An Investigation of the Impact of Race in Hall of Fame Inductions in Major League Baseball

Micheas Yimam
Math Major Class of 2019
College of Wooster
Wooster, OH
myimam19@wooster.edu

Abstract

This paper looks into the impact of race and time period for Major League Baseball players chance to get inducted into the Hall of Fame. The null hypothesis is there is no racial bias when comparing the past to now with being inducted in the Hall of Fame. By constructing a two-way ANOVA, finding which variables are significant, we will see if race or the time period the player was nominated affected them. Additional analysis will be ran to support the results from the two-way ANOVA. **Keywords:** two-way ANOVA;

Background

In 1947 Jackie Robinson broke the color barrier 18 years before the conclusion of the Civil Rights movement. He was taunted, threatened, and physically attacked during his career in the Major Leagues due to the color of his skin. From his experiences and his career Jackie Robinson was still a First ballot Hall of Famer in 1962. However, were all other players of color treated the same way? Jackie Robinson was an anomaly he was a great player and a great person who didn't wave at the test of racial slurs, but how would players who were not as successful as Jackie fair in getting inducted in the Hall of Fame. Would time affect a racial bias if there is one in the induction process for players of color? All of these questions will be addressed with the two-way ANOVA.

Process and Data

The first steps of any type of analysis is to breakdown the project into steps. The first step is to create a time frame; collecting data from 1962 (the year Jackie Robinson was on the ballot for the first time) to the present would allow me to get the largest data set possible to test my hypothesis of racial bias. The site baseball-reference.com[1] where every player from every hall of fame ballot is stored. We gathered data relating to the percentage of votes they received. For a player to be inducted into the hall of fame they need a vote percentage above 75 percent. We also used a performance stat WAR (wins above replacement). Wins above Replacement (WAR), is a statistic that generates the number of wins a player adds to a team compared to a replacement level player. WAR has become almost industry standard for evaluation players worth. With some excel formulas we were able to deduce a Boolean result if a player was inducted. However for finding if the player was player of color we had to go one by one, and verify their race. Since we are attempting to look at the racial bias throughout time we have to break my large dataset into groups. When first thinking about how we would break apart the data set we wanted to use momentous dates for racial significance in baseball as date changes. However, after Jackie Robinson most of those moments occurred before or right around 1962 anyway. The next best way was to divide the groups into three equal groups of hall of fame election years; two groups of 19 years and one group of 18 years. Below in table 1 you will see the data we used and how it was broken down.

Table 1: Breakdown of Data

1962-1980	Inducted	POC	Vote%	WAR
Group 1	0 or 1	0 or 1	0-100	+ Integer
1981-1999	Inducted	POC	Vote%	WAR
Group 2	0 or 1	0 or 1	0-100	+ Integer
2000-2018	Inducted	POC	Vote%	WAR
Group 3	0 or 1	0 or 1	0-100	+ Integer

Initial Findings

From the above data set we have a total of 2000 entries composed of each player separated into one of the three groups. The next step now is to take these data sets and create some understanding of the data by creating visualizations. Once all the visualizations are created the next steps would be to interpret that information and refill our data table to get a bigger a picture of what is going on. Then from there we can start the two-way ANOVA and begin to analyze the data. This stage is very important for any type of statistical analysis because without understanding what the data says you

can't analyze the data. First lets separate players of color and non players of color from our whole dataset.

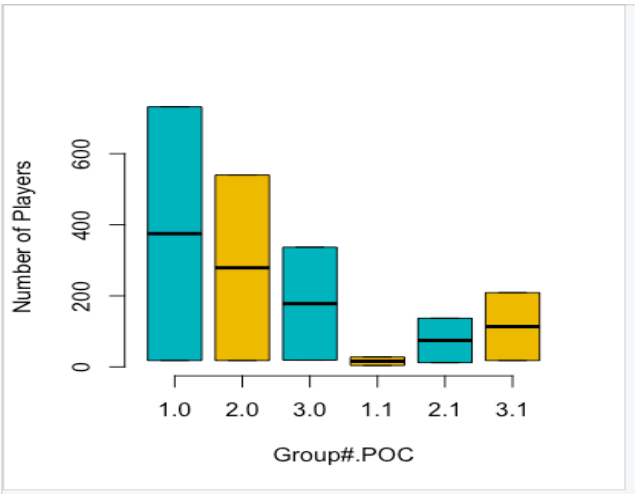

Figure 1: Number of players in each group year categorized by race

For Figures 1-3 it is important to note that the labeling on the x-axis is Group.POC. For example group 2 with non players of color would be 2.0, but group 2 with players of color would be 2.1.

What stands out is the large set of non players of color being considered for the Hall of Fame. As we can see as we move from group 1 to group 3 the colored players consideration for induction have increased significantly. This is expected and can be explained without a racial bias. The percentage of players of color playing baseball compared to Caucasian did not hit the 40% mark until 2014 [2], so the number of players being considered for the hall of fame should be skewed against players of color because they aren't evenly distributed across the league. The next step would be to look at the player's performance throughout there career's. The best way to do that is with the same groupings look at the average WAR accumulated by those players. We see this in figure 2 on the right.

From this plot we see an almost surprising outcome. Across the board the non players of color are roughly around a 58 WAR, while the players of color have a large fluctuation in there WAR ranging from 58 to 63. When looking at figure 2 the reader will take away that on average the players of color out preform the non players of color. It is important to remember the figure 1 that showed the number of players for each group, so it is possible that having a larger set of players is bringing down the average. However, we will investigate that later from this information though we should expect that the average voter percentage for players of color will be higher than non players of color because of performance. On the right in figure 3 we see the graphical representation of average voter percentage grouped.

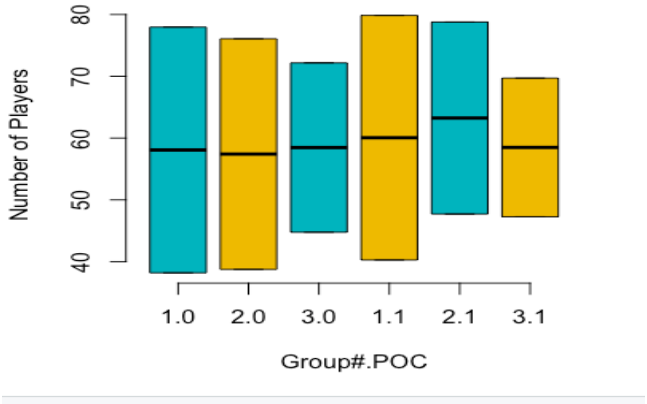

Figure 2: Average WAR in each group year categorized by race

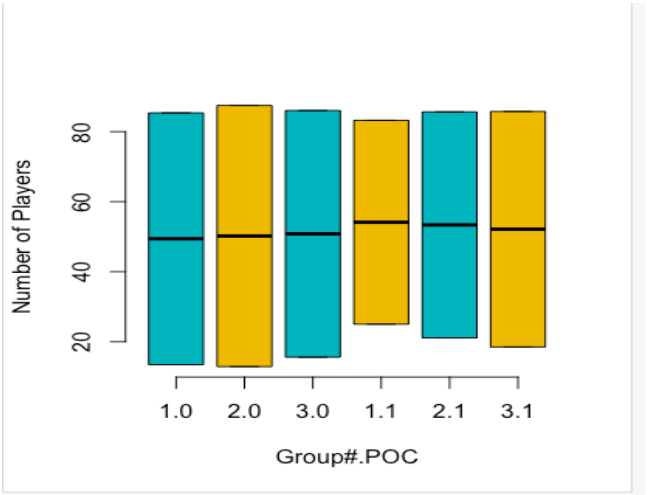

Figure 3: Average WAR in each group year categorized by race

Surprisingly the voter percentage is where it should be at a glance. The players of color have a higher vote percentage across the board. The vote percentage for the non-poc is consistently the same just like the average WAR across the time periods. Now that we have looked at our data in the three time periods and categorized for race we should break down the numbers even more. When looking at the table on the next page we will see the number of players inducted, total number of players, average vote %, and average WAR. Two tables will be made one for non players of color and one for players of color.

Table 2: Collection of Data Totals Non-POC

1962-1980	Inducted	NPOC	Vote%	WAR
Group 1	18	750	49.42	58.08
1981-1999	Inducted	NPOC	Vote%	WAR
Group 2	18	558	50.19	57.42
2000-2018	Inducted	NPOC	Vote%	WAR
Group 3	19	356	50.81	58.47

Table 3: Collection of Data Totals POC

1962-1980	Inducted	POC	Vote%	WAR
Group 1	4	32	54.12	60.06
1981-1999	Inducted	POC	Vote%	WAR
Group 2	12	149	53.38	63.25
2000-2018	Inducted	POC	Vote%	WAR
Group 3	18	227	52.14	58.48

Once again we see players of color dominate non players of color for both the vote percentage and total WAR. Yet, those two data tables alone do not tell the whole story. If we add an additional filter on our data where if the player was inducted in the hall of fame or not. We could separate the large assortment of additional non players of color who aren't inducted but included on the ballot unlike players of color.

Table 4: Collect of Data with Filters

1962-1980	Inducted	POC	Vote%	WAR
Group 1	YES	NO	85.38	77.96
Group 1	NO	NO	13.44	38.19
Group 1	YES	YES	83.258	79.85
Group 1	NO	YES	25	40.27
1981-1999	Inducted	POC	Vote%	WAR
Group 2	YES	NO	87.5	76.07
Group 2	NO	NO	12.883	38.75
Group 2	YES	YES	85.06	78.78
Group 2	NO	YES	21.09	47.71
2000-2018	Inducted	POC	Vote%	WAR
Group 3	YES	NO	86.05	72.17
Group 3	NO	NO	15.57	44.76
Group 3	YES	YES	85.77	69.7
Group 3	NO	YES	18.49	47.27

What we notice when we add the additional feature of comparing players who got into the hall of fame we see a different trend. Up to this point we would have believed that players of color on the ballots were out preforming the non players of color and in turn getting a higher vote percentage. This is only true when you look at the players who are not inducted into the hall of fame, but when you look at the players who are inducted we see a troubling occurrence. Players of color in groups 1 and 2 are out preforming non players of color in average WAR, but are receiving about 2 percent less in votes. However in group 3 this is not the case which consists of the 21st century ballots. Alas the assumption that as time goes on the racial bias on induction will impact these players less.

There are a lot of possible explanations for these results. Such as in the population sample we are testing are not equal among players of color and non players of color. Group 3 is the first time almost the same number of players of color (18) and non players of color were inducted into the hall of fame (19), yet there are 129 additional non player of colors being considered when there WAR's as a collective are almost identical 58.48 (poc) to 58.47 (non-poc). It seems that when comparing the players of color who are the cream of the crop, the ones being inducted, are not getting the same amount of votes compared to non players of color with a lower average WAR. Yet, when we compare them as a whole the larger amount of non-colored players bring the collective voter percentage down to give the misdirection that there isn't a racial bias. The only way to test this now is to introduce the two-way anova.

Analysis

ANOVA Testing

"The acronym ANOVA refers to analysis of variance and is a statistical procedure used to test the degree to which two or more groups vary or differ in an experiment. In most experiments, a great deal of variance (or difference) usually indicates that there was a significant finding from the research." [3] When comparing the F values the largest has the greatest impact. If we find that the p values are less than 0.05 after testing then we can conclude that there is a significant difference between the means and the variables have significant impact. In our case we are looking to see if the variables race and time period have a significant impact on the results. We are going to build a two-way ANOVA to test this.

Table 5: Two-Way Anova Matrix

	DF	Sum.Sq	Mean.Sq	F Value	P value
A:Group	-	-	-	-	
B:Race	-	-	-	-	
A and B	-	-	-	-	
Residuals	-	-	-	-	

This procedure usually suggests to run a single ANOVA test first to see if there is any reason that two variables would both have significant impact on the outcome. However with my initial findings we can assume both these variables will have an effect due to the vote % and WAR relationship. For me to conclude that through the years racial bias existed

among the voters the two way anova will have all p values below 0.05. Any variable with a p value above 0.05 means they don't have a significant impact on the player being inducted. Running the two-way anova test would have to be done in a statistical package, and we ran it in R. The output of the testing is below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
`Year Group`	2	0.42	0.2086	5.126	0.006015
Race	1	0.53	0.5334	13.107	0.000301
`Year Group`:Race	2	0.15	0.0735	1.805	0.164699
Residuals	2066	84.08	0.0407		

Figure 4: Two-Way ANOVA Results

The results in figure 4 from the two-way anova was not what we expected. With significantly low p values the year group (0.006) and race (0.0003) both were very significant, but when you put the variables together the p value (0.16) is greater than 0.05 meaning that when you combine the variables there is no significant impact in getting into the hall of fame. This is strange that separately both the variables held a significance but when intertwined they weren't significant. Thus we must accept our null hypothesis from our anova testing, but we have conflicting results. The F value tells us from these three effects Race is the most statistically significant because it is the largest F value. From the ANOVA results we can not say as time goes on race is less of an impact because the combined factor was not significant, but we can still say that time and race were significantly impacting inductions.

Limitations of ANOVA Testing

ANOVA testing is not a fool proof procedure there are needs to it, such as the data being used in the ANOVA testing are normally distributed.[3] There is a way to test this with the Shapiro-Wilk test. This test looks at the residuals of the ANOVA tests to create a graph (Figure 6) on the next page, and the results (figure 5) below.

From the Shapiro Wilk Test we see that our ANOVA test failed. We know this because the calculated value W has to be larger than 0.98 and the p-value has to be larger than 0.5; both of those are not the case with our data. Because our ANOVA residuals had a W value of 0.9488 and a p-value less than $2.2*10^{-16}$. Because our data set is not normally distributed we can not take the results as confidently as we

```
        Shapiro-Wilk normality test

data:  IndAnova2Residulas
W = 0.94888, p-value < 2.2e-16
```

Figure 5: Shapiro Wilk Test values

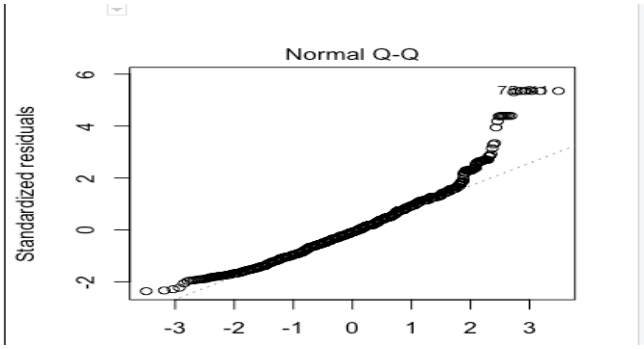

Figure 6: Shapiro Wilk Test Residual Graph

could have meaning we need to add additional evidence that race is a significant effect.

Simulations

One way to test if race is significant is by running simulations on our data set. We know the percent of inducted players of color for each year group. To do this we randomly select the number of player of colors from the entire data set of players and calculate the percentage of those players who got into the hall of fame a a multitude of times in a simulation we can see where the induction percentage for players of color stands in comparison. For example the percentage of players of color inducted in Group 1 was 12.5%, and then we select the total number players of color (32) from all the possible players in Group 1(782), and calculate the percent inducted.

From the three graphs (figures 7-9) on the next page we see that the percent of players of color inducted is significantly higher than we anticipated compared to the percent of any player being inducted into the hall of fame. In fact these simulations do validate the ANOVA results that race and year are significant variables but not as a racial bias towards players of color.

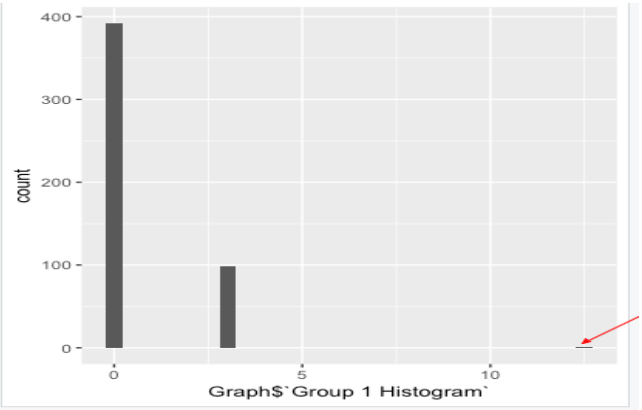

Figure 7: Group 1 Histogram, red arrow points to POC induction rate

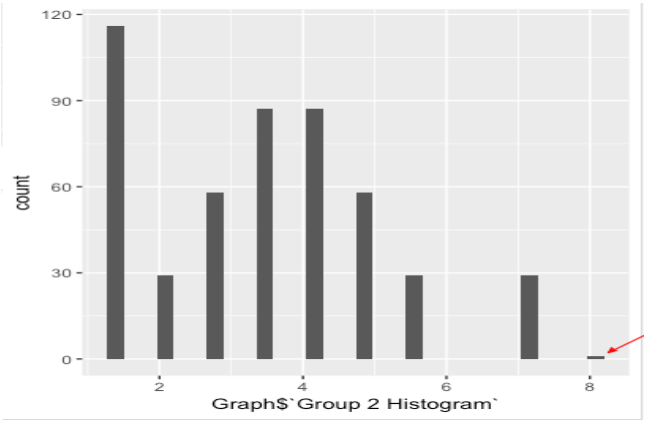

Figure 8: Group 2 Histogram, red arrow points to POC induction rate

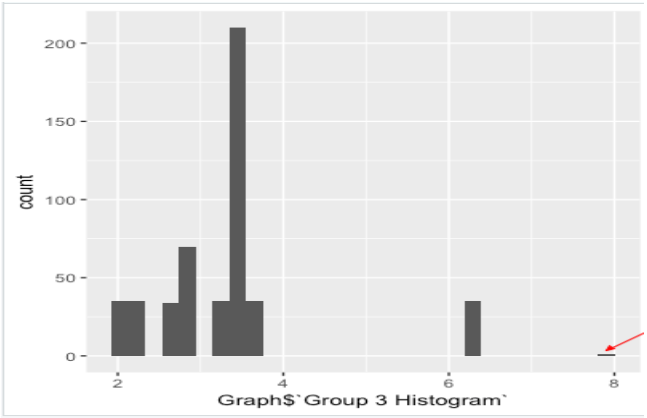

Figure 9: Group 3 Histogram, red arrow points to POC induction rate

Discussion

In the end from my experiments we can not conclude there was a racial bias towards players of color because when looking at the players who were inducted they were inducted at a higher rate then all the players as a whole. However, that doesn't exclude the possibility of a racial bias in this process. When looking at data we have to recall players of color out preformed non players of color whether they were inducted or not. Specifically when looking at Group 1 where players of color out preformed non players of color and received less votes regardless of they were inducted or not. The testing done in this paper did not look at the selection process of players being inducted, but that could be a informative extension. Another extension could be creating linear models to see performance and race as variables and see which is a better predictor for induction.

Other limitations of the dataset that need to be discussed is the voting rituals changed [2]. Many voters when the Hall of fame started would only fill out a ballot for one player instead of the 10 they were allowed to vote for. The changes in voters vote was shown in the ANOVA results because year was a significant impact because as time went on voters became less stringent on who could get inducted. In no way is this limitation the cause of inconclusive results, yet presenting this as a possibility is important in any statistical paper.

In conclusion, a racial bias throughout the years that diminishes in modern times that affects the induction of players in the Hall of Fame was not found. However statistical evidence from the ANOVA test suggests year and race both are significant factors in getting into the Hall of Fame. The simulations showed in fact from the fewer players of color on the ballot were inducted at a higher rate than non players of color. However due to the larger number of non players of color is dragging that rate down; how much that is impacting my results is unknown. Thus the next step in this analysis would be incorporating one of the above extensions to look for more definitive answers.

Key Terms

Two Way Anova is used to compare population means when populations are classified according to two factors. ANOVA assumes that the populations are normal with possibly different means and the same standard deviation and that independent SRSs are drawn from each population. ANOVA separates the total variation into parts for the model and error. The model variation is separated into parts for each of the main effects and the interaction. The calculations are then organized into an ANOVA table, where F statistics and P-values are used to test hypotheses about the main effects and the interaction.

Player of Color is a baseball player who is not Caucasian, so an African American, Latino, Asian, etc.

Shapiro Wilks Test A test on the residuals from the ANOVA with two computed values W and a. When W is below 0.98 and a is below 0.5 we know that the data set

tested in the ANOVA is not normally distributed.

Wins Above Replacement (WAR) is an all inclusive reference point that explains the worthiness of a player. WAR offers an estimation on if a player is replaced by an average player, and at the same time provides context for a player's performance throughout there career.

References

[1] baseballreference. baseballreference.com.

[2] Arna Desser. Baseball hall of fame voting: A test of the customer discrimination hypothesis. *Social Science Quarterly, vol. 80, no. 3, 1999*, page 591603.

[3] David S. Moore and George P.McCabe. *Introduction to the Practice of Statistics*. W.H Freeman and Company, 2002.