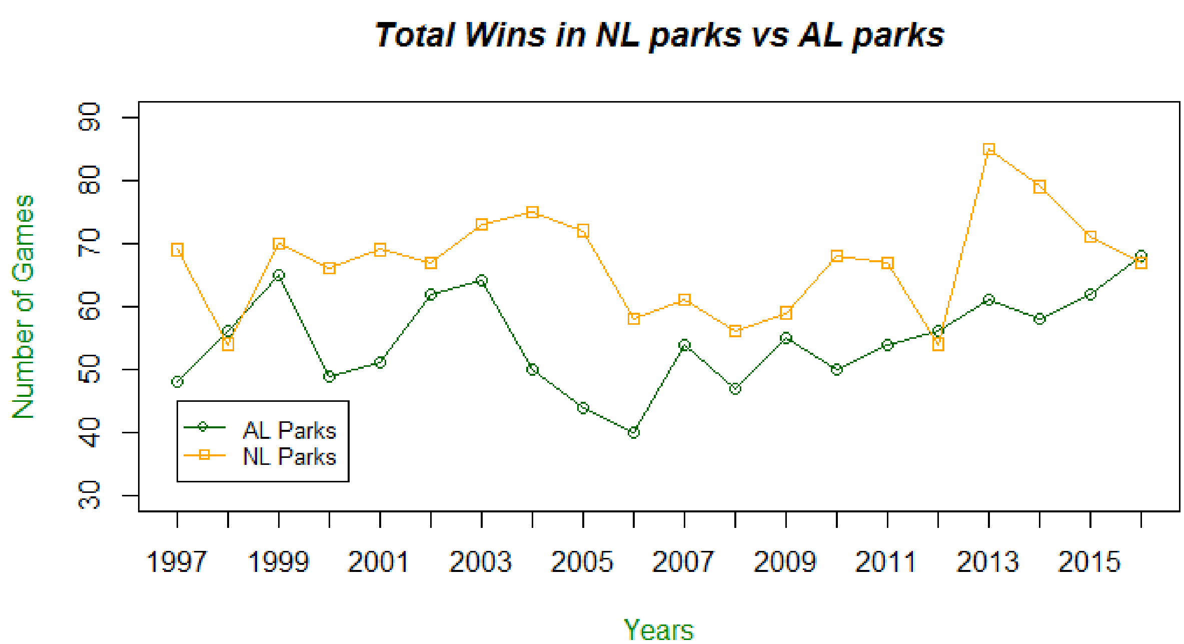
The graph below shows the number of AL wins vs the number of NL wins during interleague play. As you can see that in 16 of the 20 years, the AL won more games than the NL.



The graph below shows the number of AL wins in AL parks and in NL parks during interleague play. As you can see home advantage is real, with AL winning more games at home then away in all but two seasons (1998 and 2012 are the exceptions).

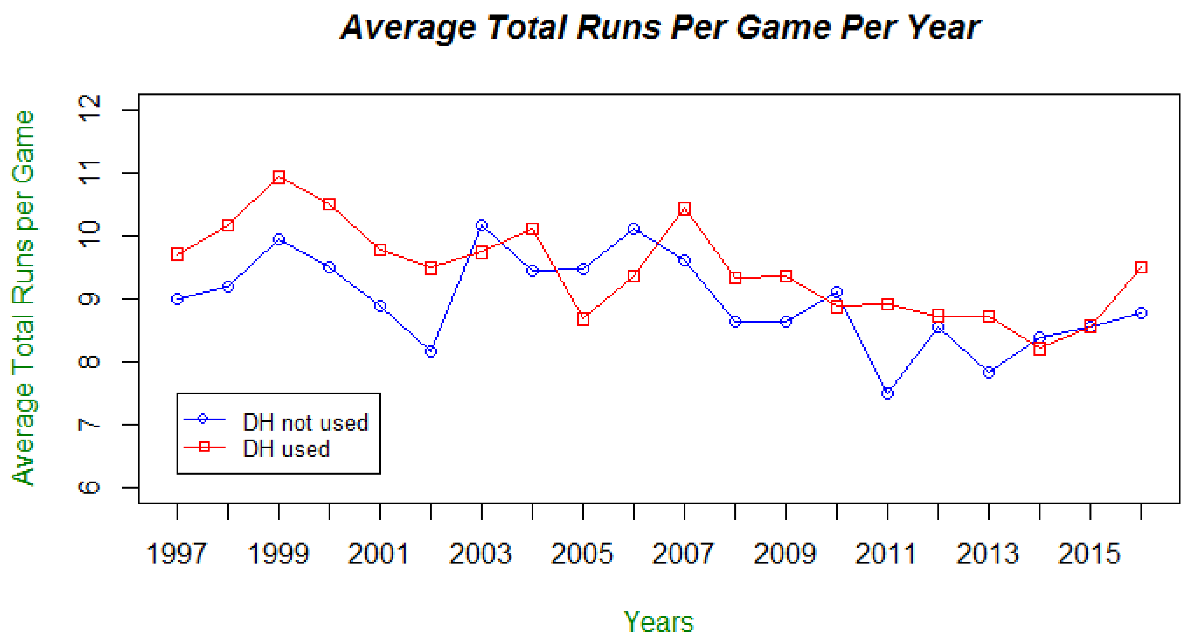


The graph below shows the number of NL wins in AL parks and in NL parks during interleague play. As you can see home advantage is real, with NL winning more games at home then away in all but two seasons (1998 and 2012 are the exceptions).



4.3.2 Average total runs scored per game in interleague play

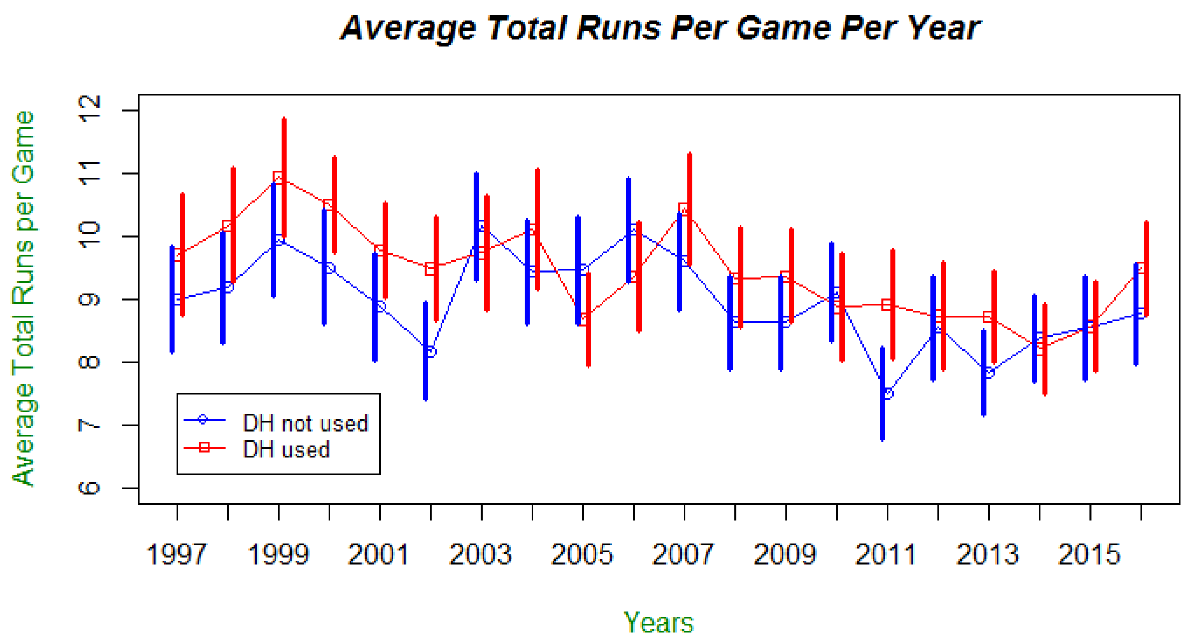
We also calculated the mean total runs scored per game in interleague games. This allows us to see the effect of the DH in match ups between teams from both leagues. Again, we address each season as an independent sample to ensure that any rule changes will not overlap any seasons. Below is the graph of the mean total runs scored in each interleague game in each season since 1997.



In 14 of the 20 years, the AL out preformed the NL in terms of average total runs per game in interleague play. Due to the steroid era beginning well before 1997, and ending in the late 2000s, we see an overall decline in average total runs scored in the graph.

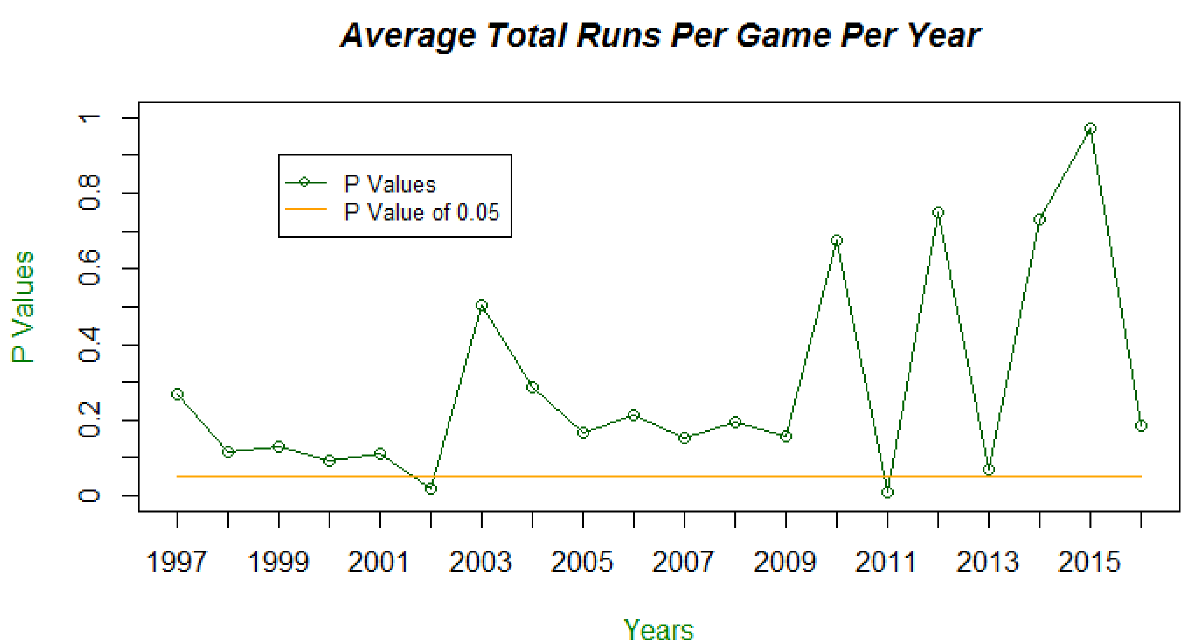
4.3.3 Confidence intervals

Using the equations from sections 3.2 and 3.3 equations we can calculate the upper and lower bounds for this data. On the next page we can see the graph of the mean total runs scored in each game in interleague play in each season since 1997 with the confidence intervals added. In all 20 seasons the confidence intervals overlapped between the AL and the NL.



4.3.4 P values

Using the equations from section 3.4, we can calculate the P values for every season from 1973 to 2016, to see if there is statistical significance that the DH influence the total runs in interleague games. Only two of the 20 seasons had a P value less than 0.05.



5 Conclusion

5.1 All games (1973 – 2016)

The P values were significant in 33 of 44 years. This means that in 33 of the 44 years we would reject the null hypothesis and accept the alternative hypothesis. In 33 of the 44 years, was there enough evidence to suggest that the DH did have an effect on the average total runs per game.

Although the confidence intervals overlapped in 22 of the 44 years, the P values from before showed us that some of these years still showed statistical significant between the means. The range of the confidence intervals is about one run per game meaning that the average total runs could be about 0.5 more runs than the actual mean or 0.5 less runs than the actual mean. Due to a wide range in the confidence intervals this indicates that these are unstable estimates.

5.2 Interleague Play (1997 – 2016)

The P values were significant in 2 of 20 years. This means that in 2 of the 22 years we would reject the null hypothesis and accept the alternative hypothesis. In only 2 of the 20 years, was there enough evidence to suggest that the DH did have an effect on the average total runs per game.

The confidence intervals always overlaped each other. The confidence intervals for interleague play also ranged about a run more than the range of the confidence intervals all games in the regular season, this was caused by a reduced number of interleague games compared to the number of games in the regular season. The range of the confidence intervals is about two runs per game meaning that the average total runs could be about 1 more run than the actual mean or 1 less run than the actual mean. Due to a wide range in the confidence intervals this indicates that these are unstable estimates.

6 References

[1] "Baseball Rule Change Timeline." Baseball Almanac. N.p., 2016. Web. 01 June 2017.

[2] "Designated Hitter" Wikipedia. Wikimedia Foundation, 31 May 2017. Web. 01 June 2017.

[3] "Download Lahman's Baseball Database." Sean Lahman — Database Journalist. Sean Lahman, 2016. Web. 01 June 2017. (Player data)

[4] "Major League Baseball." Wikipedia. Wikimedia Foundation, 30 May 2017. Web. 01 June 2017.

[5] "Play-by-Play Data Files." Retrosheet Event Files. Retrosheet, 6 Dec. 2016. Web. 01 June 2017. (Game data)

[6] "Sabermetrics." Wikipedia. Wikimedia Foundation, 31 May 2017. Web. 01 June 2017.

[7] "Tools, John Galt's. "Probability Of Playing College and Professional Baseball." Probability. N.p., 2016. Web. 01 June 2017.