The Reliance on Statistics in Baseball

by

Cody Hennequin

Submitted to the Board of Mathematics and Computer Science

School of SUNY Purchase College

in partial fulfillment of the requirements

for the degree of Bachelor of Arts

Purchase College

State University of New York

May 2021

Sponsor: Professor Shablinsky

Second Reader: Professor Zimmerman

Table of Contents

Abstract

Ever wonder how managers in Major League Baseball put together their lineups? In this project I have investigated one aspect of mathematics in the game of baseball, and that is the use of statistics as applied to the management of modern baseball teams. Statistics are able to show the performance of a player as compared to other players in the league, and their own previous seasons. Batting statistics, or offensive stats, determine how a player scores runs against opposing teams. Fielding statistics, or defensive stats, describe how good a player is at fielding the baseball and making outs, which prevents the other team from scoring runs. Pitching statistics, determine which players are good at actually throwing the ball to opposing team players. There are specific coaches that oversee these specific categories of the game, as they all help the manager to understand, interpret, and apply those statistics to make a winning lineup. These coaches are able to see that certain batters hit better against certain pitchers, and certain pitchers throw better to certain types of hitters. Statistics tell a manager which player to start or sit out a particular game. Statistics are also able to tell why teams are winning, losing, or even if they have good players. The importance and development of statistics in baseball are explained and examined to show their applications to the game by managers.

Chapter 1

Introduction

Baseball statistics have long been a part of the game as numbers that viewers can use to determine how good a player is. The sport is starting to use these numbers to their benefit however, and managers are finding new ways to win games. How are they using numbers to help them win? By assuming that a player will play close to the statistics he consistently puts up against opponents. There are so many ways to keep track of how good a player is in baseball, with the basic offensive and defensive stats that everyone has seen on television before. Offensive statistics measure how well a player can hit the ball and score runs for their team. Opposite of this is defensive statistics, which tell how well a player or team can stop the others from scoring. In a new age of baseball, there are more advanced statistics that can analyze how a player will do in the day or at night, how a pitcher would do if every fielder never made an error, and even how many runs a fielder could keep from scoring in a game. With that said, there are also outliers to these data sets. The term "regression to the mean" plays a role in determining a player's ability as well. Regression to the mean basically says that a certain player might have better hitting statistics because they were doing really well over a stretch of games, or they were having what is called a "hot streak". The same player might also have poor hitting statistics because they were having a "cold streak". The same goes for defense and pitching. Other intangible factors that can dictate a player's ability are cold versus hot weather (which plays into where a team is located or playing its games when they are not at home), day versus night games (the ability to see the ball better or worse when the sun is up), and the size of the stadium they end up playing in. This means that a player can be better or worse than his numbers say, and it leaves a manager with a lot of decisions to make. The coaches compile and sort the data, and

they have to account for intangible factors, in order to put out the best team to win games. In this data is how a player does in certain matchups, or head-to-head battles between batter and pitcher. A head-to-head matchup is very important in baseball because it can dictate whether or not your team can score runs against certain team's pitching. One topic that is going to be covered is how these matchups can play out. It is important next to discuss the outcome of the same team across different seasons, because you could have the same record and have different outcomes in a season. This topic is centered around the 2009 New York Yankees, who won the World Series, and the 2019 New York Yankees, who lost in the American League Championship Series. Both teams had the same regular season record of 103 wins and 59 losses, which is odd considering they had different outcomes to their seasons (Birken). Regression to the mean plays a key role in this chapter as relief pitching in particular was very different between the two teams. After reviewing all of these team's statistics, it can then be discussed how to make a winning lineup through using these statistics. In conclusion, there are a great many factors that must be evaluated when a baseball manager puts out a lineup for his team. How these managers set their lineup to win games using statistics will be deeply investigated.

Chapter 2

Offensive and Defensive Statistics

In total, there are 43 offensive statistics and 16 defensive statistics in the game (MLB, "Glossary"). The manager cannot keep track of the entire roster and all of those statistics himself, and in order to help him there are specific coaches for hitting and fielding. That way the manager can just get reports from his coaches on who is excelling in what categories, and who is and is not hot during a stretch of games. Also, not all of those statistics are important in making a lineup. Five offensive, or hitting, statistics hold the main keys in determining who should be in the lineup that game. Those are On Base Percentage (OBP), Slugging Percentage (SLG), Runs Created (RC), Batting Average on Balls in Play (BABIP), and On Base Plus Slugging (OPS). Defensively there are only three statistics that really make a difference and those are Fielding Percentage (FPCT), Defensive Efficiency Ratio (DER), and Range Factor.

How often a batter reaches base per plate appearance is called On Base Percentage, or OBP (MLB, "On Base Percentage"). A plate appearance is considered any time a player steps into the batter's box to hit a ball. This means that a hit, a walk, and getting hit by a pitch all count towards a player's OBP. OBP does not include errors, fielder's choice plays, drop third strikes, or sacrifice bunts. OBP includes hits scored as "H", walks denoted by "BB", and hit-by-pitches which are marked as "HBP". The formula for calculating OBP is:

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}$$

$$(Slowinski, "OPS\ and\ OPS + ")$$

OBP is important to a manager because it tells him whether or not his player is making an impact to getting on bases. In baseball scoring runs is the most important thing, and one of the few ways to score runs is by getting on base. So, if a manager sees his player with a high On Base Percentage, he could choose to put that player somewhere in the lineup where he has the ability to score several times in a game.

The next important hitting statistic is Slugging Percentage (SLG). This is the total number of bases a player records per at-bat (MLB, "Slugging Percentage"). If a player scores four bases in an at-bat, it means he hit a home run, and this is the most total bases they could score in one at-bat. They can score three bases if they hit a triple, two if they hit a double, followed by the least number of bases, the single. SLG only deals with hits and does not include walks and hit by pitches. SLG is a weighted type of stat in which the more bases you score the better the overall number is. A home run, marked as HR, is weighted the heaviest because it is one of the hardest to score. The formula for SLG is:

$$SLG = \frac{(1B * 1) + (2B * 2) + (3B * 3) + (HR * 4)}{AB}$$

$$(MLB, "Slugging\ Percentage")$$

where "AB" is At-Bat, "1B" is single, "2B" is double, and "3B" is triple. SLG is important for a manager because it measures a hitter's power. Being able to know how a hitter hits for power allows a manager to build his lineup around guys who get on base. This means that a manager can build his lineup to where someone who has a high OBP statistic hits before a hitter with a high SLG statistic. A more in-depth example would be if player A has an OBP of .400 but SLG

of .200 and player B has an OBP of .220 but a SLG of .680, then player A would hit before player B in the lineup. This is because player B has more of a chance to score player A due to the fact that player A gets on base more frequently and player B is able to drive in more runners with his high SLG.

There is then the statistic that combines the two into one number, called On-Base Plus Slugging (OPS). OPS tells how well a hitter can hit for average and for power. (MLB, "On-Base Plus Slugging"). So, OPS tells if a player can get hits, and if those hits are also for anything better than a single. Now, it may be thought that if the manager can see SLG and OBP together as OPS, then why would he need to look at them individually. The answer is because those two stats individually narrow the scope of what parts of the game a player is good at. In this sense, a manager wants his first batter to have a high OBP, and his fourth batter to be someone who has a high SLG. The manager will then use the OPS stat to determine guys in the lineup who should hit in the two and three spots because they need to be able to hit to get on base (OBP) as well as to score runs with big hits (SLG). The formula for OPS is straightforward:

$$OPS = OBP + SLG$$

$$(MLB, "On - Base\ Plus\ Slugging\ ")$$

Due to the fact that OPS is addition of two decimals, the range for these numbers will be between .500 and 1.000. As an example, a player can have .500 OBP and .500 SLG and have an OPS of 1.000. According to FanGraphs' OPS page, a 1.000 OPS is considered excellent and .570 is considered awful (Slowinski, "OPS and OPS+"). Through the use of the OPS stat, a manager can determine who his most all-around players are and put them in spots in the lineup where they

can really make a difference in the game for the offense.

Runs Created (RC) is another important statistic for managers. RC estimates a player's offensive contribution in terms of total runs (MLB, "Runs Created"). In baseball, much like any sport, you want to score more than the other team, in this case it would be more runs. RC is a statistic that tells how well one player can do just that. The formula for RC looks like this:

$$RC = \frac{TB * (H + BB)}{(AB + BB)}$$

$$(MLB, "Runs\ Created")$$

In this formula, "TB" is Total Bases, "H" is Hits, "BB" is Walks, and "AB" is At-Bats. With this formula, a manager can find out how many runs his player can create over a stretch of his plate appearances. One example of this is if a batter goes 15 for 36 over a stretch of two weeks of games. In these 15 hits, the player racked up 29 Total Bases (8 1B, 3 2B, 1 3B, 3 HR, 0 BB). This would mean that his RC stat would look like this:

$$RC = \frac{29 * (15 + 0)}{(36 + 0)} = \frac{29 * (15)}{(36)} = \frac{435}{(36)} = 12.08$$

This player's RC stat of 12.08 means that he would create 12 runs worth of offense for his team over the span of two weeks. This is a very good number over that span because one of nine players in the lineup is scoring just about one run a game over the course of 12 games. Imagine now a lineup where every player has an RC of 12.08 in 12 games. That is nine players scoring one run a game, meaning the team would score about 9 runs per game in that span. In baseball, when a team scores 9 runs, that team should win that game. In this case, RC allows managers to

see the offensive production of his hitters, especially at a quicker glance.

Batting Average on Balls in Play (BABIP) is a very powerful offensive statistic. BABIP is a player's batting average specifically on balls hit into the field of play (MLB, "Batting Average on Balls in Play"). So, if a player hits the ball and it is able to be fielded by the other team, then that at bat goes towards their BABIP statistic. This is different from normal Batting Average, which tells how often a player is getting hits. It is different because there are nuances with the Batting Average statistic where not all at bats count towards the average, or how some plate appearances are scored differently. BABIP is more describing as a player statistic because it only tells how good a player is when they actually *hit* the ball. For example, if a player hits a Sacrifice Fly, a play in which the batter hits a fly ball out to an outfielder that scores a run, that at bat does not count towards the player's Batting Average. This at bat, however, will go towards his BABIP as 0 for 1. The normal batting average statistic is important, but it accounts for every plate appearance that the player has. With BABIP, a manager can better understand if the player is getting on base on those hits they put in play or if they are just creating an out. This eliminates outcomes that are not affected by other team's fielders and that includes Home Runs and Strikeouts (MLB, "Batting Average on Balls in Play"). The equation for BABIP will look like this:

$$BABIP = \frac{(H - HR)}{(AB - K - HR + SF)}$$

$$(MLB, "Batting\ Average\ on\ Balls\ in\ Play")$$

"H" is recorded as Hits, "HR" is recorded as Home Runs, "AB" is At-Bats, and "K" is Strikeouts. "SF" is the stat considered a Sacrifice Fly, in which the batter is attempting to hit a

fly ball out in order to score a runner from third base. The reason that HR and K are being subtracted is because they are not at-bats that can be fielded by an opposing team. A K is an out, and a HR is a hit that is not considered "in play". SF is added to the denominator in the equation because it is a ball in play that is an out but counts as the player not being able to record a hit. A quick example of this would be to look at a player's stat line through 30 At-Bats. If the player went 10 for 30, his Batting Average would be .333. This is a pretty decent number considering the league Batting Average is right around this number. Now, consider for this player in his 30 at bat span, he had 10 H, 4 HR, 6 K, and 2 SF.

$$BABIP = \frac{(10-4)}{(30-6-4+2)} = \frac{(6)}{(22)} = 0.2727$$

This player's BABIP would be .273, meaning he was only really batting .273 when he hit the ball, and that is not a very high number. In this case, one can see the difference between Batting Average and BABIP, as this player was batting with a higher average but not when he was putting the ball in play. The higher the BABIP, the better chance a hitter has of getting hits. Therefore, a manager can get a better sense of how a player is hitting when he looks at their BABIP, and make a decision whether to play them based on that stat.

A manager also has to look into his player's defensive statistics in order to keep teams from scoring the same amount of runs as his team. Fielding Percent (FPCT) is a good statistic to look at to start. At its simplest, it is how often a fielder makes an out. FPCT is a defensive statistic used to tell whether or not a player can make outs on balls hit at them, balls they throw, or balls that are thrown at them (MLB, "Fielding Percentage"). The formula for FPCT is this:

$$FPCT = \frac{PO + A}{PO + A + E}$$

$$(MLB, "Fielding\ Percentage")$$

This formula shows that it is calculated using just three stats. “PO” stands for Put-Outs, which are outs recorded by a fielder by stepping on a base or tagging a runner out (MLB, “Putout”). “A” stands for Assists, which are outs recorded by a player if they touch the ball before a Put-Out (MLB, “Assist”). This means that any time a fielder fields a batted ball and throws it to another player for an out, he is awarded an Assist. “E” in this formula means Errors, and Errors are when a fielder does not field the ball cleanly and allows the hitter to get on base because of the miscue. Errors also are given on errand throws as well, meaning if a fielder throws the ball away from an out, then he is given the Error. The formula is set up to where it takes a fielder’s Put-Outs and Assists, adds them to get a sum, and then divides them by a fielder’s total chances. The way to find out how many total chances he has is by adding his Put-Outs, Assists, and Errors. FPCT is a holistic look at fielding and gives the manager a quick glance at whether or not he can put a player in the field. A player with a 1.000 FPCT is someone who has not made an error, and someone with a FPCT of 0.500 has made half of the plays that he has the opportunity to make. This comes in handy when a manager is deciding on a Designated Hitter, or DH. The DH position is used because they are hitting in place of the pitchers, who typically hit poorly. When deciding on a DH, the manager usually chooses players who are good at hitting but may not be strong in the field. A good example of this is New York Yankee’s slugger Giancarlo Stanton versus veteran Yankee outfielder Brett Gardner. Stanton’s career batting average is .268 (“Giancarlo Stanton Stats”, Sean Forman) and Gardner’s is .259 (“Brett Gardner Stats”, Sean Forman). Stanton is better at hitting for average than Gardner is, but the problem then becomes

the difference between their FPCT. Stanton's FPCT as an outfielder over the course of his career is .980 ("Giancarlo Stanton Stats", Sean Forman), while Gardner's FPCT is .992 ("Brett Gardner Stats", Sean Forman). With Stanton being better at the plate and Gardner better in the field, the Yankee manager will want to keep Stanton's bat in the lineup as the DH but play Gardner in the outfield so that he can maximize his hitting. All in all, FPCT is a good indicator for a manager to determine which players will give him the best chance at making outs.

Managers can also use Defensive Efficiency Ratio (DER) to determine how well his team is fielding, as a whole. DER is how often a defense is able to make an out on a ball in play (MLB, "Defensive Efficiency Ratio"). Defensive Efficiency Ratio is geared towards telling how the whole team does as opposed to how well any single player is doing. It could be seen as the defensive BABIP, as the stat is evaluating the defense based on how many times batters reach base on the balls put in play (MLB, "Defensive Efficiency Ratio"). The DER stat looks like:

$$DER = 1 - \frac{(H + ROE - HR)}{(PA - BB - SO - HBP - HR)}$$

$$(MLB, "Defensive\ Efficiency\ Ratio")$$

DER is subtracted from 1 because it is meant to be a percentage of a defense converting an out. "H" is hits, "ROE" is Reached on Error, "HR" is Home Run, "PA" is Plate Appearance, "BB" is Walk, "SO" is considered Strike-Out, and "HBP" is Hit by Pitch. The higher the DER, the better. DER becomes important when a manager is making defensive substitutions, or replacing some fielders for others, and needs to know how the whole team will do with players that are not usually in the starting lineup. This is because a lot of positions depend on others in order to make outs. For instance, a shortstop and second basemen have better team chemistry if they have been

playing together for a while. If the manager starts another second basemen due to an injury, he needs to know how the team would play as a whole with that new player in. The manager can get a glimpse of the whole team using the DER statistic.

Range Factor, or RF, is the last major statistic that managers can use to help make a winning lineup. It is the measure of how many plays a fielder can make, or how much range a fielder has (MLB, "Range Factor"). ESPN always has the top plays where a baseball player is running a long way for a ball in the air, or they dive five feet to their right to stop a ground ball. This is what Range Factor determines, as it can tell the amount of times they are making a play that they can get to. It will not say how far they can run for a ball or how long their dive is, but it will tell if they make the out on the ball they start running to catch. The formula looks like this:

$$RF = \frac{PO + A}{DGP}$$

$$(MLB, "Range\ Factor")$$

"PO" is this formula means Putout, "A" means Assist, and "DGP" is Defensive Games Played. DGP is important to this formula because there are times a manager will enter a player into a game but only to hit. DGP just means that the formula only takes into account the times that the player actually plays the field in that game. This statistic is important to managers because it tells them if their players are making outs on balls hit to them, no matter how difficult the play. If the RF of a player is higher, the better chance they have of making an out. A manager can couple this stat with the team's DER to determine which players to put in the field. The RF stat can also be used to tell which position a player is better at. For example, many third basemen can transition to play first base. RF allows a manager to compare his other third basemen on his

depth chart to that player, as well as to compare the player to his other first basemen to see if they are a good fit.

While these statistics are important individually, in order to create a winning lineup a manager needs to play those players with the best numbers in most of these statistical areas. For example, a player is more valuable and helpful to winning if they have a high OBP, high BABIP, high RC, and an excellent FPCT with a great DER. This becomes tough to determine over the course of a 162-game season, as players do not always play to their statistical line. This means that over the first twelve games, a player could have an OBP of .417 (he got on base 15 out of 36 times in a two-week span). Then through the next twelve games he got on base 7 out of 36 times, giving him an OBP of .194 in those two weeks. His OBP for the season would be .305 (22 times out of 72), a respectable number but not indicative of the hot streak he was on to begin the season, or the cold streak that followed. Most players experience streaks all season, and this means that a manager has to recognize when a player is hot and cold. One way to recognize this is through "regression to the mean". Regression to the mean basically says that a performance is always varied around an average true performance, and that extreme performances tend to get less extreme over the course of the next attempt ("Regression to the Mean", Farnam Street). This can be attributed to testing measurements never being exactly the same for each case ("Regression to the Mean", Farnam Street). A simple example of this is playing during the day as opposed to night. Some players have better daytime statistics than they do night and that is because the sun allows them to see the ball better. This goes the same for weather conditions, home and away games, and even who is in the lineup behind a pitcher. Altogether, those eight statistics are among the most important hitting and fielding statistics in the game because they are more accurate in determining whether or not a player can help a manager win games.

Chapter 3

Pitching Statistics

Pitching is an intricate part of baseball. It is an art to some, and it requires an extreme amount of focus and energy. There are 58 pitching statistics alone, and these range from basic averages to advanced statistics that tell the spin speed on any given pitch (MLB, "Glossary"). Again, not all of them are important to making a winning lineup. With the offensive and defensive stats, there is a little leeway in whether a manager should follow the statistical evidence on who to play because of regression to the mean. With pitching however, a manager should stick really close to what pitcher's statistics say. Pitchers do not have as many factors as the position players do in terms of regression to the mean, mainly because they are only worrying about pitching. While some pitchers hit (only in the National League, a rule that may change in the coming years), and all pitchers field, there is not enough data in those categories to count against a pitcher's pitching ability. A pitcher's throwing mechanics almost never change during the course of a season, they usually throw the same types of pitches during the length of their career, they throw from the same distance to home plate in every stadium no matter if they are home or away, and they are typically pitching off the same types of mounds (in terms of materials such as clay or dirt). In these reasons it is easy to see how pitchers pitch close to their stat lines.

All of the following stats can be applied to starters and relievers alike. Starters are pitchers who start the game on the mound for their team. They can usually pitch multiple innings in a game, as managers aim to throw them at least five or six innings at the minimum. Relievers are pitchers who come in after the starter and are usually good for a maximum of three innings in a game. A team can carry anywhere between six or seven starters, and a minimum of eight

relievers. Starters throw once every five games, whereas relievers can come in and pitch multiple games in a row. This is because starters usually throw a lot more pitches and need more time off to rest their arms before their next start. Relievers can go into a game, throw a low number of pitches, and still be able to relieve for the game the next day. An important type of reliever is called a closer. A closer is someone a manager calls in to record the last three outs of a game in order to secure a win. These closers are usually the best relievers and can do really well under pressure. Closers can be considered as "outliers" to pitching stats, and this is because they do not record the same number of innings as every other pitcher on the team. Closers are used to keep the game tied or to ensure a win, but some games a closer is not needed. Some of these situations include if that closer's team is winning, or losing, by a lot of runs, if a starting pitcher is throwing really well and ends up pitching all nine innings, or if the manager decides to save the closer for a more important game such as a series deciding win.

The most important discussion when looking at a pitcher's stats and determining whether they should be in the starting rotation or bullpen is the difference between Earned Run Average (ERA) and Fielding Independent Pitching (FIP). ERA is the number of earned runs a pitcher allows per nine innings (MLB, "Earned Run Average"). It is basically saying how many times the other team will score earned runs off of that pitcher over the course of a game. Earned runs are runs that a pitcher allows but he may not have control over whether they score or not. FIP however, focuses solely on the events a pitcher has the most control over, the events that he caused or created in order to allow the other team to score (MLB, "Fielding Independent Pitching"). FIP removes results on the balls that are hit into the field of play because the pitcher cannot control whether his fielders make errors or not (MLB, "Fielding Independent Pitching"). ERA does not account for this, and while it is important to the game, it is not as accurate of a

representation of how well a pitcher can pitch as FIP is. Essentially, FIP is a raw pitching stat to determine how good a pitcher is at his most basic form. This is helpful to a manager when trying to determine which pitchers to call up for the season, who to trade for, and who is going to give him the best chance to win because he can see how many raw runs a pitcher will give up in a game. The formula for ERA is:

$$ERA = \frac{(9 * \text{ER})}{\text{IP}}$$

$$(MLB, "Earned\ Run\ Average")$$

This formula shows "ER" which is Earned Runs being multiplied by nine and then being divided by "IP" which are Innings Pitched. The Earned Runs are multiplied by nine because the number of innings in a single game is nine, so as to make it an average across a nine-inning game. The formula for FIP is a little more complex:

$$FIP = \frac{\Big((HR * 13) + \big(3 * (BB + HBP)\big) - (2 * K)\Big)}{IP} + FIP\ Constant$$

$$(MLB, "Fielding\ Independent\ Pitching")$$

This formula has more parts to it, but it is more indicative of what is controllable for a pitcher. "HR" is Home Run, "BB" is Walk, "HBP" is Hit by Pitch, "K" is Strikeout, and "IP" is Innings Pitched. Essentially, these events, the walks, strikeouts, and hit by pitches, are all at the discretion of the pitcher and his abilities because no fielder has the ability to allow them from their position. Home Runs are the only balls in play tallied in this formula and this is because no fielder can make an out or play on a ball hit over the fence. If a player hits a Home Run, that is the responsibility of the pitcher. Strikeouts are subtracted in this formula because they are recorded outs that the pitcher is responsible for, meaning they do not count as outs fielders make, and they are not credited hits. When looking at the FIP Constant, the number is derived based on

league averages that regulate the factors across all ballparks and teams. The constant is required because Major League pitchers often have larger K totals than they do HR, BB, and HBP totals combined. In this case, the numerator is often a negative number, and the FIP Constant balances this out and regulates all pitchers to be on the same level. Both ERA and FIP can be read in a box score as being a number of runs a pitcher gives up in a game. For example, if a pitcher has an ERA of 3.19, he is said to give up about 3 runs a game. The same goes for FIP, as a pitcher could have a FIP of 3.19 and he would be giving up about 3 runs a game. The trick between these two stats is that if a pitcher has an ERA of 3.19, he could be giving up 3 runs a game that are not all necessarily his fault, meaning his FIP would be lower. If a pitcher has a FIP of 3.19, he is giving up 3 runs a game on his own, and his ERA would be higher. When a manager is looking for how many runs he can expect a pitcher to give up without a defense on the field, he will look at his pitcher's FIP, which is usually always lower than his ERA. A reason for this is because a manager could have different players in the field behind his pitcher, so in order to see how the pitcher will do without this issue, FIP is necessary to read. ERA is the generic stat to glance at when preparing for a game because that allows a manager to prepare for how many runs a pitcher will give up with his fielders behind him. This is a part of the statistics where closing pitchers are outliers. The data for a closer is collected in a snapshot, meaning if they give up one run in the one inning they pitch, their ERA will be really high. For reference, if a closer goes into a game and throws one whole inning giving up one earned run, his ERA would look like:

$$ERA = \frac{(9 * 1)}{1} = \frac{9}{1} = 9.00$$

This closer would then have to throw 8 consecutive scoreless innings in order to get his ERA back down. This would look like this:

$$ERA = \frac{(9 * 1)}{9} = \frac{9}{9} = 1.00$$

It is obviously very hard to throw 9 consecutive scoreless innings in the major leagues, especially when you are pitching in stressful situations with runners on base. In this case, it is very hard to judge a closer on his ERA due to the fact that they have a lack of innings pitched as opposed to those who throw a lot of them. While ERA will show up on television as the widely used and accepted stat to gauge pitching, FIP is a much more telling and indicative stat to judge pitchers on.

Following this, a manager must measure his pitcher's WHIP, or Walks and Hits per Inning Pitched. This is how well a pitcher keeps runners off the bases by looking at the walks and hits a pitcher gives up over the course of the innings that they pitch (MLB, "Walks and Hits Per Inning Pitched"). A pitcher's job is to not allow the other team to score runs, and a team's best chance of scoring runs is by putting players on the bases. The ways to do this are by getting walks and hits. Errors also can grant a player a base, however that is not calculated here because errors are not a pitcher's responsibility. The formula for WHIP is:

$$WHIP = \frac{BB + H}{IP}$$

$$(MLB, "Walks\ and\ Hits\ Per\ Inning\ Pitched")$$

In this formula, "BB" is walk, "H" is Hit, and "IP" is Inning Pitched. WHIP calculates whether a pitcher can avoid these walks and hits as best they can over the total number of innings they have pitched, meaning the lower the total the better. If a pitcher has a WHIP close to 1.000 it means that they allow at least one base runner for every inning they have thrown whether that be

through a hit or walk. On the other end, if a pitcher has a WHIP of 5.000 more, they are allowing 5 runners for every inning they throw. Starting pitchers usually have slightly higher WHIPs due to the fact that they have several more innings pitched, and more ability to walk batters because of their length in games. Relievers usually have lower WHIPs, but it is important to look at for a reliever because they do not typically pitch long. If a reliever has a high WHIP, he will most likely not find his way into important spots in games until he gets better because he is allowing too many baserunners and therefore does not give the team a chance to win. Due to this, a manager wants his pitchers to have WHIP stats that are lower.

When looking at important stats for pitchers, Innings per Start (I/GS) is very important for starting pitching. I/GS is the average number of innings a pitcher throws per game started (MLB, “Innings per Start”). This statistic is more important for starting pitching because it will tell the manager how many innings his pitcher can throw before he has to put in a reliever. A manager wants his starter to give him at least 6 innings in the game so that he can give the ball to his best relievers for a shorter amount of time. If the starter throws 4 innings, whether that be because of a bad day or high pitch count, the manager needs to use 5 innings worth of relievers. Considering that a manager only has so many relievers in the bullpen, he has to pick and choose when to use his best ones. This is also important because the team plays games every day with very few off, so fatigue is an issue and that keeps a manager from playing the same reliever for four days in a row. The formula for I/GS is:

$$\frac{I}{GS} = \frac{IP}{GS}$$

$$(MLB, "Innings\ per\ Start")$$

This formula is straight forward as it takes “IP”, or Innings Pitched, and divides it by “GS” which is Games Started. A manager needs his starting pitchers to have higher I/GS stats, and if he has starters with lower I/GS numbers, he needs to plan to use his bullpen wisely. While I/GS is a simple statistic and formula, its meaning and implications can dictate how a manager runs his entire pitching staff.

Another simple statistic with a simple formula and great meaning is Runs Allowed Per Nine Innings Pitched, or RA9. This stat tells how many runs a pitcher allows per nine innings that they pitch (MLB, “Runs Allowed Per Nine Innings Pitched”). The numbers for this stat are only counting the runs that cross the plate. It is not “how” a team scores runs against pitching, but just that the team is scoring runs against that pitcher. An easier way to think about it is that it is a pitcher’s Earned Run Average without the “E” (MLB, “Runs Allowed Per Nine Innings Pitched”). The formula for RA9 looks like:

$$RA9 = \frac{(9 * R)}{IP}$$

$$(MLB, "Runs\ Allowed\ Per\ Nine\ Innings\ Pitched")$$

In this case, “R” is Runs and “IP” is Innings Pitched. RA9 tells a manager how many runs he can expect to cross the plate in a game against a specific pitcher. ERA and FIP are different than RA9 because they measure an individual pitcher’s contribution to the game. Those two stats are more specific and important in explaining what is going to happen when the fielders and pitcher are both on their game. RA9 is essentially the stat that tells the manager what will happen if things go wrong. In other words, a pitcher’s RA9 stat will tell a manager what runners score no matter the circumstances. These circumstances include runs that score on errors, hit by pitches,

fielder's choice plays, and even bad or missed calls by umpires. While RA9 will not tell a manager how or why those runs scored, it is important in terms of telling him how to prepare for how many runs *could* score in the game. One thing to note with this stat as well is that a starting pitcher's RA9 or a reliever's RA9 is not as helpful as the whole pitching staff's RA9. Thinking about the game in this way, a manager can plan on looking at his staff's RA9 before making his lineup. He is looking at his team's RA9 in order to generate his lineup to score more runs than his pitchers will allow. A holistic look at the RA9 makes more sense because the team's fielding is generally the same across every game, and all relievers are technically available to pitch in every game. The manager would then go about generating his lineup in this way: look at his team's RA9, then his starting pitcher's RA9, and comparing the two, and then generating his lineup to score more than his pitchers allow. Take this example, a team's RA9 through one week of games is 4.50. Starting the next game is a starting pitcher with an RA9 of 5.62 and ERA of 3.97. A manager will compare these two and see that he can expect anywhere between 4 and 6 runs in the upcoming game because the whole staff gives up just about 5 a game, his starter gives up close to 6 per game, and his starter has an ERA that is just under 4. A manager will always create a lineup to score a lot of runs every game, but if he can expect and prepare a lineup to score 4 to 6 runs in that game then he has a really good chance at winning. It is through this that the manager is basically using RA9 to create upper and lower bounds in order to prepare how to score.

Up until now most of the discussion has been on starting pitching statistics. All of these stats are good at measuring any pitcher's performance whether that be for a starter or reliever, but the question becomes how a manager can better assess the pitchers that he can trust coming out of the bullpen. Some relievers are better under pressure and when the stakes are high,

whereas others are better when they are given a clean inning or with very low stakes. The best stat for a manager to look at for this situation is IR-A%, or Inherited Runs Allowed Percentage. The formula looks like:

$$IR - A\% = \frac{IR_S}{IR} * 100$$

$$(MLB, "Inherited\ Runs\ Allowed\ Percentage")$$

The "IR_S" number means Inherited Runners Scored, and "IR" is Inherited Runners. IR-A% tells a manager how often a pitcher allows runners to score after he comes into the game with runners on (MLB, "Inherited Runs Allowed Percentage). In simplified terms, if a starting pitcher comes out of the game in the middle of an inning with a runner or more on base, then the reliever that comes into the game is inheriting that runner, or those runners. If the reliever allows those runs to score for any reason, that is the Inherited Runners Scored stat. When a manager needs to know which pitcher to bring in in a high leverage situation, which is any situation in which the opposing team has runners on base in scoring position, he will need to look at the IR-A% stat. A low IR-A% is better in this case because it means there is less of a chance for the pitcher to give up the runs, and it is not tied into a pitcher's ERA or FIP. However, if the pitcher puts baserunners on after he allows the others to score, those do not count towards his IR-A%. In this case, IR-A% is the percentage of runners that score on a pitcher after he has come into the game with them on base. A reliever with a high percentage is more likely to be brought in in a clean inning, and a reliever with a low percentage is more likely to be brought in with the bases loaded. The closers, typically the team's best relievers, usually have the lowest IR-A% because they can come into the game in any situation and very rarely allow runners to score. This stat is good to

have in the pocket of the manager so he can generate a plan for which pitchers to use in which situations.

There are too many pitching stats to sift through and pay attention to for a manager creating a lineup and setting up a game plan for his pitching staff. ERA, FIP, WHIP, I/GS, RA9, and IR-A% are the most important and all-encompassing stats for a manager to look at in order to help him win. These six stats cover stats for starters and relievers alike and allow for the manager to see the immediate and important information to his upcoming game as well as for future games.

Chapter 4

Batter v. Pitcher Matchups

A manager has a lot of decisions to make based on a lot of statistics. After detailing the fourteen most important stats to a manager, he must look at specific batter versus pitcher matchups in order to orchestrate his lineup. Until now, the discussion has been about stats and their applications to a manager's decisions. Specific batter versus pitcher matchups plays into these stats and further help a manager make his lineup because some hitters do better against some pitchers than others. When looking at these matchups, managers need to look at and account for regression to the mean. A large part is due to the handedness of the batter and pitcher. Traditionally, left-handed batters do not hit well against left-handed pitchers, and right-handed batters do not hit well against right-handed pitching. The reason for this is because both the hitter and pitcher are playing on the same side of the plate, so it is harder for the batter to see the ball. Now, left-handed pitching and hitting is not as prominent in baseball as right-handed pitching and hitting, so right-handed batters are typically hitting against right-handed pitchers often. In this case, a manager will almost always have a right-handed batter in the lineup who will end up hitting against a right-handed pitcher at some point during the game. Due to this, a manager needs to consider the matchup. If a specific pitcher is starting and is lights out against right-handed hitters, the manager should consider sitting out most of his righty batters to start the game.

To apply this batter versus pitcher matchup to a manager and the decisions he has to make, a snapshot of some left- and right-handed batters and pitchers are assessed. The Los Angeles Dodgers, New York Mets, Atlanta Braves, and Washington Nationals are all Major League teams that play in the National League. The Mets, Braves, and Nationals all play in the

same division, and the Dodgers play all of these teams at least once a season. Top pitchers Clayton Kershaw (Dodgers) and Jacob deGrom (Mets) have been in the league for 14 and 8 years respectively. Hitters Ryan Zimmerman (Nationals) and Freddie Freeman (Braves) have been in the league for 15 and 11 years respectively. Both Zimmerman and deGrom are right-handed, and Freeman and Kershaw are left-handed. Table 4.1 holds the career stats for Kershaw and deGrom while Table 4.2 holds the career stats for Freeman and Zimmerman.

Table 4.1

Player Name	ERA	FIP	WHIP	I/GS	RA9
Clayton Kershaw	2.44	2.76	1.00	6.59	2.66
Jacob deGrom	2.55	2.70	1.04	6.41	2.82

(MLB, "Clayton Kershaw Stats") / (MLB, "Jacob deGrom Stats")

Table 4.2

Player Name	BA	OBP	SLG	OPS	RC
Freddie Freeman	0.294	0.382	0.508	0.891	1006.11
Ryan Zimmerman	0.279	0.343	0.475	0.818	1049.09

(MLB, "Freddie Freeman Stats") / (MLB, "Ryan Zimmerman Stats")

Here a manager can see that deGrom has a slightly higher ERA, yet he has a lower FIP than Kershaw. This can be due to a few reasons, one being that Kershaw has pitched for six more years than deGrom and has more numbers to show for it. Another reason this could be is because the Dodgers have been a statistically better team than the Mets, and therefore Kershaw has had an easier time pitching than deGrom. Either way, both pitchers boast very low WHIP numbers and high I/GS numbers, meaning both starters give up very little base runners per inning and length in their starts by pitching, on average, into the seventh inning. Offensively, Freeman and

Zimmerman are also closely aligned. Their batting averages are added in this table as a simple gauge for reference with the other stats. Freeman has better OBP, SLG, and OPS numbers than Zimmerman, yet Zimmerman has more RC. This could be due to the fact that Zimmerman has four more years of experience and therefore more stats to show. Between these two players though, Freeman is the better hitter based on stats alone. Now a manager must look at the more intricate stats to make his lineup. Table 4.3 shows the stats from the 2019 season for Kershaw and deGrom and depicts how they would pitch against batters of left and right handedness. Table 4.4 explains the 2019 stats for Freeman and Zimmerman against left- and right-handed pitchers.

Table 4.3

Player Name	IP v. LHB	ERA v. LHB	WHIP v. LHB	RA9 v. LHB	IP v. RHB	ERA v. RHB	WHIP v. RHB	RA9 v. RHB
Kershaw	39.1	3.66	0.997	3.913	139.0	2.85	1.058	2.978
deGrom	84.0	2.79	1.143	3.107	120.0	2.18	0.850	1.750

(MLB, “Clayton Kershaw Statcast”) / (MLB, “Jacob deGrom Statcast”)

Table 4.4

Player Name	AVG v. LHP	OBP v. LHP	SLG v. LHP	OPS v. LHP	AVG v. RHP	OBP v. RHP	SLG v. RHP	OPS v. RHP
Freeman	0.255	0.309	0.441	0.750	0.310	0.416	0.589	1.005
Zimmerman	0.367	0.415	0.551	0.966	0.213	0.285	0.361	0.646

(MLB, “Freddie Freeman Statcast”) / (MLB, “Ryan Zimmerman Statcast”)

In these tables, a manager can see a few things. In Table 4.3, a manager can see how both pitchers do not have many recorded innings against left-handed batters, and this is because lefties are rare. A manager can also see that deGrom pitches better against lefties, despite the higher WHIP. This is a rare case, as Kershaw being left-handed should mean that he is better against left-handed hitting. However, if a manager is hitting against Kershaw, he will want to put in

more lefties in his lineup for better success. When looking at these pitchers against right-handed hitting, they both have great numbers. Where Kershaw lacks against lefties, he makes up for against righties. His ERA is almost a whole run lower, as is his RA9. deGrom is also better against righties, as would make sense due to the fact that he is right-handed. The other impressive stat for deGrom is that his RA9 and WHIP are extremely low against righties. deGrom's RA9 suggest that he may allow at most 2 runs a game, and his WHIP says he would allow less than a base runner an inning. In Table 4.4, manager can see that Freddie Freeman is not as good at hitting against left-handed pitching as Zimmerman is. Zimmerman's OBP is at 0.415, which means that in almost half of his at bats he is getting on base against left handers. Zimmerman's SLG is also high, which means he can also hit for power against lefties as opposed to Freeman's. When comparing the hitters to right-handed pitching, Freeman is a different hitter. While his batting average against righties does not show his abilities, it proves that a manager needs to look deeper than the basic batting average. The key stat to look at here is Freeman's OPS, because it shows he can hit for power and average. When a hitter has an OPS above 1.000 it means he is very good. When a manager is trying to choose between these two pitchers, deGrom is a better starting pitcher than Kershaw and the manager can expect to start deGrom the first game and then Kershaw the second, no matter who is in the lineup. When deciding between hitters, the manager can expect to start Freeman when a righty pitcher is starting, and Zimmerman against lefty pitching. What happens when these hitters face off against these specific pitchers?

With these four players playing against each other often, they have a history and stats. Tables 4.5 and 4.6 show how these players would then do against each other. In Table 4.5 is

Freeman and Zimmerman and their stats hitting against Kershaw. In Table 4.6 is Freeman and Zimmerman's hitting stats against deGrom.

Table 4.5

Player Matchup	BA	OBP	SLG	OPS	RC	BABIP
Freeman v. deGrom	0.259	0.348	0.448	0.797	9.06	0.308
Zimmerman v. deGrom	0.185	0.167	0.407	0.574	2.04	0.176

(MLB, "Jacob deGrom Stats")

Table 4.6

Player Matchup	BA	OBP	SLG	OPS	RC	BABIP
Freeman v. Kershaw	0.273	0.360	0.455	0.814	3.60	0.417
Zimmerman v. Kershaw	0.323	0.333	0.484	0.817	5.16	0.333

(MLB, "Clayton Kershaw Stats")

The manager already established that Freeman is a better hitter against righties and Zimmerman better against lefties. These tables then establish these hitter's specific stats against the pitchers that are tabbed as starters for games one and two. deGrom was the best starter, and it shows in Table 4.5 as well. Both stat lines for the hitters are low due to the fact that deGrom is so good, however when deciding who to start with respect to hitting against deGrom, Freeman is the guy. Over the course of his career, Freeman has accumulated around 9 RC against, as opposed to Zimmerman's 2. Also important for a manager to note is that Freeman's BABIP is much higher than Zimmerman's. This shows that Freeman has a better chance to get hits than Zimmerman against deGrom, which means that Freeman is the better choice for the lineup. When looking at these hitters against Kershaw (Table 4.6), the story is a little different. Freeman does not typically hit well against lefties, however his specific stat line against Kershaw is

relatively good. Both Zimmerman's and Freeman's stat lines are close, and they all favor Zimmerman when choosing which batter to start in the lineup. However, if a manager needs to decide on whether he needs more chances at hits, then he would start Freeman. This is because the two hitter's stat lines are so close that if a manager starts Freeman over Zimmerman, he is not losing as much in his hitter. In other words, if Zimmerman's SLG was 0.300 points higher, his OPS was 0.200 points higher, or he had 4 more RC than Freeman, then Zimmerman would be the starter hands down. Since Freeman is only batting 0.029 SLG lower, 0.003 OPS lower, and 1.56 RC less, the manager is not losing much if he starts him. Also, Freeman has a better OBP stat and BABIP stat, meaning the manager must note who is hot and who is not.

This is where regression to the mean comes into play. When looking at regression to the mean for hitting, it is safe to look at streaks. A hot streak would mean every ball a hitter hits seems to get them on base whether it is falling in for a single, or down the line for extra bases. When a hitter is cold, he seems to never be able to get on base or is always hitting the ball at someone and making outs. A hitter is on a hot streak if his stat line looks like 15 for 25 with 8 extra base hits and 12 RBIs, or Runs Batted In. A hitter is on a cold streak if his stat line is something along the lines of 1 for 25 with no extra base hits and 10 strikeouts. If a few of the manager's batters are on hot streaks, they should stay in the lineup no matter what against any pitcher the other teams throw. When player's stat lines are as close as Zimmerman's and Freeman's, the streak and regression to the mean needs to be considered. If Freeman's stats are worse against left-handed pitching overall, but Kershaw is starting, and his stats are not as bad then the manager would choose Zimmerman due to the overall fact that he is consistently better against lefties. However, if Freeman is close in stats and he is having a hot streak, Freeman would make the start. The same goes for cold streaks, as if Freeman is on a cold streak, he would

not be making the start. Streaks for hitters come and go, so a manager cannot count on Freeman being able to consistently start over Zimmerman every game. Streaks also apply to pitching, however not as often as it does for hitting. If Kershaw is on a pitching hot streak, the manager would not want to start Freeman.

Overall, when a manager is making his decision for the lineup card, he is going to want to start Freeman against deGrom and Zimmerman against Kershaw. Against other lefties, Zimmerman will also get the start and Freeman will get the start against righties. If a manager has deGrom and Kershaw as his pitchers, he will want to name deGrom as his game one starter, and Kershaw as his game two starter.

Chapter 5

2009 New York Yankees v. 2019 Yankees

To this point, it has been easy to see how a manager can setup his lineup using statistics. Baseball clearly has a heavy reliance on players and their performance where it creates numbers in order to measure and gauge how well they are doing. While stats can tell a manager how well a player, or team, is playing over the last couple of games or seasons, they can also help explain the difference between why teams are winning. In 2009, the New York Yankees won the World Series against the Philadelphia Phillies. Since then, the Yankees have not been to the Fall Classic, having made the playoffs eight times and missing them completely for three years (Birken). They have come close, reaching the American League Championship Series (if a team wins this series, they go to the World Series) in four of the years and having two 100-win seasons in 2018 and 2019 (Birken). When a team wins 100 games in a season, it means they are really good in that particular season since it is so hard to do. In their only 100-win season where they made the Championship Series, 2019, the Yankees had one of the best teams they have had in ten years. The 2009 Yankees and 2019 Yankees have a lot in common, so much in fact that they had the exact same regular season record at 103 wins and 59 losses (Birken). Some may say this is not so much of a coincidence, but the rest of the teams' statistics say otherwise. Below are table comparisons of the two Yankee teams in hitting, pitching, and defense. Table 5.1 compares the offensive stats of the two teams while Table 5.2 compares the pitching. Table 5.3 compares the fielding between the teams.

Table 5.1

Team Offense	AVG	OBP	SLG	OPS	TB	BABIP	RC
2009 NYY	0.283	0.362	0.478	0.840	2703	0.306	969.11
2019 NYY	0.267	0.339	0.490	0.829	2735	0.306	916.71

("2009 New York Yankees Statistics") / ("2019 New York Yankees Statistics")

Table 5.2

Team Pitching	ERA	FIP	WHIP	I/GS	RA9	IR-A%
2009 NYY	4.26	4.32	1.352	6.76	4.674	27.13%
2019 NYY	4.31	4.47	1.304	6.12	4.609	28.05%

("2009 New York Yankees Statistics") / ("2019 New York Yankees Statistics")

Table 5.3

Team Defense	FPCT	DER
2009 NYY	0.985	0.690
2019 NYY	0.982	0.677

("2009 New York Yankees Statistics") / ("2019 New York Yankees Statistics")

In Table 5.1, both teams are separated by decimals in each category. Their batting averages are separated by 0.016, their OPS by 0.011, their RC by 32, and their BABIP stats are exactly the same. In any 162-game span, a team's stats across seasons are usually never this close, and if they are its in one or two categories. These two teams and their offensive stats are close in *every* category. The similarities do not stop there. When looking at their pitching (Table 5.2) and defensive (Table 5.3) stats they are separated by small margins as well. Their ERA stats are separated by 0.05, FIP by 0.15, WHIP by 0.048, and I/GS by 0.64 innings. The team's RA9

stats have a difference of 0.065 runs, FPCT by 0.003%, and DER of 0.013. The only blemish on these records is the 0.92% in the IR-A%, and even this stat is a close contest. The question then becomes what all of these stats really have to say about the difference between these two teams.

The first way to look at this is which team is "better" by those small decimals. Both team's stats for 162 games are really good, regardless of who is better. In this case, the better team statistically between the years are the 2009 Yankees. They have the better stats in ten of the fifteen categories above (AVG, OBP, OPS, RC, ERA, FIP, I/GS, IR-A%, FPCT, and DER). Some may say that makes sense because they won the World Series, so of course they have better stats. The argument can actually be made that they won their World Series *because* of their stats. Here are some reasons why:

Table 5.4

RC	The 2009 RC stat is higher which means more runs created, which leads to more chances to score.
DER	The 2009 DER stat is higher which means more defensive cohesion, less errors, and better chance at keeping unnecessary runners from reaching base due to those errors.
I/GS	The 2009 I/GS stat is higher which points to the starters going later into games. When the starter goes later into games that means the relievers get to rest more often, and when they get more rest, they come out fresher and

	crisper for their next appearance. This then leads to less late inning runs being allowed and better overall appearances, which is reflected in the IR-A% stat.
IR-A%	The 2009 IR-A% stat is lower which means that less inherited runners are scoring. This can be attributed to the starters going later into games and giving the relievers a chance to rest more often.

Now the team should take into account this as well, that all the stats are close, but the difference lies in those nitty and gritty decimals. If a team is winning games by 1or 2 runs every night, the games become more stressful over time. It creates more wear and tear on those players, leading to more injuries and worse performance. If a team is winning games by more runs with less effort in terms of less errors, extraneous runs, and generally having an easier time hitting, then it is easier to play loose from game to game and win more often. The 2009 Yankees did just that and it shows through their stats edging out their 2019 counterparts. It truly expresses that the difference between a World Series appearance (and win) and playoff appearance is a matter of a few decimals.

Another way to view the difference between these teams, aside from the numbers, is the data that does not show up in a box score. Previously touched on was that of outliers and regression to the mean. A huge part of the 2009 Yankees success came from their closer, and Hall of Fame reliever Mariano Rivera. The top closer for the 2019 Yankees was left-handed

pitcher Aroldis Chapman, who can throw fastballs consistently at 99-102 miles per hour, some of the fastest and hardest in the game. Why these guys are outliers to the data, and why they can be part of the reason that the 2009 Yankees and 2019 Yankees were different is because Rivera is one of the best closers to ever pitch in the Major Leagues. In Table 5.5, the seasons of Rivera and Chapman are compared.

Table 5.5

Player Name	ERA	SV	IP	SO	WHIP	IR-A%
Rivera (2009)	1.76	44 (in 46 opportunities)	66.1	72	0.90	25.00%
Chapman (2019)	2.21	37 (in 42 opportunities	57.0	85	1.11	18.18%

(MLB, “Mariano Rivera Stats”) / (MLB, “Aroldis Chapman Stats”)

There are a few things that should catch a manager’s eye in these dueling stat lines. The first is that Rivera had 46 save opportunities and made 44 saves (SV). A save is when the closer comes in and "saves” the game from being lost or ensures the win for his team in the last inning if there are tying and/or winning runs on base. A blown save is when the closer cannot complete the inning without giving up the tying or winning runs. In this case, Chapman had 5 blown saves, which means the 2019 Yankees lost 5 more games whereas the 2009 Yankees lost only 2. Other key differences between these two arms are the ERA, IP, and WHIP of Rivera. Rivera’s sub 1 WHIP is a key difference because it shows that he allowed less than 1 runner per inning pitched. This becomes important in postseasons because every runner means more than anything in the postseason, and when a closer is barely allowing one per inning he pitched, that is excellent. Chapman does lead in strikeouts and IR-A% however and can be attributed to much of the team’s success in 2019 due to his ability to leave runners stranded on base.

The last way to view the two teams and their differences has absolutely nothing to do with either team's stats actually. While the main discussion and differences between the two teams is the statistical decimals that separate the two, it is worth mentioning that the Houston Astros, the team of which the 2019 Yankees lost to in the Championship Series, was caught cheating. It was discovered that the Astros knew which pitches were coming and were notified of them during their at-bats, which means they had better chances at hitting. A key moment in the series came when Astros second baseman Jose Altuve hit a game winning homerun off of closer Aroldis Chapman. Of course, had the Astros not cheated who would know who would have actually won the series, but with the way the 2009 and 2019 Yankees were so similar statistically it points in favor of the 2019 Yankees in being able to pull off a World Series appearance and maybe a win. In conclusion, the difference between a World Series win and playoff appearance just comes down to a small margin of decimals in key stats, a couple of powerhouse closers with the ability to change the outcome of a game late, and losing to a team that cheats in order to help them win.

Chapter 6

Making Your Own Lineup

Now that it is understood that statistics are a huge part of a manager's decision for past seasons and present ones, it is easier to see how many decisions he has to make based on numbers. In this chapter, how to put together a lineup using all of this information is going to be detailed. It will not account for streaks or regression to the mean and will only consist of looking at the stat lines of players on a manager's roster. The first step is deciding between two players competing for the same position. A manager could have several position battles to decide between, meaning he could have two solid players for every position on the field. For this example, though, the stat lines for just one position battle are detailed. Table 6.1 is going to show two shortstops on the team and their stats.

Table 6.1

Player Name	BA	OBP	SLG	OPS	RC	BABIP	FPCT	DER	RF
A	0.333	0.380	0.460	0.840	6.78	0.392	0.977	0.705	4.781
B	0.322	0.368	0.442	0.810	4.99	0.380	0.984	0.727	4.802

The manager has to decide who gets the starting spot at shortstop, and whoever does not play in the field will start at the DH position meaning he only hits for that game. Through these two stat lines in Table 6.1, a manager can see that Player A is the better offensive player. His BA, OBP, SLG, OPS, RC, and BABIP stats are all higher than Player B's. Where Player A is good at hitting however, is where Player B is better at fielding. Player B has higher FPCT, DER, and RF meaning he has more range, is better in efficiency, and makes more plays in the field. This manager does not want Player A fielding over Player B, but he cannot lose Player A's bat in

the lineup. Good for the manager, Player A can stay in the lineup as his DH. The manager must establish that Player B has to be the starting shortstop due to his better fielding stats. In some cases for a manager, he must choose the guys that will perform in all aspects instead of just hitting. For instance, if a manager has to make his position battles for every position, he might consider picking a few of his better hitters over his fielders in some positions because he wants to ensure his better hitters give him chances to score. Essentially, as long as his hitter can outplay his fielding, he may find himself in the lineup. In most cases however, the manager is going to choose to start the better fielder over the better hitter. With these two players, Player A carries the hitting stats to be a great 3 hitter, and this is because his OPS, RC, and BABIP are very high. A manager wants his 3 hitters to hit for average and power, create a good number of runs, and get on base more often when he puts the ball in play. Player B still has respectable hitting statistics, and his OBP and BABIP are high enough to be considered the manager's 9 hitter. A 9 hitter is someone who can get on base but may not be as strong of a hitter and is someone who can get the lineup turned over so that the top of the order guys can hit again. In this first step, the manager has penciled in Player B as the starting shortstop batting 9th, and Player A as the starting DH batting 3rd.

The next step, once all position battles are decided, is to organize and generate the lineup to score runs. This has to come after setting his defenders because a manager cannot play all of his players on the field at the same time. He is only allowed ten players at a time, eight fielders, one pitcher, and one DH. The pitcher does not hit in the lineup in this example, so the manager will have nine hitters making up his lineup. Consider Table 6.2, where nine players and their stat lines are displayed. Player's A and B are already decided as third and ninth respectively.

Table 6.2

Player Name	BA	OBP	SLG	OPS	RC	BABIP
A	0.333	0.380	0.460	0.840	6.78	0.392
B	0.322	0.368	0.442	0.810	4.99	0.380
C	0.310	0.440	0.420	0.860	5.05	0.350
D	0.375	0.315	0.505	0.820	6.90	0.319
E	0.325	0.395	0.435	0.830	5.56	0.344
F	0.324	0.375	0.425	0.800	5.18	0.347
G	0.327	0.370	0.445	0.815	5.25	0.330
H	0.372	0.312	0.495	0.807	5.95	0.322
I	0.360	0.327	0.485	0.812	5.70	0.327

Looking at these stat lines a manager can organize his lineup like this:

1	C
2	E
3	A
4	D
5	H
6	I
7	G
8	F
9	B

Breaking it down in Table 6.3:

Table 6.3

1. Player C	Leads off due to his high OBP, and good BABIP. A leadoff hitter needs to get on base often and hit balls into play and have them fall for hits. Player C has this ability and would be leadoff in this case.
2. Player E	Player E has solid all around stats with a slightly higher OBP to follow leadoff hitter. Often times in baseball the second hitter is "second leadoff", so a manager wants someone with similar stats to his leadoff guy but may not be as strong as him. The second hitter also carries an average RC stat, so as to be able to score a runner in the chance that the leadoff guy gets on base.
3. Player A	Already established as the three hitter from above. Player A has a good BA, great OBP, high SLG, and good OPS which means he is a solid hitter for average and power. The three hitter is usually one of the best hitters in the lineup, so it makes sense to put Player A in

	this spot as well because he has some of the best hitting stats on the team.
4. Player D	The highest SLG usually belongs to the fourth hitter in the lineup. High SLG means great power hitting to drive in runs, and a high RC stat means they create a lot of runs. Player D has high stats in these categories and that is why he would be the manager's decision for fourth in the lineup.
5. Player H	The fifth batter is to the fourth batter like the second batter is to the first. If the four hitter cannot slug in a run with an extra base hit, the five hitter has to do just that. The five hitter usually has a lower SLG than the 4 hitter but higher than rest of team. Player H in this case is that second four hitter, with a slightly lower SLG and a high RC stat.
6. Player I	Player I is a solid all-around player with a higher SLG and RC to pick up runs left by the 4th and 5th hitters. He also has a higher-than-average BABIP, so he comes through as the manager's sixth hitter in the lineup.

7. Player G	Player G has a good BA and good OBP to continue an inning. The 7th, 8th, and 9th hitters in the lineup are not the best hitters on the team, but they are not terrible either. The seven hitter is usually a decent hitter. His stats do not propel him to the middle or top of the order, but they grant him the ability to spark the bottom of the lineup. When thinking technically about the game, if the first 6 hitters get out the game would be in the 3rd inning. This would mean that the seven hitter is leading off the inning and would need to provide the team with a little push to get everything going. Player G has the stats to do just that.
8. Player F	The eighth spot in the order does not have any specific criteria for a manager. Player F has a good BA and BABIP for bottom of the order to get on base. The more players on base in the bottom of the order means the better chance for the lineup to turnover. The more times the top of the order comes around to hit, the better chance a team has of winning

	because they are the best hitters on the team. In this case Player F's ability to get hits is best measured by his decent BABIP and therefore nets him 8th.
9. Player B	Player B was already placed in ninth from before. His solid BA and OBP, with higher BABIP means he is good at getting on base and turning the lineup over for the top of the order guys.

Now that the manager has his starting lineup and defense established, he can choose his pitching. This step can come first or last but is usually set the first week of the season because of starting rotations. Starting rotations are the five starting pitchers on the pitching staff that rotate every game. So, the first starting pitcher starts the first game, the second starting pitcher the second game, and so forth until they come back around to the first starter again. Managers usually set this at the beginning of the season so starters have an idea of when they are throwing and can be adjusted at the middle of the season where the players have multiple days off during the "All-Star Break". This All-Star Break allows managers to reset their rotation if they need to because it gives all five starters a chance to rest before starting back up in the second half of the regular season. Relieving pitching usually gets the most attention from a manager though due to the fact that some guys are better in certain situations than others and that can change from week to week or even day to day. In Table 6.4, there are two stat lines for the manager's two closing pitchers.

Table 6.4

Player Name	ERA	FIP	WHIP	RA9	IR-A%
X	2.40	2.32	0.92	2.55	22.40
Z	2.22	2.00	1.02	2.44	29.60

(MLB, "Aroldis Chapman Stats") / (MLB, "Kenley Jansen Stats")

Since his defense, lineup, and rotation are set, he just needs to pick between who is going to close the game if and when it gets to that point. Looking at their stat lines, Player X and Player Z are very good relievers. Player X has higher ERA, FIP, and RA9 stats, but also has incredible WHIP and IR-A%. Player Z has great stats in the area that Player X lacks, however the stat that hurts Z from being the closer is his IR-A%. At 7.20% higher, his IR-A% is what is going to make the manager decide to throw Player X as the closer. In baseball there is a position called the "Set-up". This position is a reliever who bridges the gap between a starter or other reliever, and the closer. They usually get a clean inning, and it is usually the eighth inning of the game. This is actually the position that Player Z would fall into due to the fact that his stats are good in the other categories. If Player Z ends up leaving runners on base, Player X can come in and save the game. In the end of this process, the lineup card for the manager for this game would look like this:

Table 6.5

Batting Order	Pitchers
1. C	
2. E	
3. A	

4. D	**Set up:**
5. H	Z
6. I	
7. G	**Closer:**
8. F	X
9. B	

A few things to understand about this small example. First is that it does not account for hot streaks or cold streaks. Any one of these players could be experiencing a bad or good stretch of games. The second is that the example does not account for handedness. The batters in this example are just being examined for their raw stat lines. When going in depth for a manager however, the player's stats against left and right handers are presented to them so they can better prepare their lineup. This example is just about the basic numbers a manager has to look at in order to put together his lineup, and as it can be seen there is a lot to account for.

Chapter 7

Conclusion

A manager has a lot to account for when trying to make a winning lineup. His lineup does not include just the hitters. When creating his lineup, he has to account for his fielders and even pitchers. There are coaches that oversee each stat category. For offense there are only five stats a coach should home in on to evaluate his players. These five offensive stats are the most important because they detail which hitters will get on base, get hits, get power hits, and create runs. The three defensive stats that a manager needs help make a winning lineup because they show which players can play their positions well along with being able to tell how the whole team does on the field as a whole. When evaluating players who hit and play the field, the manager needs to look at all eight stats. As for pitching, coaches and managers need to pay attention to six main stats. These stats look at starters and relievers alike and are important when setting a pitching rotation. The six main pitching stats will determine which pitchers should start first and even when a reliever should come into the game. When the coaches and manager get together, they discuss these stats holistically. In this holistic review, they look at specific batter versus pitcher matchups. The matchups and the stats that come with them further decide how the lineup should be shaped. Managers should consider that players may have hot or cold streaks they are going through and would need to judge his player's performance based on this as well. Stats in baseball are important to the present but can indicate the future by looking at the past. They can tell how teams did by looking at past seasons and compare them year to year to indicate the team's direction of winning. A manager can find that decimals can make the difference between a World Series and missing the playoffs. Baseball statistics are important to the game because they decide the good players from the bad. They determine which players

should get paid more than others. If a manager can put together a winning lineup, make the playoffs, and win the World Series, then he can get a raise on his next contract. The best way a manager can do that is by following his team's statistics, and by judging and evaluating his players based on the main stats presented. With all of this information, anyone could go out and put together a winning lineup for high school baseball and beyond!

Bibliography

Birken, Dan. "New York Yankees Playoff History: 1903 - 2020." *New York Yankees Playoff History | 1903 - 2020*, Champs or Chumps, champsorchumps.us/team/mlb/new-york-yankees. Accessed 17 May 2021.

"Brett Gardner Stats." Edited by Sean Forman, *Baseball*, 2021, www.baseball-reference.com/players/g/gardnbr01.shtml. Accessed 17 May 2021.

"Giancarlo Stanton Stats." Edited by Sean Forman, *Baseball*, Sports Reference, 2021, www.baseball-reference.com/players/s/stantmi03.shtml. Accessed 17 May 2021.

MLB. "Aroldis Chapman Stats, Fantasy & News." *MLB.com*, MLB, www.mlb.com/player/aroldis-chapman-547973. Accessed 17 May 2021.

MLB. "Assist: Glossary." *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/assist. Accessed 18 May 2021.

MLB. "Batting Average on Balls in Play (BABIP): Glossary." *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/babip. Accessed 18 May 2021.

MLB. "Clayton Kershaw Statcast, Visuals & Advanced Metrics: MLB.com." *Baseballsavant.com*, MLB Advanced Media, baseballsavant.mlb.com/savant-player/clayton-kershaw-477132?stats=splits-r-pitching-mlb&season=2019. Accessed 17 May 2021.

MLB. “Clayton Kershaw Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/clayton-kershaw-477132. Accessed 17 May 2021.

MLB. “Defensive Efficiency Ratio (DER): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/defensive-efficiency-ratio. Accessed 18 May 2021.

MLB. “Earned Run Average (ERA): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/earned-run-average. Accessed 18 May 2021.

MLB. “Fielding Independent Pitching (FIP): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/fielding-independent-pitching. Accessed 18 May 2021.

MLB. “Fielding Percentage (FPCT): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/fielding-percentage. Accessed 18 May 2021.

MLB. “Freddie Freeman Statcast, Visuals & Advanced Metrics: MLB.com.” *Baseballsavant.com*, MLB Advanced Media, baseballsavant.mlb.com/savant-player/freddie-freeman-518692?stats=splits-r-hitting-mlb&season=2019. Accessed 17 May 2021.

MLB. “Freddie Freeman Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/freddie-freeman-518692?stats=career-r-hitting-mlb&year=2021. Accessed 17 May 2021.

MLB. “Glossary.” *Major League Baseball*, MLB, 2021, m.mlb.com/glossary. Accessed 17 May 2021.

MLB. “Inherited Runs Allowed Percentage (IR-A%): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/inherited-runs-allowed-percentage. Accessed 18 May 2021.

MLB. “Innings Per Start (I/GS): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/innings-per-start. Accessed 18 May 2021.

MLB. “Jacob DeGrom Statcast, Visuals & Advanced Metrics: MLB.com.” *Baseballsavant.com*, MLB Advanced Media, baseballsavant.mlb.com/savant-player/jacob-degrom-594798?stats=statcast-r-pitching-mlb. Accessed 17 May 2021.

MLB. “Jacob DeGrom Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/jacob-degrom-594798. Accessed 17 May 2021.

MLB. “Kenley Jansen Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/kenley-jansen-445276. Accessed 17 May 2021.

MLB. “Mariano Rivera Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/mariano-rivera-121250. Accessed 17 May 2021.

MLB. “On-Base Percentage (OBP): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/on-base-percentage. Accessed 18 May 2021.

MLB. “On-Base Plus Slugging (OPS): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/on-base-plus-slugging. Accessed 18 May 2021.

MLB. “Putout (PO): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/putout. Accessed 18 May 2021.

MLB. “Range Factor (RF): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/range-factor. Accessed 18 May 2021.

MLB. “Runs Allowed Per Nine Innings Pitched (RA9): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/runs-allowed-per-nine-innings-pitched. Accessed 18 May 2021.

MLB. “Runs Created (RC): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/advanced-stats/runs-created. Accessed 18 May 2021.

MLB. “Ryan Zimmerman Statcast, Visuals & Advanced Metrics: MLB.com.” *Baseballsavant.com*, MLB Advanced Media, baseballsavant.mlb.com/savant-player/ryan-zimmerman-475582?stats=splits-r-hitting-mlb&season=2019. Accessed 17 May 2021.

MLB. “Ryan Zimmerman Stats, Fantasy & News.” *MLB.com*, MLB, www.mlb.com/player/ryan-zimmerman-475582?stats=career-r-hitting-mlb&year=2021. Accessed 17 May 2021.

MLB. “Slugging Percentage (SLG): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/slugging-percentage. Accessed 18 May 2021.

MLB. “Walks And Hits Per Inning Pitched (WHIP): Glossary.” *MLB.com*, MLB, www.mlb.com/glossary/standard-stats/walks-and-hits-per-inning-pitched. Accessed 18 May 2021.

“Regression Toward the Mean: An Introduction with Examples.” Edited by Farnam Street, *Farnam Street*, 1 Oct. 2020, fs.blog/2015/07/regression-to-the-mean/. Accessed 17 May 2021.

“Range Factor (RF).” Edited by Seneca Labs, *Captain Calculator*, Seneca Labs, 2014, captaincalculator.com/sports/baseball/range-factor-calculator/. Accessed 17 May 2021.

Slowinski, Piper. “FIP.” *FIP | Sabermetrics Library*, FanGraphs, 15 Feb. 2010, library.fangraphs.com/pitching/fip/. Accessed 17 May 2021.

Slowinski, Piper. “Guts!: FanGraphs Baseball.” *Guts! | FanGraphs Baseball*, FanGraphs, www.fangraphs.com/guts.aspx?type=cn. Accessed 17 May 2021.

Slowinski, Piper. “OPS and OPS+.” *OPS and OPS+ | Sabermetrics Library*, FanGraphs, 16 Feb. 2010, library.fangraphs.com/offense/ops/. Accessed 18 May 2021.

“2009 New York Yankees Statistics.” Edited by Sean Forman, *Baseball*, Sports Reference, www.baseball-reference.com/teams/NYY/2009.shtml. Accessed 17 May 2021.

"2019 New York Yankees Batting Statistics." Edited by Sean Forman, *Baseball*, Sports Reference, www.baseball-reference.com/teams/NYY/2019-batting.shtml. Accessed 17 May 2021.