On a potential mercy rule for Major League Baseball

by

Rajiv Jung K C

B.S., Texas Tech University, 2020

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. Michael Higgins

Copyright

Abstract

Major League Baseball (MLB) games are getting lengthier, having gone up from an average of 2.5 hours in the 1970s to over three hours since 2012. Meanwhile, the average attendance in an MLB game has recently decreased from around 30,800 in 2012 to less than 28,300 in 2019. In this report, we consider a solution to shorten games and increase their entertainment value through implementing a mercy rule, also called a run rule, in which a game may end early in the case of a blowout. We analyze data from 24,296 MLB regular season games between 2010 and 2019 to identify potential mercy rules for which early termination is unlikely to affect the final result. We find that calling a game after the seventh inning if a team is up by seven or more runs will not change the outcome of a game 99.9% of the time, and this rule may impact about 10% of all games each season. In addition, this mercy rule will decrease the game time by an average of about 40 minutes. We then discuss the practical implementation of this rule.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Words are not enough to describe how grateful I am to the Department of Statistics at Kansas State University for this tremendous opportunity to pursue my graduate studies in Statistics with being fully funded by a Graduate Teaching Assistantship.

I am thankful to Dr. Christopher Vahl and Dr. Michael Higgins for their perpetual guidance from day one throughout my graduate studies, not to mention for the former's role as a committee member and the latter's as the major professor in this research. This work would not have been possible without Dr. Michael Higgins' patience and active mentorship. I also thank Dr. James Higgins for developing this exciting research idea and advising on the analysis.

Last but not least, I would like to thank my parents and my sister for their endless love and encouragement.

Chapter 1

Introduction

A mercy rule ends a two-competitor sports competition earlier than the scheduled point if one contestant has a vast and seemingly insurmountable scoring lead over the other. Because it spares further humiliation for the losing team, it is called the mercy rule. It is common in amateur leagues in the US, where running up the score is considered unsporting. It is prevalent in baseball and softball, where there is no game clock and a dominant team, in theory, could continue an inning ceaselessly. As the name implies, such rules save further embarrassment to the losing team players and keep needless injuries from occurring to both sides when the game's outcome is a foregone conclusion.

In Kansans High School baseball, where regulation play is seven innings, the mercy rule is the popular 10-run rule. The rule states that if the home team is ahead by 10 runs after $4\frac{1}{2}$ innings or the visiting team is ahead by 10 runs after five innings, the game is called. The threshold for this rule may vary in regional contests. (KSHSAA, 2022)

However, as the most profitable sports league on the planet after NFL (National Football League) (Randjelovic, 2020), with an average per season player salary of $4.4 million (Blum, 2022), the MLB has different dynamics. The games are played for the entertainment of paying fans and are supported by paying sponsors, so not playing the games for the full nine innings might cause a concern that fans and sponsors are not getting their money's worth. Also, the element of "mercy" should not be a consideration for the professional players as they are paid to perform regardless of the score.

Yet increasingly, there are concerns that MLB is losing its entertainment value. (Panacy, 2013) The games are getting lengthier, having gone from an average of about $2\frac{1}{2}$ hours in the 1970s to over three hours since 2012. (Baseball-Reference, 2022) A descriptive analysis of our data shows that the average MLB game length has increased in recent years while the average MLB game attendance has decreased. This result is concerning given that the US population has increased by 7.5% (23 million) from 2010 to 2020. (US Census, 2022)

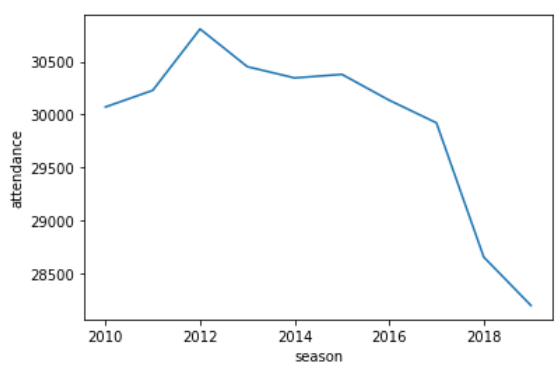

Figure 1.1: *Average attendance*

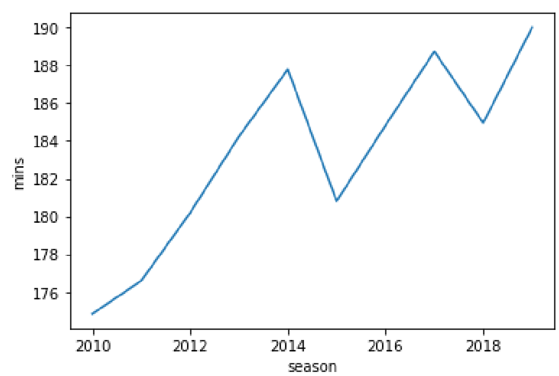

Figure 1.2: *Average game length*

1.1 Reasons to consider exploring mercy rule

While the embarrassment for the players or showing mercy to the losing team is not a consideration, there are several good reasons for calling the game when the outcome is a foregone conclusion. The first is to prevent boredom among the fans. When one team is far ahead in the late innings, the fans often leave the stadium because no entertainment value is left, which diminishes the game's worth for sponsors and broadcasters. Nothing may be more tedious and difficult for broadcasters than to fill in airtime when nothing of any significance is happening on the field. The quality of play can also diminish. For instance, teams may use position players to do the pitching to save the regular pitchers for truly contested games. New York Yankees manager Aaron Boone believes the mercy rule is worth considering for this situation. (Horrobin, 2021) Players' injuries are also a consideration. Why risk injuries when the game's outcome is unlikely to change with further play?

1.2 Purpose of this study

If a mercy rule were to be considered by MLB, two obvious questions must be answered:

i. The number of innings[1] that should be played before the rule is invoked?

ii. How significant a run differential[2] must there be for the game to be called?

Hence, this study aims to provide the necessary statistical information and analysis to MLB to make informed choices for the inning and run differential for the mercy rule to be applied should such a rule be implemented.

1.3 Data used for this study

The data set used for our research is acquired from Retrosheet[3], a nonprofit organization whose website freely features box scores of MLB games from 1906 to the present and play-by-play narratives for almost every contest since the 1930s. Retrosheet informally began in 1989 through the efforts of Dr. David W. Smith, a biology professor at the University of Delaware, and fellow baseball enthusiasts. (Smith, 2000) By 2007, Retrosheet had been cited as a source in multiple large American newspapers, including The Boston Globe (The Boston Globe, 2007), Chicago Tribune (Chicago Tribune, 2003), Detroit Free Press (Detroit Free Press, 2003), and Pittsburgh Post-Gazette (Pittsburgh Post-Gazette, 2003).

Our research uses data from the 2010-2019 MLB seasons, with box score data on 24,296 contests. One of the contests between the Chicago Cubs and Pittsburgh Pirates played on September 29, 2016, which ended in a 1-1 tie due to rain, is removed. (Cubs Vs. Pirates, 2016, 2022) While Retorsheet contains data on 161 different fields/variables[4] from a MLB

[1] An inning is a basic unit of play consisting of two halves: the top and the bottom.
[2] A team's run differential = the total runs allowed - the total runs scored
[3] https://www.retrosheet.org
[4] https://www.retrosheet.org/gamelogs/glfields.txt

game, only the following variables deemed most relevant to our research are used for this study.

Table 1.1: *Variables used*

Field	Meaning
1	Date in the form "yyyymmdd"
4	Visiting team
7	Home team
10	Visiting team score (final)
11	Home team score (final)
18	Attendance
19	Time of game in minutes
20	Visiting team box score
21	Home team box score

1.4 Literature Review

The state of existing research on the mercy rule in baseball seems to be limited to the philosophical aspects of the game primarily intended for amateur leagues. More specifically, researchers from the intellectual landscape have argued for or against the anti-blowout thesis (AB) to reason whether or not to use a mercy rule in baseball. AB states, "it is intrinsically unsporting for players or teams to maximize the margin of victory after they have secured victory in a one-sided contest." (Dixon, 1992) Some philosophers have also introduced a term called a revised AB to argue that running up the score is unsporting as long as it is not justified. The revised AB states, "It is prima facie unsporting for players or teams to maximize the margin of victory after they have secured the win in a one-sided contest". (Feezell, 1999) While the literature supporting the AB or the revised AB seems to favor mercy rule in some ways, the ones arguing against it are not for it.

Dixon (1992), argues against a widely held view concerning sportsmanship, the AB. He concludes that no complete arguments give any good reason for condemning the pursuit of victory by a wide margin as fundamentally unsporting. He also emphasized that dramatic comebacks are possible in competition until the final inning has been played as a reason for the leading team to run up the score to reduce the likelihood of such a comeback. (Dixon, 1992)

As a response to Dixon, Feezell (1999) argues by creating a modified version of AB which he calls a revised anti-blow-out thesis. Feezell does not deny that there are circumstances in which pursuing blowouts might be justified. Blowouts might have some strategic significance for future games played in a series. For example, in professional baseball, a blowout in a playoff series's first game might give the opponent a psychological blow and shake their confidence for the next round. On the other hand, trashing an opponent and then easing up, in a strategic sense, can also have the same effect on their confidence. However, Dixon (2000) claims that in baseball, it is not clear that the revised AB comes into effect. In contrast to "time-based" sports, where a significant lead becomes unassailable when only a small amount of playing time remains, surprise late-inning comebacks are always possible in baseball.

In defense of the revised AB, Hardman et al. (1996) raises concerns if the winning is only the thing that matter the most. He states that the dominant view of running up the score when the win is secured is meaningless and even points out that it is a symptom of our culture rather than a cause of sport's dilemma. Thus, he seems to believe in some form of the mercy rule.

Moreover, Sailors (2010) examines various mercy rules used in different sporting contexts to avoid prolonging athletic contests when they become blowouts. Sailors argue that such regulations should be upheld and perhaps expanded. Sailors object to the claim by stating that fans quickly become bored at best and unruly at worst during lopsided contests. Sailors conclude that while blowouts may not be intrinsically morally wrong, there are good reasons to encourage the use of mercy rules.

While these past studies are insightful in exploring mercy rule, they are subjective. They don't provide quantitative information on the most critical questions, i.e., when is a game considered a blowout to invoke a mercy rule if we were to implement it? How big of the lead is considered too big to overcome? Technically, it is possible to overcome a deficit of 10 runs after the seventh inning but is it probable? How frequently does it occur in every 1000 games? Hence, we answer these questions based on data analysis through our research.

1.5 Current state of mercy rule

Although commonly used in amateur leagues, the mercy rule and other rules to shorten a game do not seem uncommon in professional baseball competitions, including MLB. Here, we discuss the cases where such practices prevail.

1.5.1 MLB

i. Rained out rule in the event of inclement weather - If a game has completed the top half of the fifth inning and the home team is ahead, the game can be considered an official game where the home team is declared the winner. On the other hand, if a game has completed the bottom half of the fifth inning and either team is ahead, the game can be considered an official game where the leading team is declared the winner. And if the teams are tied after the fifth inning, the game is suspended and will be resumed at a future date. (MLB Rules, 2019)

ii. Mercy rule in spring training games - MLB has several rules to shorten their spring training matches. One of such rules, dubbed as mercy rule, allows defensive teams to end an inning before three outs if the pitcher has pitched at least 20 pitches. Upon mutual agreement of the competing teams' managers, another rule change allows the game to be shortened to five or seven innings. (Bonesteel, 2021)

iii. Seven innings double headers - Due to the Covid-19 pandemic, the double headers were limited to seven innings each during the 2020-2021 regular season. (Wormeli, 2021)

1.5.2 WBC and Softball

The World Baseball Classic (WBC), an international baseball tournament endorsed by World Baseball Softball Confederation (WBSC) in collaboration with MLB since 2013 (MLB.com, 2022), calls off a game by invoking its mercy rule, with the exception during the championship round, if one team is ahead by:

i. Ten or more runs after any complete inning, beginning with the completion of the seventh inning, or;

ii. Fifteen or more runs after any complete inning, beginning with the completion of the fifth inning. (WBC, 2022)

In softball, rules are different for fast/modified and slow pitch. In World Baseball Softball Confederation (WBSC) - sanctioned competitions, the run-ahead rule (the WBSC terminology for a mercy rule) is, for fast or modified fast pitch, 20 runs after three innings, 15 after four, or eight after five. In slow pitch, the margin is 20 runs after four innings or 15 after five. (Official Rules of Softball, 2005)

Chapter 2

Analysis and Results

In this chapter, we discuss the data analysis to answer our study's purpose - to make the choices for the innings and run differential for the mercy rule to be invoked should such a rule be implemented. Throughout our research, we denote an inning with i and run differential with r. In addition, we state the mercy rule in the form (r, i) which means a game where a team is ahead by r or more runs at the end of the i^{th} inning. The data collection and analysis are categorized into the following sections.

2.1 Processing and cleaning data

The raw data on Retrosheet was contained in text format. Therefore, we merged all the season data into one file and saved it as a CSV file in tabular form. We used Python programming language and its Numpy and Pandas libraries to carry out this task. One notable work on processing data was to convert team box scores into a tabular format where we have a score for each innings in a separate column. Initially, the box scores were in a string format like this - "010000(10)0x" which indicates the game where the home team scored a run in the second inning, ten in the seventh, and didn't bat in the bottom of the ninth. While this may seem as simple as converting a string to a list of characters, a double-digit run can be scored in any inning, making the brackets in the original box score dynamic that need to be removed before converting it to the list of characters.

After this, we worked on data cleaning. We started with checking the missing, duplicates, and irrelevant data. Then, we converted the data types into their respective format as all the data was in a text format. For example, the date column was converted to date type, numbers to integers, and time to minutes. Afterward, we ran a summary statistic of the numerical data to check for outliers.

Finally, we calculated the run differential between the home and visiting teams by subtracting their respective cumulative runs scored at the end of each inning for all the games. For instance, suppose Team A has scores of 1, 2, 2, 1, 4, 6 and Team B has 0, 0, 1, 2, 3, 4, then their cumulative run scores are 1, 3, 5, 6, 10, 16 and 0, 0, 1, 3, 6, 10, and finally the run differential for Team A on that game is $1, 3, 4, 3, 4, 6$.

2.2 Choosing the inning and the run differential

One important thing to start with when suggesting the number of innings and run differential for a mercy rule is how it could impact the outcome of a game. For example, how often a trailing team overcomes a particular run deficit to win the game? Or, what is the probability that a team remains ahead by a specific run for the remainder of the game and wins? We are interested in finding such instances whose likelihood of occurrence is almost given.

We obtained $p_{i,r}$, the proportion of games where a team would remain ahead by r or more runs after the i_{th} inning for the remainder of the game and win the game, to identify possible candidates for the i and r. The $p_{i,r}$ was calculated from the run differential data from table 2.1 using the following form.

$$p_{i,r} = P(W \mid r \geq r_i) = \frac{P(W \,\&\, r \geq r_i)}{P(r \geq r_i)} = \frac{X}{Y} \tag{2.1}$$

where:

W = a team wins the game

r_i = a game where a team leads by at least r runs at the end of an i^{th} inning

X = total number of games where a team wins after leading by at least r runs after i^{th} inning

Y = total number of games where a team leads by at least r runs after i^{th} inning

Table 2.1: *The proportion $p_{i,r}$ of games where a team would remain ahead for the remainder of the game and win*

		Innings							
		1	2	3	4	5	6	7	8
Run Differential	1	0.6914	0.7217	0.7536	0.7926	0.8280	0.8716	0.9139	0.9542
	2	0.7771	0.8100	0.8340	0.8636	0.8935	0.9267	0.9549	0.9787
	3	0.8508	0.8678	0.8949	0.9159	0.9370	0.9604	0.9788	0.9909
	4	0.8999	0.9123	0.9339	0.9488	0.9612	0.9769	0.9891	0.9965
	5	0.9310	0.9461	0.9620	0.9656	0.9757	0.9889	0.9937	0.9979
	6	0.9912	0.9600	0.9712	0.9765	0.9861	0.9923	0.9966	0.9987
	7	1.00	0.9784	0.9785	0.9869	0.9938	0.9969	$\boxed{0.9992}$	$\boxed{1.00}$
	8	1.00	0.9769	0.9898	0.9894	0.9942	0.9984	1.00	1.00
	9	1.00	1.00	0.9932	1.00	0.9940	$\boxed{1.00}$	1.00	1.00
	10	1.00	1.00	1.00	1.00	0.9964	1.00	1.00	1.00
	11	1.00	1.00	1.00	1.00	$\boxed{1.00}$	1.00	1.00	1.00

.

In the above table, for $p_{i,r} = p_{7,7}$, the proportion of win when a team is ahead by at least seven runs after seventh inning is 0.9992 or 99.992%.

We choose the candidates for the proposed mercy rule from among those values of i and r for which $p_{i,r} \geq 0.999$. We want to determine a blowout game as soon as possible. At the same time, we do not want the rule to change the outcome of that game. Hence, the earliest a mercy rule can be invoked is when a team is ahead by r or more runs at the end of the i^{th} inning given in the form (r,i) are: (11,5), (9,6), (7,7), and (7,8). Only the choices of $i \geq 5$ were considered consistent with the regulation game.

Another factor in recommending values for i and r for a mercy rule is determining the proportion of the total games impacted by each possible mercy rule. The proportion of the

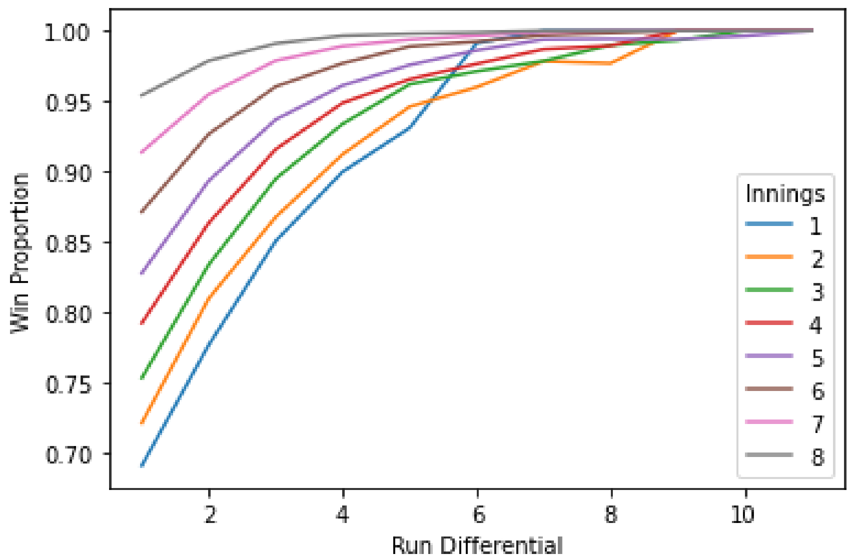

Figure 2.1: $p_{i,r}$

total games where each possible mercy rule can be applied is computed using:

$$\frac{\text{Y (from equation 2.1)}}{\text{Total number of games in our study}} \tag{2.2}$$

Table 2.2: *Proportion of the games affected by mercy rule*

Run Differential	Innings 1	2	3	4	5	6	7	8
1	0.4761	0.6554	0.7513	0.8134	0.8511	0.8716	0.8860	0.8978
2	0.2201	0.3408	0.4544	0.5298	0.5899	0.6336	0.6694	0.7011
3	0.0971	0.1738	0.2635	0.3283	0.3948	0.4428	0.4878	0.5235
4	0.0391	0.0878	0.1508	0.2000	0.2580	0.3034	0.3471	0.3852
5	0.0143	0.0435	0.0833	0.1209	0.1642	0.1996	0.2370	0.2693
6	0.0047	0.0226	0.0458	0.0701	0.1005	0.1287	0.1578	0.1851
7	0.0022	0.0114	0.0230	0.0409	0.0600	0.0800	$\boxed{0.1026}$	$\boxed{0.1230}$
8	0.0012	0.0054	0.0121	0.0233	0.0358	0.0502	0.0651	0.0820
9	0.0005	0.0025	0.0061	0.0128	0.0205	$\boxed{0.0300}$	0.0409	0.0525
10	0.0002	0.0008	0.0028	0.0064	0.0113	0.0167	0.0248	0.0333
11	0.0001	0.0004	0.0015	0.0037	$\boxed{0.0067}$	0.0099	0.0140	0.0202

If there are too few games where we can apply the mercy rule, the rule would have no practical effect, while having too many of such games would change the essential nature of

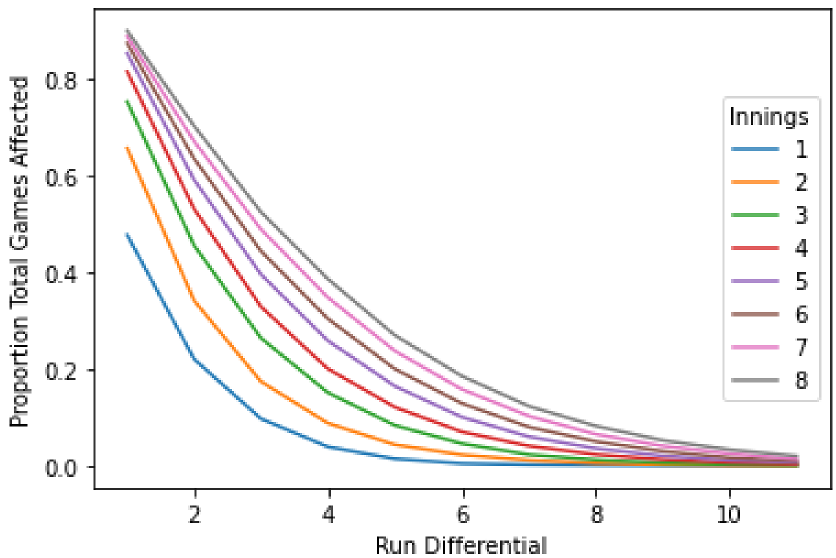

Figure 2.2: *Proportion of the games affected*

the game. Therefore, among the candidates for the mercy rule, (7,7) and (7,8) seem to affect the most number of games while keeping the change in the game's outcome at a minimum, respectively, affecting roughly 10% and 12% of games per season.

If MLB were to implement these two rules, how would it change the outcome of a game? The following table shows the total number of games from the 2010-2019 seasons whose output would have been changed by a mercy rule.

Table 2.3: *Total number of the games whose outcome would get changed by each mercy rule*

Run Differential	Innings: 1	2	3	4	5	6	7	8
1	3570	4431	4497	4099	3557	2720	1854	998
2	1192	1573	1833	1755	1526	1128	733	362
3	352	558	673	671	604	426	251	116
4	95	187	242	249	243	170	92	33
5	24	57	77	101	97	54	36	14
6	1	22	32	40	34	24	13	6
7	0	6	12	13	9	6	$\boxed{2}$	$\boxed{0}$
8	0	3	3	6	5	2	0	0
9	0	0	1	0	3	0	0	0
10	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0

Some fascinating results are observed in table 2.3. For the (7,7) rule, only two games' outcomes would be changed from the 2010-2019 season had this rule been implemented. For example, in the game played between New York Yankees (NYA) and Tampa Bay Rays (TBA) on 09/28/2011, NYA were leading by seven runs at the end of the seventh inning, but TBA scored seven runs in the eighth and ninth inning and one decisive run at the end of the 12^{th} inning to win the game by 8-7. Similarly, in the game played between San Diego Padres (SDN) and Colorado Rockies (COL) on 06/14/2019, COL were leading by seven runs at the end of the seventh inning, yet SDN scored seven runs in the eighth and ninth inning and five runs at the end of the 12^{th} inning to win the game by 16-12. Moreover, no team has ever come back to win the game after being down by seven runs at the end of the first inning.

2.3 Determining effect of mercy rule on game length

Once we have the choices of (r, i) for a mercy rule, the next factor we want to consider is how much time we would save had we invoked the mercy rule. We want to estimate the increase in game length past the blowout threshold of (7,7) and (7, 8) due to the additional runs and innings. We started with estimating how much a run and an inning add to a game's total time using a linear regression model below.

$$Y_k = c + \beta_1 \cdot X_{1k} + \beta_2 \cdot X_{2k} + \epsilon_k \tag{2.3}$$

where:

Y_k = total game length in minutes in the k^{th} game

X_{1k} = total runs in the k^{th} game

X_{2k} = total innings in the k^{th} game

c, ϵ_k = constant and error term respectively

We fit the regression model using "statsmodel's ols" function in Python to estimate the coefficients β_1, and β_2. The summary of the results is as follows:

Table 2.4: *Summary of regression model*

	coef	std err	t	P> |t|	[0.025	0.975]
const	-24.7857	1.183	-20.948	0.000	-27.105	-22.466
run	2.6699	0.026	103.416	0.000	2.619	2.720
innings	20.5956	0.129	159.580	0.000	20.343	20.849

The results from the above table show that a run and an inning add $2.669 + 20.595 = 23.264$ minutes on average to the total game time. Then, using the same results, we computed the average minutes saved for each blowout game due to the additional runs and innings past the blowout threshold using the following equation and calculated their mean.

$$\hat{Y}_k^* = 2.669 \cdot X_{1k}^* + 20.595 \cdot X_{2k}^* \tag{2.4}$$

where:

$\hat{Y}_k^*$ = average time saved in the k^{th} game past blowout threshold

X_{1k}^* = total additional runs in the k^{th} game past blowout threshold

X_{2k}^* = total additional innings in the k^{th} game past blowout threshold

Using equation 2.4, we find $\hat{Y}_k^*$ for each game that meets the blowout threshold of (7,7) and (8,7). Then, we computed the average of each rule's $\hat{Y}_k^*$. For instance, on how $\hat{Y}_k^*$ was computed - in the game played between Boston Red Sox (BOX) and Seattle Mariners (SEA) on 03/28/2019, BOX were up by eight runs after the seventh inning. The game was played until ninth inning and lasted for three hours and 34 minutes, where BOX went to win the game by 12-4. (Baseball-Reference, 2019) Then using (7,7) as the blowout threshold, $\hat{Y}_k^* = 2.669 \cdot X_{1k}^* + 20.595 \cdot X_{2k}^* \rightarrow \hat{Y}_k^* = 2.669 \cdot 0 + 20.595 \cdot 1.5 \rightarrow \hat{Y}_k^* = 30.893$ minutes. In other words, around 31 minutes could have been saved had MLB called the (7,7) rule on this game. In this example, $X_{1k}^* = 0$ because either team scored no runs after seventh inning, and $X_{1k}^* = 1.5$ because SEA did not bat going into the second half of the ninth inning as the winner.

After finding the mean of all the $\hat{Y}_k^*$, on average, the (7,7) rule is expected to reduce a blowout game by 39.55 minutes, and the (8,7) rule is expected to decrease the game length by 16.40 minutes. The amount of time saved can be taken further to narrow down the choice of our mercy rule. When a team is up by at least seven runs after seven innings, and it is already implausible for the trailing team to pull a surprise come back, we want our mercy rule to be invoked as early as possible. Hence, we propose the (7,7) as our final mercy rule.

2.4 Determining if blowout games tend to be longer than non-blowout games

To address the concern that the average baseball game length is getting longer, we want to analyze if the blowout games, i.e., games for which the mercy rule would have been applied, tend to be longer than non-blowout games and if so by how much. Here, a positive difference means that blowout games were longer than the non-blowouts, while a negative difference means vice-versa.

Table 2.5: *Difference in average game length between blowout and non-blowout games*

Season	Time Diff (in Mins)
2010	-0.01
2011	-0.05
2012	-4.06
2013	-3.11
2014	-1.88
2015	-0.22
2016	-2.41
2017	0.42
2018	-0.90
2019	-2.01

The result obtained from this analysis is surprising, as, in almost all the seasons from 2010-2019, the average game length between blowouts and non-blowouts appears to be practically the same. The primary reason is that we used the total game minutes in the analysis based on the actual number of innings, not the (7,7) threshold. Another reason might be that a blowout game could also occur when all runs are scored in one inning while the rest of the game goes quickly. Lastly, the non-blowout games may have a lot of on-field drama, while the blowouts may lack such.

2.5 Determining whether the proportions of blowout games have increased over the years

One factor in explaining the diminishing entertainment value of the game and whether its effect could be mitigated by invoking a mercy rule is determining whether the proportions of blowout games have increased over the years. Here, the proportion of the blowout games in a year is computed by dividing the total number of games in a season, where we can apply the mercy rule by the total number of games in a season.

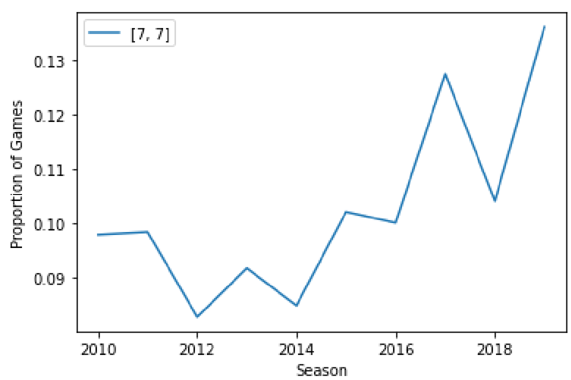

Figure 2.3: *Proportion of the blowouts over the years*

The proportion of the blowout games seems to gradually increase over the years for our proposed mercy rule, i.e., (7,7). Is this increase in blowout games a reason why MLB losing its popularity? (Enten, 2022)

Chapter 3

Discussion

In this chapter, we discuss the practicality of our mercy rule, where we highlight the implementation of the rule. In addition, we also comment on some benefits and some possible drawbacks of our mercy rule.

3.1 Implementation

3.1.1 Case I : Game is a big blowout

When a team leads by ten or more runs at the end of the seventh inning, there likely won't be anything significant going on in the field. If trailing, the defensive team may bring a positional player to pitch. An MLB rule allows a positional player to pitch only when the game goes into extra innings or their team is either ahead or trailing by at least six runs. It may further add boredom to the fans as they are likely in attendance to watch their favorite pitcher pitch. The fans of either team may already know the game's outcome, and at this point, they are just waiting for the official score. However, this situation can be prevented by calling the mercy rule so that everyone gets home early by at least an average of about 40 minutes. Going home would be even earlier if the home team is the one who leads by 7+, as the seventh-inning stretch can be used to walk to the parking lot.

3.1.2 Case II: Team up by five or six runs at the end of the sixth inning

When a home team is up by five or six runs at the end of the sixth inning, the seventh inning may be more competitive than usual. The visiting team may try their best to score at least a run to decrease the run differential and take the game to the following two innings. Similarly, the home team, batting in the bottom half of the seventh inning, may also try to score a couple of extra runs to close the game by invoking the mercy rule. This can be a nail-biter for both teams, and the seventh-inning stretch can be used to hype up the crowd to end the game early.

3.1.3 Case III: Team up by seven or more runs before the seventh inning

When a home team is up by seven or more runs before the seventh inning, they are probably interested in closing the game. As a result, the game may become more competitive as it again puts pressure on the visiting team to score at least a run in the seventh inning so that they can reduce the run differential and take the game to the 8th and 9th inning. If the away team reduces the run differential by a run or two, the home team can again use the seventh-inning stretch to rally the fans. It will potentially look to close the game at the bottom half of the seventh inning by stretching the run differential to at least seven runs.

Similarly, when a home team is down by 7+ runs going into the bottom half of the seventh inning, they may use the seventh-inning stretch to rally fans behind them to play more baseball. Again, the reason is that they may eagerly look to reduce the run differential and take the game to the following two innings.

3.2 Advantages

Mercy rule may prevent embarrassment to the fans and reduce the chances of the players and the crowd getting unruly due to the opposition poking fun at the another.

Another significant benefit of the mercy rule is that it can reduce players' injuries, especially the ones that occur on the field in the late innings of the game. Posner et al. (2011) show that, from 2002 to 2008, 3072 MLB players were placed on the disabled list, with an average of 439 players each year. Once on the DL, a player can't return before a minimum of 15 days. The financial responsibility for this seems exceptionally large. Conte et al. (2016) state that in 2014, the prorated salary cost to players designated to the DL and their replacements was $579,568,059 without even including the medical expenses incurred to treat and rehabilitate the players' injuries.

3.3 Potential drawbacks

While the mercy rule can prevent fan boredom by making a blowout game more competitive, it may have some potential drawbacks.

3.3.1 Financial reasons

Calling a game off at the seventh inning equals fewer innings, fewer television advertisements, and less time for fans to spend money roaming around the ballpark. Hence, a mercy rule seems to decrease MLB's revenue. However, it is also likely that the viewership may go down on a blowout game whose outcome is a foregone conclusion. Fans attending the stadium may stay until the end, but the audience watching from home may get distracted and leave, which may not help MLB further generate viewership and advertisement money.

3.3.2 Unhappy purist fanbase of the game

Shortening a game by a couple of innings might not sit well with the hardcore fanbase of the game. They may not be ready for the change and might take it as a start to radically change the essential nature of the game into something else they are not used to. It may prevent some of them from attending a baseball game and hence less money to MLB due to fewer viewers and attendees. That said, this new change may also mark a symbol of dynamism in MLB and may help attract new fans who otherwise were not interested in the sport due to the long game length and a superior team running score over the inferior.

3.3.3 Less game time for positional players

Ending a game early by two innings may rob the positional players' game time and in-game experiences. It is because they will only be able to pitch when their team is trailing or leading by six or more runs unless it is an extra inning. However, fans may also leave the stadium once the game heads into a blowout which can embarrass the positional players lose their confidence, and harm their growth.

Chapter 4

Conclusion

We purpose mercy rule to call off an MLB game when a team is ahead by at least seven runs at the end of the seventh inning. 99.92% of the 24,295 games analyzed from the 2010-2019 MLB seasons had a team ahead by seven or more runs after the seventh inning for the remainder of the game won the match. We found only two contests in that period whose outcome would get changed by the rule. On average, there are around 10.26% of the games each season where we can apply our mercy rule. Invoking the mercy rule may make the gameplay more competitive and reduce the chances of injuries when the result is a foregone conclusion. The (7,7) rule is expected to facilitate a blowout game's length by an average of around 40 minutes.

Our research may provide a solid starting point for further evaluating the mercy rule for MLB. The availability of data on players' injuries and when they occurred, especially those that appear on the field, may help make a more precise choice of a mercy rule. In addition, the attendance and the viewership data as a game nears the end and the data on the time it takes for an inning may help make a mercy rule more practical.

Bibliography

KSHSAA. Baseball Regulations, 2022. URL https://www.kshsaa.org/Public/Baseball/Regulations.cfm.

Dusan Randjelovic. 11 MOST PROFITABLE SPORTS LEAGUES – THEIR VALUE WILL SURPRISE YOU, June 2020. URL https://apsportseditors.org/others/most-profitable-sports-leagues/. publisher: AP Sports Editor.

Ronald Blum. AP exclusive: MLB average salary up 5.9% after lockout, April 2022. URL https://apnews.com/article/mlb-sports-business-los-angeles-angels-max-scherzer-55c4961f957839a6ae798364c8e32c14.

Peter Panacy. MLB: How Major League Baseball is Losing its Appeal in the Modern Era, 2013. URL https://bleacherreport.com/articles/1566029-mlb-how-major-league-baseball-is-losing-its-appeal-in-the-modern-era.

Baseball-Reference. Major League Miscellaneous Year-by-Year Averages and Totals, 2022. URL https://www.baseball-reference.com/leagues/majors/misc.shtml.

US Census. U.S. Census Bureau QuickFacts: United States, 2022. URL https://www.census.gov/quickfacts/fact/table/US/PST045221.

Jordan Horrobin. Boone: Let's talk about the mercy rule, 2021. URL https://www.mlb.com/news/aaron-boone-says-mercy-rule-is-worth-considering.

David W Smith. The retro sheet. *Retrosheet. org*, 2000.

The Boston Globe. Protecting a big lead. *The Boston Globe*, page 44, May 2007. URL https://www.newspapers.com/clip/75513778/protecting-a-big-lead/.

Chicago Tribune. A cut above .500. *Chicago Tribune*, pages 3–2, September 2003. URL https://www.newspapers.com/clip/75513345/a-cut-above-500/.

Detroit Free Press. Worst After 80. *Detroit Free Press*, page 24, July 2003. URL https://www.newspapers.com/clip/75513481/worst-after-80/.

Pittsburgh Post-Gazette. Baseball (column). *Pittsburgh Post-Gazette*, page 47, June 2003. URL https://www.newspapers.com/clip/75513637/baseball-column/.

Cubs Vs. Pirates, 2016. Cubs vs. Pirates - Game Summary - September 29, 2016 - ESPN, 2022. URL https://www.espn.com/mlb/game/_/gameId/360929123.

Nicholas Dixon. On sportsmanship and "running up the score". *Journal of the Philosophy of Sport*, 19(1):1–13, 1992.

Randolph M Feezell. Sportsmanship and blowouts: Baseball and beyond. *Journal of the Philosophy of Sport*, 26(1):68–78, 1999.

Nicholas Dixon. The inevitability of disappointment: reply to feezell. *Journal of the Philosophy of Sport*, 27(1):93–99, 2000.

Alun Hardman, Luanne Fox, Doug McLaughlin, and Kurt Zimmerman. On sportsmanship and "running up the score": Issues of incompetence and humiliation. *Journal of the Philosophy of Sport*, 23(1):58–69, 1996.

Pam R Sailors. Mercy killing: sportsmanship and blowouts. *Journal of the Philosophy of Sport*, 37(1):60–68, 2010.

MLB Rules. Rules | Glossary, 2019. URL https://www.mlb.com/glossary/rules.

Matt Bonesteel. Spring training 'mercy rule,' explained: MLB teams take advantage of 'rolling the inning'. *Washington Post*, 2021. ISSN 0190-8286. URL https://www.washingtonpost.com/sports/2021/03/02/mlb-mercy-rule/.

Ryan Wormeli. How the coronavirus pandemic changed baseball's rules, 2021. URL https://www.nbcsports.com/washington/nationals/how-coronavirus-pandemic-changed-baseballs-rules-possibly-forever.

MLB.com. About | World Baseball Classic, 2022. URL https://www.mlb.com/world-baseball-classic/about.

WBC. World Baseball Classic, 2022. URL https://www.mlb.com/world-baseball-classic.

Official Rules of Softball, 2005. URL http://www.internationalsoftball.com/english/rules_standards/Rulebook_2002.pdf.

Baseball-Reference. Boston Red Sox at Seattle Mariners Box Score, March 28, 2019, 2019. URL https://www.baseball-reference.com/boxes/SEA/SEA201903280.shtml.

Harry Enten. Why is baseball no longer America's Game?, 2022. URL https://www.cnn.com/2022/04/07/sport/mlb-opening-day-baseball-popularity-spt-intl/index.html.

Matthew Posner, Kenneth L Cameron, Jennifer Moriatis Wolf, Philip J Belmont Jr, and Brett D Owens. Epidemiology of major league baseball injuries. *The American journal of sports medicine*, 39(8):1675–1691, 2011.

Stan Conte, Christopher L Camp, and Joshua S Dines. Injury trends in major league baseball over 18 seasons: 1998-2015. *Am J Orthop*, 45(3):116–123, 2016.

Appendix A

Code

Note: The following code is in Python 3.10. For future researchers - if you need any clarification on the code, please email me at rajiv-kc (at) outlook.com with subject line "K-State Mercy Rule Research".

```
# ================================================================================
# Go to https://www.retrosheet.org/gamelogs/index.html to download
# the individual season data 2010-2019 in zip format
# ================================================================================
# ================================================================================
# Extract individual zip files
# ================================================================================
import glob, os
import zipfile
import pandas as pd
# file path to the folder where the zip files are saved
data_file_path = r"C:\...\Master's Report\scraped data"
os.chdir(data_file_path + "\historic_data\zip_files")
for zip_file in glob.glob("*.zip"):
    my_zip = zipfile.ZipFile(data_file_path + "\historic_data\zip_files"
```

```
    + "\\" +zip_file)
    storage_path = data_file_path + "\historic_data\\txt_files"
    for file in my_zip.namelist():
        my_zip.extract(file, storage_path)

# ==============================================================================
# Combine all the season data into one dataframe
# ==============================================================================

path = "C:/.../Master’s Report/scraped data/historic_data/txt_files/"
os.chdir(path)
# Get the season files for the last 12 seasons
file_name_by_season = [file for file in glob.glob("*.txt")][-12:]
file_name_by_season.reverse()

mlb_game_log = pd.read_csv(path + file_name_by_season[0], header = None,
                           parse_dates=[0],
                           dtype = { 19 : str, 20 : str})
games_per_season = [len(mlb_game_log)]
for file_name in file_name_by_season[1:]:
    mlb_season = pd.read_csv(path + file_name, header = None,
                             parse_dates=[0],
                             dtype = { 19 : str, 20 : str})
    games_per_season.append(len(mlb_season))
    mlb_game_log = mlb_game_log.append(mlb_season, ignore_index = True)

sum(games_per_season)
mlb_game_log
```

```
# find and remove tied games
mlb_game_log[mlb_game_log[9] == mlb_game_log[10]].index.values
mlb_game_log = mlb_game_log.drop(index=12995).reset_index(drop=True)

# save the data until above point
# save as parquett to save some space
import fastparquet
# parquet needs string as column names
mlb_game_log.columns = [str(i) for i in range(len(mlb_game_log.columns))]
mlb_game_log = mlb_game_log[(mlb_game_log['0']>'2009-12-31') &
(mlb_game_log['0']<'2020-01-01')]
# mlb_game_log.to_parquet(data_file_path + "\mlb_game_log.parquet")

data_file_path = r"file path to the saved folder"

# load the data
import pandas as pd
mlb_data_par = pd.read_parquet(data_file_path +
"\mlb_game_log.parquet", engine='fastparquet')

# ===============================================================================
# Data Cleaning & Preparation
# ===============================================================================

# select data since 2010 season excluding 2020 and 2021 seasons due to COVID-19
mlb_data_par.info() # structure of the variables

# column name and the number of missing values
cols_with_nans = {i:mlb_data_par[i].isnull().sum() for i in
```

```
mlb_data_par if mlb_data_par[i].isnull().sum()>0}

# check duplicate rows
mlb_data_par.duplicated().sum()

# ===============================================================================
# Get box scores data
# ===============================================================================

# get the max no. of innings for H/V games
max_inn_v = max([(len(i)) for i in mlb_data_par['19']]) - 2
# becasue there are two paranthesis
max_inn_h = max([(len(i)) for i in mlb_data_par['20']]) - 2

# check no. of games that had more than one double digits score in the game
max([i.count(')') for i in mlb_data_par['19']])
# 1 means there were no games that had more than one two digits score in a game

# rename columns to distinguish home/away matches
column_names = ['V'+str(i) for i in range(1,max_inn_v+1)]
df_visit = pd.DataFrame(columns = column_names)

# Get box score for visit teams
''' this chunk of code until line 139 takes more than
        40 mins to run '''
for score in mlb_data_par['19']:
    test = list(score)[1:-1]
    if '(' and ')' in test:
        start = test.index('(')
```

```
        end = test.index(’)’)+1
        test[start:end] = [‘’.join(test[start : end])]
        test = [i.strip(‘(|)’) for i in test]
        df_visit = df_visit.append(pd.Series(test, index=df_visit.columns[:len(test)]),
                                   ignore_index=True)

    else:
        df_visit = df_visit.append(pd.Series(test, index=df_visit.columns[:len(test)]),
                                   ignore_index=True)

# df_visit

# Get box score for home teams
column_names = [‘H’+str(i) for i in range(1,max_inn_h+1)]
df_home = pd.DataFrame(columns = column_names)
for score in mlb_data_par[‘20’]:
    test = list(score)[1:-1]
    if ‘(‘ and ’)’ in test:
        start = test.index(‘(‘)
        end = test.index(‘)’)+1
        test[start:end] = [‘’.join(test[start : end])]
        test = [i.strip(‘(|)’) for i in test]
        df_home = df_home.append(pd.Series(test, index=df_home.columns[:len(test)]),
                                 ignore_index=True)

    else:
        df_home = df_home.append(pd.Series(test, index=df_home.columns[:len(test)]),
                                 ignore_index=True)
# df_home
```

```
# save as csv
df_visit.to_csv(data_file_path + "\mlb_box_scores_visit.csv")
df_home.to_csv(data_file_path + "\mlb_box_scores_home.csv")

# Get run differential between home and visit team
df_h = df_home.fillna(0)
df_h = df_h.replace('x', 0)
df_h = df_h.astype(int)

df_v = df_visit.fillna(0)
df_v = df_v.replace('x', 0)
df_v = df_v.astype(int)

df_v.columns = [i for i in range(1,21)]
df_h.columns = [i for i in range(1,21)]

import numpy as np
# cumulative score for each vist and home team after each inning
visit_team_cum_sum = np.cumsum(df_v, axis = 1)
home_team_cum_sum = np.cumsum(df_h, axis = 1)
visit_team_cum_sum.to_csv(data_file_path + "\\visit_team_cum_sum.csv")
home_team_cum_sum.to_csv(data_file_path + "\home_team_cum_sum.csv")

# load cummulative score - home and visit teams
home_cscore = pd.read_csv(data_file_path + "\home_team_cum_sum.csv",
names= [str(i) for i in range(1,21)])
visit_cscore = pd.read_csv(data_file_path + "\\visit_team_cum_sum.csv",
names = [str(i) for i in range(1,21)])
```

```
# assign column name to home and visit team
# home_cscore.columns = [‘H’+str(i) for i in range(1,20+1)]
# visit_cscore.columns = [‘V’+str(i) for i in range(1,20+1)]

# ===============================================================================
# Compute run differential
# ===============================================================================
# home team is ahead by r or more runs after i innning
# -ve run score means home team trailed
home_run_diff = home_cscore - visit_cscore

# visit team is ahead by r or more runs after i inning
# -ve run score means visit team trailed
visit_run_diff = home_run_diff * -1 # or visit score - home score

# ===============================================================================
# Create a subset of data set that only include the variables relevant to
our work
# ===============================================================================

subset = [0, 3, 6, 9, 10, 11, 16, 17, 18, 19, 20]
mlb_data = mlb_data_par[[str(i) for i in subset]]

# find total runs scored in each game
game_len = []
for i in mlb_data[‘18’]:
    test = i.split(‘:’)
    t_mins = int(test[0]) * 60 + int(test[1])
```

```
    game_len.append(t_mins)
mlb_data['18'] = game_len

mlb_data['total_runs'] = mlb_data['9'] + mlb_data['10']

# At this point also append home_run_diff in the mlb_data dataset

# find no. of innings in each game
no_innings = []
for i in range(len(home_score)):
    if ('x' in home_score.iloc[i].values):
        innings = home_score.iloc[i].count() - 0.5
    else:
        innings = home_score.iloc[i].count()
    no_innings.append(innings)

mlb_data['no_innings'] = no_innings

# ==============================================================================
# Analysis and Results
# ==============================================================================

import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

mlb_data = pd.read_csv(data_file_path + "\mlb_data.csv",
                       parse_dates=[0],
                       dtype = { 19 : str, 20 : str})
```

```
# MLB game time and attendance through years
df1 = mlb_data[['date', 'attendance', 'time_mins']]
df2 = df1.groupby(df1.date.dt.year)['attendance', 'time_mins'].mean().round(0)
norm_df = (df2-df2.mean())/df2.std()
plt.plot(norm_df)
plt.gca().legend(('attendance','game_time'))

# this is a universal function for almost all of our analysis
# uncomment the return statement below as needed

def get_analysis(df, i, r):
    counter = []
    # total number of games where Mercy rule could be applied (Blow-outs)
    sub_df = df[((df['X'+str(i)] >= r) | (df['X'+str(i)] <= - r)) &
    (df['no_innings']>=i-0.5)].reset_index(drop=True)
    for game in range(len(sub_df)):
        # games that were not affected
        if (sub_df['X'+str(i)][game] > 0 and sub_df['X20'][game] > 0):
            affected = 0
        elif (sub_df['X'+str(i)][game] < 0 and sub_df['X20'][game] < 0):
            affected = 0
        # games that were tied at the end of ith innings
        elif (sub_df['X'+str(i)][game] == 0):
            affected = 1
        # games that would be affected by Mercy rule
        else:
            affected = 2
        counter.append(affected)
```

```
    # table 2.1
    # prop of games where a team remains ahead after i/r and wins the game
    # return round(counter.count(0) / len(sub_df), 4)

    # table 2.2
    # returns prop of games where mercy rule can be applied
    # return round(len(sub_df) / len(df), 4)
    # return counter.count(1)

    # table 2.3
    # # returns prop of games whose outcome would change given r, i
    # try:
    #     return round(counter.count(2) / len(df), 4)
    # except ZeroDivisionError:
    #     return 'NaN'

# usage of the function for each possible mercy rule
df = pd.DataFrame()
for i in range(1,9):
    for r in range(1,12):
        df.at[i,r] = get_analysis(mlb_data, i, r)
df.columns=[r for r in range(1,12)]
df.index=[i for i in range(1,9)]
df.T

# plot function
# change the title as needed
ax = df.T.plot()
plt.legend(['1', '2', '3', '4', '5', '6', '7', '8'],
```

```
            loc='center right', title='Innings')
ax.set_xlabel('Run Differential')
ax.set_ylabel('Proportion Total Games Affected')

''' Determine the proportion of the total games in a season
that would be impacted (7,7) mercy rule. '''
ir = [[7, 7]]
df1 = pd.DataFrame()
for k in ir:
    for year in range(2010,2020):
        df2 = mlb_data[pd.DatetimeIndex(mlb_data['date']).year == year]
        df1.at[str(k), year] = get_ngames_affected(df2, k[1], k[0])
df1
ax = df1.T.plot()
ax.set_xlabel('Season')
ax.set_ylabel('Proportion of Games')

# ===============================================================================
# Determine effect of mercy rule in game length
# ===============================================================================

# fit a simple regression model
x = mlb_data[['total_runs','no_innings']]
y = mlb_data['time_mins']

x = sm.add_constant(x)

model = sm.OLS(y,x).fit()
print(model.summary())
```

```
‘‘‘
How does one run and an additional inning add up to a game time?
’’’

def get_time(df, i, r):
    ‘‘‘
    This function takes a specific mlb_data_set,
    no. of innings i and run differential r
    to calculate the average time (mins)
    due to a run and additional inning
    ’’’
    sub_df = df[((df[‘X’+str(i)] >= r) | (df[‘X’+str(i)] <= - r))&
    (df[‘no_innings’]>=i-0.5)].reset_index(drop=True)
    visit_count = sub_df.loc[:, sub_df.columns.str.startswith(‘V’)].iloc[:,i:20].
    astype(float).sum(axis=1)
    home_count = sub_df.loc[:, sub_df.columns.str.startswith(‘H’)].iloc[:,i:20].
    apply(lambda x: pd.to_numeric(x,errors=‘coerce’)).sum(axis=1)
    total_runs = visit_count + home_count
    ninngs = sub_df[‘no_innings’] - i
    sub_df[‘avg_time’] = 2.6699 * total_runs + 20.5956 * ninngs #- 24.785

    return sub_df.reset_index(drop=True)

sevenrule = get_time(mlb_data, 7, 7)[‘avg_time’]
sevenrule.mean()
eightrule = get_time(mlb_data, 8, 7)[‘avg_time’]
eightrule.mean()
```

```
# ================================================================================
# Determine if blowout games tend to be longer than non-blowout games
# ================================================================================
def get_ngames_affected2(df, i, r):
    # total number of games where Mercy rule could be applied (Blow-outs)
    sub_df = df[((df['X'+str(i)] >= r) | (df['X'+str(i)] <= - r)) &
    (df['no_innings']>=i-0.5)].reset_index(drop=True)
    blow_out_time = round(sub_df['time_mins'].mean(), 2)
    # Non-blowouts
    sub_df1 = df[~(((df['X'+str(i)] >= r) | (df['X'+str(i)] <= - r)) &
    (df['no_innings']>=i-0.5))].reset_index(drop=True)
    nblow_out_time = round(sub_df1['time_mins'].mean(), 2)

    return (blow_out_time-nblow_out_time)

ir = [[7, 7]]
df1 = pd.DataFrame()
for k in ir:
    for year in range(2010,2020):
        df2 = mlb_data[pd.DatetimeIndex(mlb_data['date']).year == year]
        df1.at[str(k), year] = (get_ngames_affected2(df2, k[1], k[0]))
df1.T.plot()

# ================================================================================
# Determine if blowout games tend to be longer than non-blowout games
# ================================================================================
def get_ngames_affected2(df, i, r):
    # total number of games where Mercy rule could be applied (Blow-outs)
    sub_df = df[((df['X'+str(i)] >= r) | (df['X'+str(i)] <= - r)) &
```

```
    (df['no_innings']>=i-0.5)].reset_index(drop=True)
    blow_out_time = round(sub_df['time_mins'].mean(), 2)
    # Non-blowouts
    # sub_df1 = df[(df['X'+str(i)] < r) & (df['X'+str(i)] > - r) &
    (df['no_innings']>=i-0.5) & (df['X'+str(i)] != 0)].reset_index(drop=True)
    sub_df1 = df[~(((df['X'+str(i)] >= r) | (df['X'+str(i)] <= - r)) &
    (df['no_innings']>=i-0.5))].reset_index(drop=True)
    nblow_out_time = round(sub_df1['time_mins'].mean(), 2)

    return (blow_out_time-nblow_out_time)

ir = [[7, 7]]
df1 = pd.DataFrame()
for k in ir:
    for year in range(2010,2020):
        df2 = mlb_data[pd.DatetimeIndex(mlb_data['date']).year == year]
        df1.at[str(k), year] = (get_ngames_affected2(df2, k[1], k[0]))
df1.T.plot()
```