An Analysis of Utilizing the Cutoff Man from the Outfield in Baseball

https://github.com/kulic1ja/SMT-Data_Challenge_Kulich

Abstract

Baseball has always been a game where defenders are trying to get outs. To get those outs the defenders must line up in the field to defend it when the ball is hit in play. Defenders have always interacted with each other through the throwing of the ball. The outfielder almost always throws the ball back in play towards the infield. This paper goes into whether a cutoff man is important in terms of throwing speed. Does hitting the cutoff man or going straight to the intended target make a difference in terms of speed? With this question in mind, this paper goes into a deep dive into figuring out if the cutoff man should be utilized and in what situations. Using the mean in the difference between intended target and the outfield along with confidence intervals we can see if this makes a difference. Using these we can see that it truly depends on the situation that an outfielder is in that determines whether the cutoff man should be utilized. This benefits the game because it helps us to see if hitting the cutoff man makes the ball travel to the intended target faster or not.

Introduction/Background

Since the beginning of baseball, defenders have lined up in various ways to try and retrieve the ball when it is put into play. One of these two things happens either the ball is put into play and is recorded as an out since it was hit toward a defender, or the ball is not hit at a defender and the batter advances to a base. Most of the time the defenders interact with each other through throwing the ball. Whenever the ball is hit into the outfield the player must throw the ball back into the infield. Sometimes the ball is cut off by a player or the ball goes straight to the intended target without any cutoffs. Does hitting the cutoff man or going straight to the intended target make a difference in terms of speed? With this question in mind, the objectives for this project are to investigate if it is quicker to throw the ball to the intended target or should the cutoff man be utilized in certain situations. There are many different situations where a cutoff man could be utilized. This could be in a situation where it is thrown to the catcher, third baseman, shortstop, second baseman, and first baseman. By looking into this we can get a good understanding of the cutoff man and its importance in the game of baseball. The solution that I used for this was by using the data in the folders for the SMT Data Challenge and looking into the average time in differential between both cutoffs and when the ball went straight to the intended target. This solution should help give us an idea of whether a cutoff man is worth it in terms of throw speed. I will now dive deeper into what data was used and why I used it.

Data

The data that was utilized for this project mainly came from the game events, ball position, and player position spreadsheets. The game events data showed each play from a game from a pitch to a pickoff. This was used because I needed to find the plays that included a throw from the outfielder back into the infielder. I grabbed all the plays starting from the pitcher pitching the ball to the outfielder acquiring the ball and then to the ball reaching the intended target of any player in the infield. I grabbed all the data from that spreadsheet and put it into a “plays needed” spreadsheet with left field, center field, and right field tabs. The tabs meaning which outfield threw the ball back into the infield. The next data that I used was from the ball position spreadsheets. I used ball_position_x, ball_position_y, and ball_position_z. This was needed so I would know where the ball position was when the ball was thrown and where the ball ended at its intended location. I also grabbed position data from the player position spreadsheet so I could know where the players were as well when looking at the analysis. All these data points were essential to making sure I could analyze where the ball is and look at the time it took to get from point A to point B with or without a cutoff man. After all the plays were brought into a spreadsheet, I needed to investigate what the intended target was of the throws from the outfield and grab the plays that had all the same intended target so I could investigate the average difference of time with or without a cutoff.

The way that I defined a cutoff man was when the outfielder threw the ball and if it was cutoff by a player before it reached the intended target. For example, if the right fielder was throwing the ball with an end point of the third baseball and the second baseman caught the ball and threw it to third base, the second baseman would be considered the cutoff man. When analyzing this I did not take any base runners into account because I did not deem it necessary for them to be taken into consideration especially in the context of my research question only looking at the throw speed.

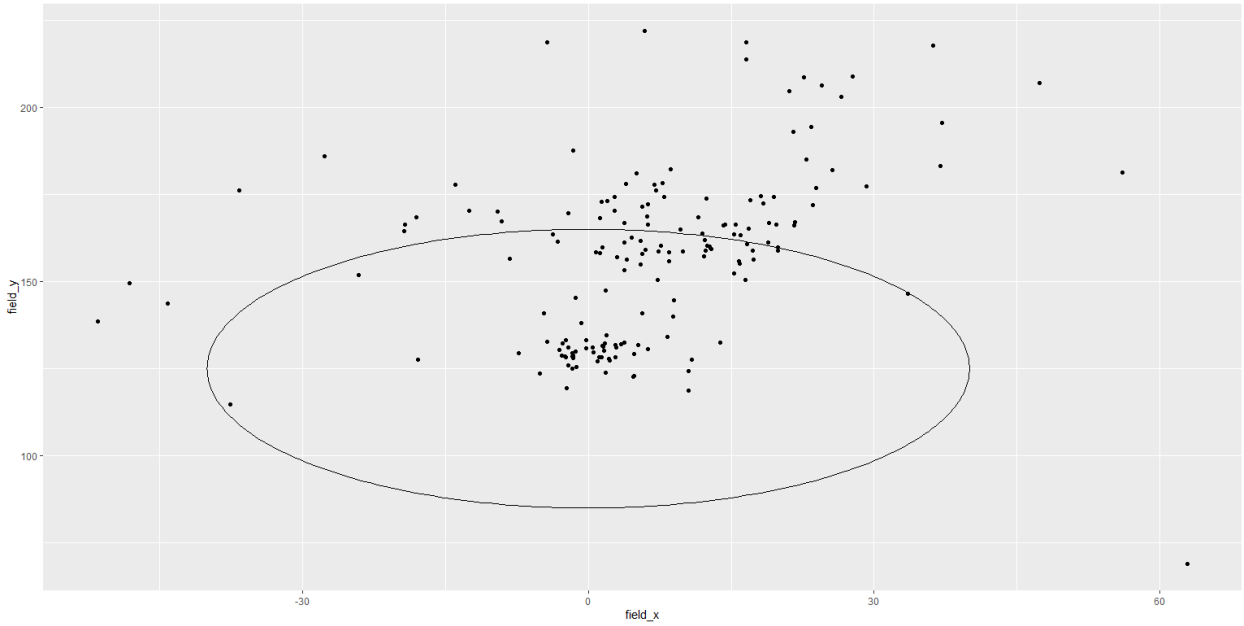
Using this data, I was able to investigate and take a difference between when the ball was acquired at the intended target and when the ball was acquired by the outfielder. I used subsets of data for when the ball was thrown from a fielder to a specific player in the infield. For example, when the ball

was thrown to the catcher, there is a dataset that has throws from left field to the catcher, center field to the catcher, and right field to the catcher with an average time difference between cutoff man and no cutoff man.

Methodology

Since the problem I am looking to solve is regarding the speed of the ball coming from the outfield back to the infield, what we really need to investigate is the time difference between when the ball gets to the intended target and when the ball was acquired in the outfield. I then took an average of these time differences with a split between no cutoff man and with a cutoff man involved. This approach to the analysis made it a simplified way to looking at whether a cutoff man makes a difference in terms of the speed it took to get from the outfield to the intended target. With this simplified approach in mind, I was hoping it would be easier to look at and make assumptions. By looking at all the plays that were given in the dataset we can get a true average in the difference between hitting the cutoff man or not. Another way that I analyzed this data was by using a confidence interval of 0.95. A confidence interval of 0.95 gives us a range of where the population or actual mean could end up with 95% confidence. The method I used to determine the confidence interval was since I wanted to figure out the population mean. Using these ranges, we can see if there is an overlap of the range of means between cutting the ball off and not. This helps us to see if there is value or no value in hitting the cutoff man.

Some of the assumptions that I made were about a general area in which baseball was acquired and where the different bases were. For example, on a play that was thrown to second base, the data had general area that ranged from different locations, so I just assumed that those plays were right in being near second base. Below is a visual of plays that ended near second base with a circle of where I considered plays to be an end target of second base.



Another assumption that I made was that the talent of each outfielder was the same. I know different outfielders have different talent but for the sake of making this analysis simple I assumed equal talent among all the outfielders that made plays in this dataset.

Discussion/Conclusion

With the question and methodology now defined, let's investigate the analysis of the data that I have. The tables below show what the intended target or result of the play was, where it was thrown from and the average time and confidence interval that it took for the ball to reach that target with or without a cutoff man. The 'lci' means that it is the lowest limit where the mean can be and the 'uci' means that it is the highest limit where the mean can be. Here are the tables shown in milliseconds:

Catcher Intended Target

	left_field_catcher$Type	mean	lci	uci
1	Cutoff	5045.250	4106.241	5984.259
2	No Cutoff	3713.158	3364.311	4062.005

	center_field_catcher$Type	mean	lci	uci
1	Cutoff	5585.714	5232.821	5938.607
2	No Cutoff	3684.692	3385.025	3984.359

	right_field_catcher$Type	mean	lci	uci
1	Cutoff	6219.571	4308.226	8130.916
2	No Cutoff	3955.030	3553.687	4356.374

Third Baseman Intended Target

	left_field_third$Type	mean	lci	uci
1	Cutoff	5569.000	4730.287	6407.713
2	No Cutoff	3482.039	3302.238	3661.840

	center_field_third$Type	mean	lci	uci
1	Cutoff	5650.500	4161.604	7139.396
2	No Cutoff	3567.682	3222.652	3912.711

	right_field_third$Type	mean	lci	uci
1	Cutoff	5024	4697.693	5350.307
2	No Cutoff	4240	3635.021	4844.979

Shortstop Intended Target

	left_field_short$Type	mean	lci	uci
1	No Cutoff	3059.783	2978.544	3141.022

	center_field_short$Type	mean	lci	uci
1	Cutoff	5366.667	873.993	9859.340
2	No Cutoff	3224.371	3142.534	3306.208

	right_field_short$Type	mean	lci	uci
1	Cutoff	5240.778	4586.513	5895.043
2	No Cutoff	3579.690	3419.006	3740.374

Second Baseman Intended Target

	left_field_second$Type	mean	lci	uci
1	Cutoff	5592.571	4997.748	6187.395
2	No Cutoff	3531.449	3369.989	3692.910

	center_field_second$Type	mean	lci	uci
1	Cutoff	4773.077	3988.598	5557.556
2	No Cutoff	3172.245	3078.085	3266.406

	right_field_second$Type	mean	lci	uci
1	Cutoff	7025	5436.724	8613.276
2	No Cutoff	3028	2956.926	3099.074

First Baseman Intended Target

	left_field_first$Type	mean	lci	uci
1	Cutoff	4800	3529.380	6070.620
2	No Cutoff	4080	3338.112	4821.888

	center_field_first$Type	mean	lci	uci
1	Cutoff	5828.400	4590.415	7066.385
2	No Cutoff	3702.455	3054.959	4349.950

	right_field_first$Type	mean	lci	uci
1	Cutoff	4844.000	-11597.829	21285.829
2	No Cutoff	3331.786	3073.725	3589.846

Looking at these tables we can see that every time a cutoff man is utilized in an outfield assist it slows down the time it takes for the ball to reach its intended target. Even with the ball bouncing before it reaches the intended target, the average time is still faster than hitting the cutoff man. Below is an example of this from the left fielder to the catcher. As you can see when the ball is bounced with no cutoff it is still faster than the no bounce with the cutoff.

	LFCatherNoBounceCutoff	LFCatcherBouncedCutoff	LFCatcherBouncedNoCutoff	LFCatcherNoBounceNoCutoff
1	4.52	5.920667	3.625	3.811111

This is since the cutoff man must catch the ball, turn, and throw the ball to where the intended target is which adds more time to this occurrence. Now while the means shows that not hitting a cutoff man is a quicker approach when throwing a ball to its intended target, this does not always mean that the cutoff man should never be utilized. When the ball is hit into the deeper half of the outfield, an average outfielder should hit a cutoff man because there is little to no difference in the amount of time in that situation whether a cutoff man is hit or not. For example, with a play that had the same amount of deep depth hit to the right fielder the time for a cutoff was 5200 milliseconds and for no cutoff was 6100 milliseconds. This shows that the cutoff man is still essential to when the ball is hit to the deeper part of the field.

Looking at the intervals we see overlaps in the right field to catcher, right field to third, center field to short, and left field to first plays. With seeing these overlaps, we can say that the cutoff man is still an important part to the game. This makes sense because it depends on where the ball is hit into the outfield. As from the example above, it does not make a huge difference if the ball is cutoff or not when it is hit into the deeper part of the outfield. The overlap in confidence intervals shows just how close together all the plays are whether a cutoff man is hit or not.

This benefits the game because it helps us to see if hitting the cutoff man makes the ball travel to the intended target faster or not. From this analysis we can see that on average hitting the cutoff man slows down the ball in terms of reaching the intended target. But we also see that the cutoff man is still an essential part of the game. If the ball is hit to the deeper part of the outfield the cutoff man could help speed it up in some cases. Knowing the situation, where the ball is going, and where the ball was hit to in the outfield play important factors in determining whether the cutoff man should be utilized or not.

Improvements/Future Work

There are different players who play the game, but this analysis is done based on assuming that all players are of the same talent level to keep the analysis fair. Now in the real world this is not true, some outfielders have better arm talent than others. The answer to the question is that not hitting the cutoff man makes the ball travel faster, but it can vary based on different players. A future enhancement to this project is to do an analysis on each individual outfielder to see if there are better outfielders than the others.

Another future enhancement I could do with this project is to make a model that you input where the ball is hit and where you want to throw the ball to and make a prediction of whether the cutoff man should be hit or not.

Another future enhancement would be to take a deeper dive into the ball positions and player positions and determine where the ball is going. Mentioned above, I only used a generalized area for where the ball went and where the players were located.