

# How Important is Putting the Ball in Play in Baseball?

Marcelo Hernandez

Carthage College

[mhernandez2@carthage.edu](mailto:mhernandez2@carthage.edu)

November 17, 2017

## Abstract

Baseball is a game in which the goal is to score more runs than your opponent. However, with the 2017 Major League Baseball season breaking home run and strikeout records, does putting the ball in play make a difference? In this paper, we will introduce a new statistic, contact average, and using this, show that it does not.

## 1 Introduction

What do all of the following have in common: a grand slam, a sacrifice bunt, a groundball double play, or a ball dropped by the outfielder? They all involve a baseball batter hitting the ball in play. As a current youth baseball coach, the author has found that when the ball is put in play, it greatly increases the chance of getting on base and scoring runs due to the pressure placed on the opponent's defense. This in turn results in a greater chance of winning. However, can the same be said about Major League Baseball? This past 2017 baseball season saw a record number of both home runs and strikeouts. Does this mean that putting the ball in play is not as important if there will be a home run to cover for the guy who just struck out and failed to put the ball in play? The purpose of this research is to determine whether putting the ball in play helps determine a team's ability to win games both in season and over the course of a couple of years.

## 2 Definitions and Development

To determine how many times a baseball player puts the ball in play, it must be determined how many times he makes contact. Using some key baseball statistics, making contact will be defined. These statistics will be used to formulate a new statistic.

**Definition 1** Making **contact** means for a player to hit the ball with his bat causing the ball to become in play. This does not include foul balls that are not caught. Foul balls that are caught are included because the team on defense is able to make a play on the ball and a baserunner could advance.

**Definition 2** A **plate appearance (PA)** is a baseball statistic for the number of times a baseball batter completes his turn batting in a game.

**Definition 3A** An **at-bat (AB)** is a baseball statistic that comes when a batter reaches base via fielder's choice, hit, or an error (not including catcher's interference) or when a batter is put out on a non-sacrifice. (Whereas a plate appearance refers to each completed turn batting, regardless of the result.)

**Definition 3B** A **fielder's choice (FC)** is when an infielder attempts to make a play at a base other than first base on a ground ball.

**Definition 3C** A **hit (H)** is when a baseball player hits the ball and reaches base safely without an error or fielder's choice from the opposing team.

**Definition 3D** An **error (E)** is when the opposing team is unable to make a routine play to get a play on offense out.

**Definition 3E** A **sacrifice (SAC)** is when a player puts the ball in play through either a bunt or a fly ball that causes base runners to advance.

**Definition 4** A **base on balls (BB)** occurs when a baseball batter is awarded first base by drawing four balls from the pitcher. A base on balls may be intentional from the opposing pitcher.

**Definition 5** A **hit by pitch (HBP)** occurs when a baseball batter is awarded first base because he was hit by the pitched ball.

**Definition 6** A **strikeout (SO)** is recorded when there are three strikes against the batter and he is called out. When this occurs, it means that the batter did not put the ball in play.

Note: When a pitcher strikes out a player, the common shorthand is K. However, when it comes to offensive statistics as seen in this section, the strikeout is referred to as SO.

**Definition 7** **On base percentage (OBP)** is a statistic that measures how often a player gets on base.

**Definition 8** A **home run (HR)** is a type of hit in which the player who hit the ball is able to score without the benefit of an error from the opposing defense.

We will now define our new statistic, contact average, for this paper.

**Definition 9** **Contact average (CAVG)** is determined by taking the batter's plate appearances subtracted by base on balls, hit by pitches and strikeouts all divided by the batter's plate appearances subtracted by base on balls and hit by pitches. Written more formally,

$$CAVG = \frac{PA - BB - HBP - SO}{PA - BB - HBP}.$$

Note: It is customary to round baseball statistics to the nearest thousandth when applicable.

**Example 10** Using Babe Ruth's career statistics, we shall computer his contact average:

$$CAVG = \frac{6474 - 614 - 62 - 989}{6474 - 614 - 62} \approx 0.844.$$

**Example 11** The player with the record for the least number of strikeouts is Joe Sewell. Let's take a look at what his contact average was:

$$CAVG = \frac{8333 - 842 - 80 - 114}{8333 - 842 - 80} \approx 0.985.$$

**Example 12** The player with the record for the most number of strikeouts is Reggie Jackson. Let us determine what his contact average was:

$$CAVG = \frac{11418 - 1375 - 96 - 2597}{11418 - 1375 - 96} \approx 0.739.$$

Surprisingly, just because Reggie Jackson had the most number of strikeouts, does not mean he had the lowest contact average in the history of the game. Finding the lowest contact average in the history of full time players is an exercise left for the reader.

**Example 13** Now looking at the 2001 Seattle Mariners (the team with the best record in the last century), we shall computer the team contact average:

$$CAVG = \frac{10623 - 2062 - 43 - 1330}{10623 - 2062 - 43} \approx 0.829.$$

**Example 14** The only other team to have as many wins as the 2001 Seattle Mariners, was the 1906 Chicago Cubs. We shall computer their team contact average:

$$CAVG = \frac{5742 - 448 - 45 - 516}{5742 - 448 - 45} \approx 0.902.$$

**Example 15** To give perspective on what a high and low team contact average is, the following table shows the contact averages of all the Major League Baseball teams from the 2017 baseball season. Included are the highest, lowest, and average contact averages. For individual contact averages of past and present players, see Figure 12 in the appendix.

### 2017 Team Contact Averages

Team	CAVG	Team	CAVG	Team	CAVG
<b>Houston Astros*</b>	<b>0.809</b>	New York Mets	0.769	Chicago White Sox	0.750
Cleveland Indians	0.793	Detroit Tigers	0.765	Chicago Cubs	0.749
Kansas City Royals	0.791	Washington Nationals	0.765	Los Angeles Dodgers	0.748
Atlanta Braves	0.791	Cincinnati Reds	0.762	Philadelphia Phillies	0.747
San Francisco Giants	0.786	<b>Toronto Blue Jays**</b>	<b>0.761</b>	Arizona Diamondbacks	0.740
Boston Red Sox	0.786	<b>Minnesota Twins**</b>	<b>0.761</b>	Oakland Athletics	0.730
Los Angeles Angels	0.781	St. Louis Cardinals	0.758	Texas Rangers	0.728
Pittsburg Pirates	0.781	New York Yankees	0.756	San Diego Padres	0.724
Miami Marlins	0.775	Baltimore Orioles	0.752	Tampa Bay Rays	0.723
Seattle Mariners	0.774	Colorado Rockies	0.750	<b>Milwaukee Brewers***</b>	<b>0.716</b>

\*Maximum. \*\*Average. \*\*\*Minimum.

Figure 1: Strikeouts Over the Years (1871-2017)

### 3 Results

Figure 2A shows that strikeouts have been growing at a nearly exponential rate since the inauguration of professional baseball in 1871. Figure 2B considers the statistic contact average. There is a strong negative correlation as the correlation coefficient for this figure as it is  $-0.925$ . Looking at the past twenty seasons (1998-2017) in Figure 3, there is again a strong negative correlation, but slightly weaker than the correlation coefficient from Figure 2B, with a correlation coefficient of  $-0.907$ . However, the correlation coefficient for the past 13 seasons (Figure 4) has a near perfect negative correlation with a correlation coefficient of  $-0.993$ .

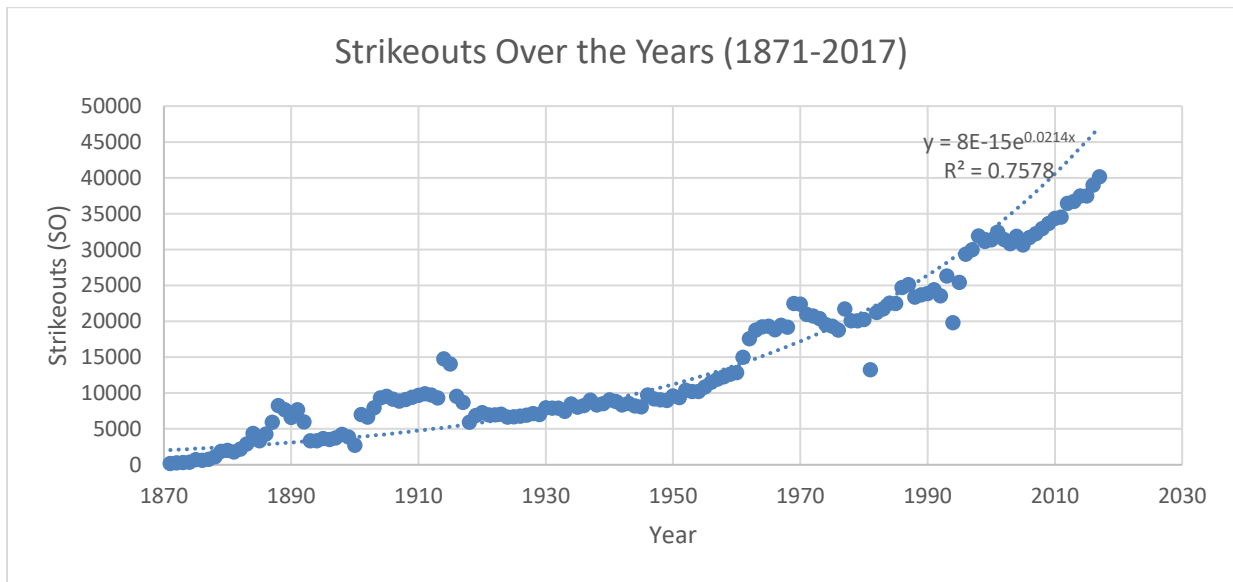


Figure 2A: Strikeouts Over the Years (1871-2017)

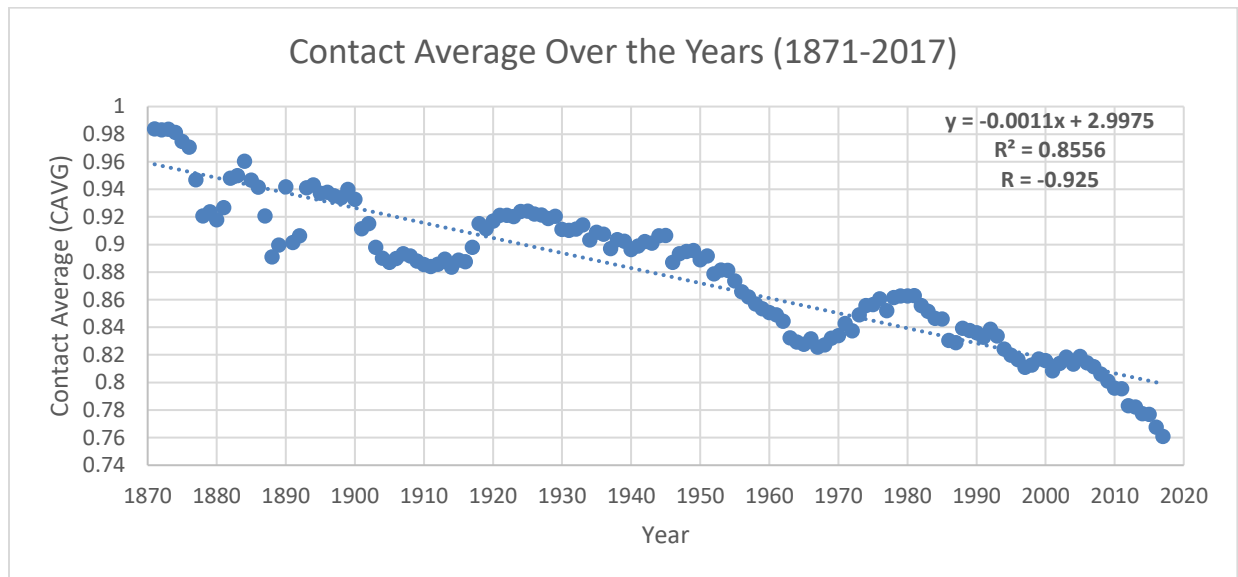


Figure 2B: Contact Average for the Past 20 Seasons

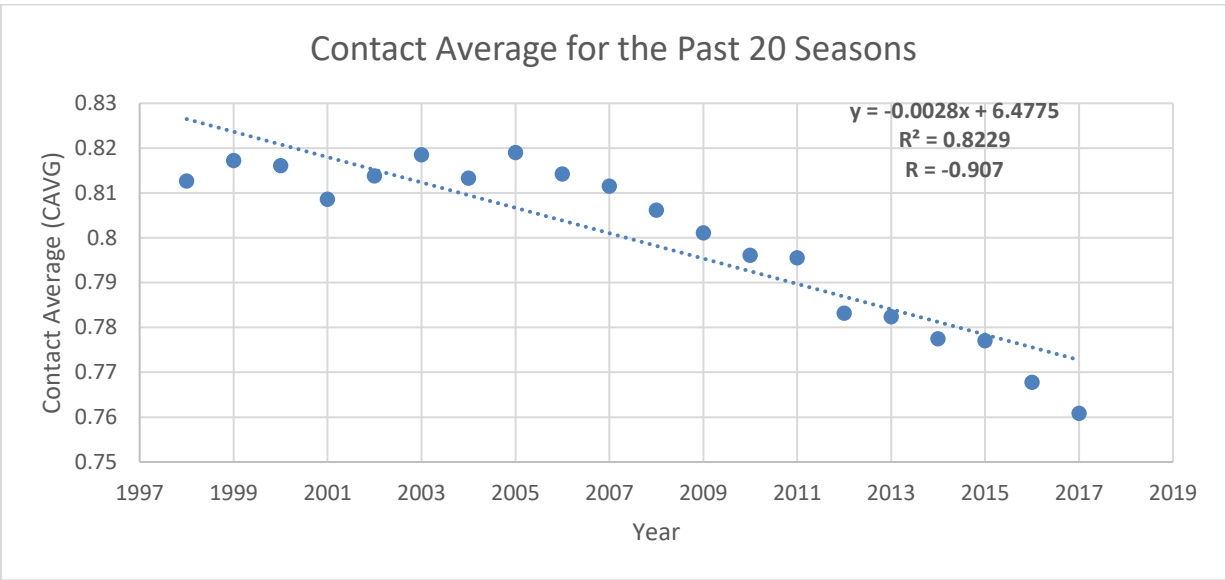


Figure 3: Contact Average for the Past 20 Seasons

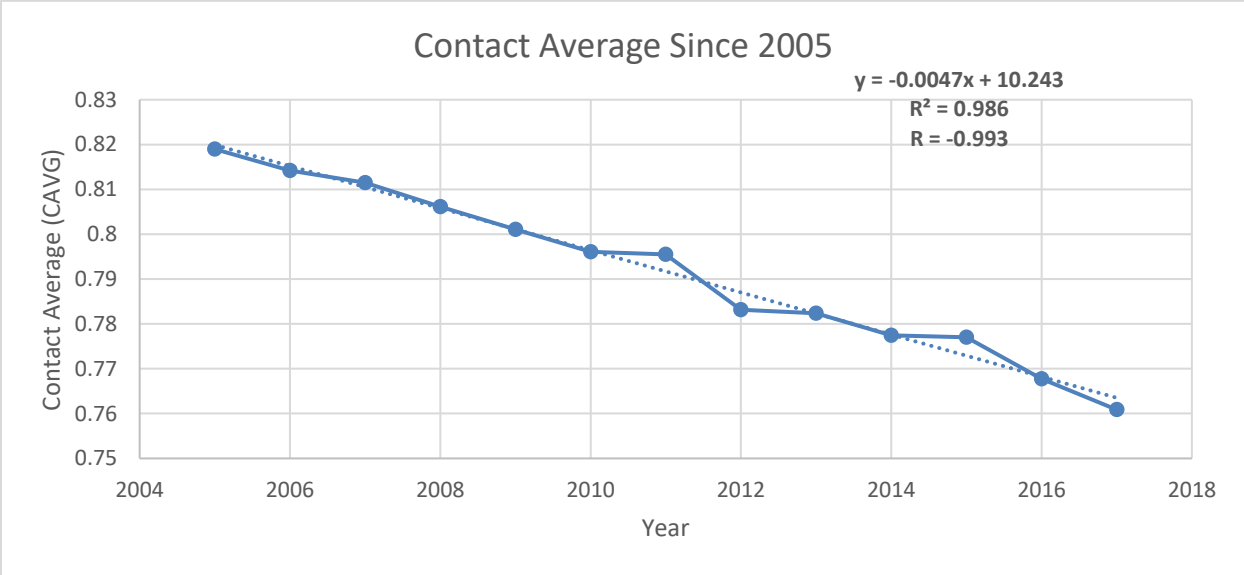


Figure 4: Contact Average for the Past 13 Seasons

### 4 Conclusion and Directions for Further Research

The team with the highest contact average in 2017 was the Houston Astros, who just so happened to win the 2017 World Series. At first glance, Figure 1 seems to show that there could be a correlation with team success and contact average. However, the teams with the third, fourth, and fifth highest contact averages had more losses than wins in the 2017 season. Surprisingly, the range in which major league baseball teams make contact is between 0.700 and 0.800, much lower than one would expect from a professional league.

Figure 2A shows, there has been an almost exponential increase in strikeouts since the inauguration of professional baseball. However, one must account for the fact that there were fewer teams in the early days of the game. As such, since the statistic of contact average will be considered for the entire league, it does not matter how many teams are playing. When Figure 2B is considered, there is still alarming news. Over the past 146 years, baseball players have been making less and less contact with the baseball. The correlation coefficient for this 146-year time period is -0.925 meaning that it is a very strong negative correlation.

Looking at Figure 2B, there are two apparent valleys— both could be explained through the two deadball eras in baseball history. The first Deadball Era was from 1901 until 1920. During this time there was some difference in major rules. One of them was the “Foul strike” rule. This rule stated a batter could hit as many foul balls as they wished without having to worry about them being called strikes. The spitball was another contributing factor as it was legal at this time. A spitball is a ball that has been tampered with by the pitcher’s saliva. It changes the aerodynamic properties of the ball as well as reduces the friction between the pitcher’s fingers and the ball. One final factor was the lack of replacement of the baseball used. Often times, a game would be played with one or two baseballs and by the end of the game it would be so damaged it would be hard to hit, giving the advantage to the pitcher. Eventually this Deadball Era was overcome with a couple rule changes. The first rule change was the banishment of the spitball. The next was the requirement of cleaner baseballs, allowing the batter to have a better view of it towards the end of the game. This rule change came from the untimely death of Ray Chapman from a pitch that was hard to see. Finally, baseballs were made with help from a machine rather than completed by hand. Pitchers claimed that this caused the ball was to be springier, but the manufacturers denied this [3]. Nevertheless, it helped improve offensive output.

The Second Deadball Era was roughly from 1964 to 1972. Many claim that offensive output was low due to the expanded strike zone and the elite pitchers who were at their prime at this time. Such names included Sandy Koufax, Bob Gibson, Nolan Ryan and many more. Others claimed it was due to the expansion of teams and new ballparks. Earned run averages were low and so were batting averages. However, in the mid 1970s, offensive statistics began to increase again. As a result, some say that the Second Deadball Era was just a phase in which pitchers dominated for an extended amount of time. In 1969, the pitcher’s mound was also dropped five inches and strike zone was also reduced in size [5]. This may also have contributed to increased offensive success.

Since the focus of this research was to look at the past 20 seasons, the author decided to include Figure 3 to see how major league baseball has fared with making contact since introducing 30 teams. The results had a slightly weaker correlation than the past 146 years which showed hope that perhaps the strong negative trend would be weakening. The opposite turns out to be true when looking more closely at the data. Figure 4 shows how terrible making contact has been since 2005. The correlation coefficient is nearly perfect at -0.993. It makes sense though with every season since 2008 having more strikeouts than the year before it. And not just from an arbitrary amount, but a record amount each year.

Figures 5 and 6 (in the appendices) compare Contact Average with on base percentage, home runs, and team wins, respectively. The first figure shows that a weak to moderate strength in correlation between contact average and on base percentage. This makes sense because as much as putting the ball in play increases the chances of getting on base by putting pressure on the defense, it is still not the only way a player could get on base. Figure 6 goes along with the cliché in baseball that players who hit home runs tend to make less contact due to their focus on power. Again, there is not a strong correlation, just a weak to moderate one.

Now the natural question arises, does contact average make a difference when it comes to team success? Like the previous two figures, there really was not a strong correlation between a high contact average and a higher chance of team success. There is a back and forth in the data in recent years, which shows that between both sides of the field there is a constant competition of who can beat who. One year it could be the pitcher's year while the other it could be the hitter's. Thus, it can be concluded that the relationship between the two statistics is insignificant. This makes sense when considering the many different manners in which a team could make contact. This statistic makes a home run, which is an automatic run, on the same level as a ground ball out. This means that a team that makes weak contact may have the same contact average as a team that makes harder contact or even contact that results in balls that do not result in an out.

The goal of this statistic was to see how often a team put a ball in play and its correlation with winning. A strikeout does not put the ball in play, but does it really make a difference? It may seem counterintuitive that a team that strike outs often has the same chance of winning as their opponents, but it is important to realize that the majority of teams were very similar in terms of their contact average. It seems that baseball teams tend to be near the 80% mark. As mentioned earlier, the different ways that the team makes contact would actually make a difference. Nevertheless, the data shows that putting the ball in play does not make a difference.

What if contact average was combined with another statistic? In baseball, there is a statistic called On Base Percentage Plus Slugging (OPS) that was created in the 1980s and is considered by some to be one of the first Saber metric statistic. "Saber metrics" comes from the acronym SABR (The Society of American Baseball Research) [4]. Basically, OPS combines two important baseball statistics and places them into one category. It measures a player's ability to hit for power and get on base. Figure 8 shows how important OPS is when it comes to team wins with a pretty moderate correlation between the two statistics. Moreover, Figures 9 and 10 show that On Base Percentage Plus Contact and Contact Plus Slugging (CPS) are not quite as significant to team wins as OPS is. Figure 11 combines OPS with Contact Average and finds that it really did not improve the chances of team success.

There are a couple future considerations for contact average that may be of interest to look into. Since not all contact is equal and often times it results in an out, perhaps a statistic that determines a productive out could be created. A productive out would be defined as one where a player makes contact and causes a baserunner to be moved over. It would be based off of how many opportunities a player had. The hypothesis would be that a team with a higher productive out average would have a higher chance to win. The next future consideration would be to look at how the alleged "Steroid era" of the 90's and early 2000's fare with contact average. Did steroids raise the amount of contact? Finally, it would be interesting to look at whether or not the current era of baseball is in the midst of a third dead ball era. This would involve looking at the previous two dead ball eras and determining what caused them as well as a way to get out of it. With the first two deadball eras having about a 50-year space between them, it may be that there is a learning curve that is just part of the game.

## References

[1] At-bat (AB), retrieved on November 30, 2017, from <http://m.mlb.com/glossary/standard-stats/at-bat>

[2] Lind, Marchal, Wathen. (2011). *Statistical Techniques in Business and Economics*. New York, NY: McGraw-Hill/Irwin.

[3] Baseball Reference, retrieved on October 15, 2017, from <https://www.baseball-reference.com>.

[4] Wisewell, Aaron (2013) *The Baseball Coach*. Raleigh, NC; Lulu Press, Inc.

[5] Baseball Rule Change Timeline, retrieved on November 10, 2017 from <http://www.baseball-almanac.com/rulechng.shtml>.

## Appendices

Original spreadsheets are available upon request.

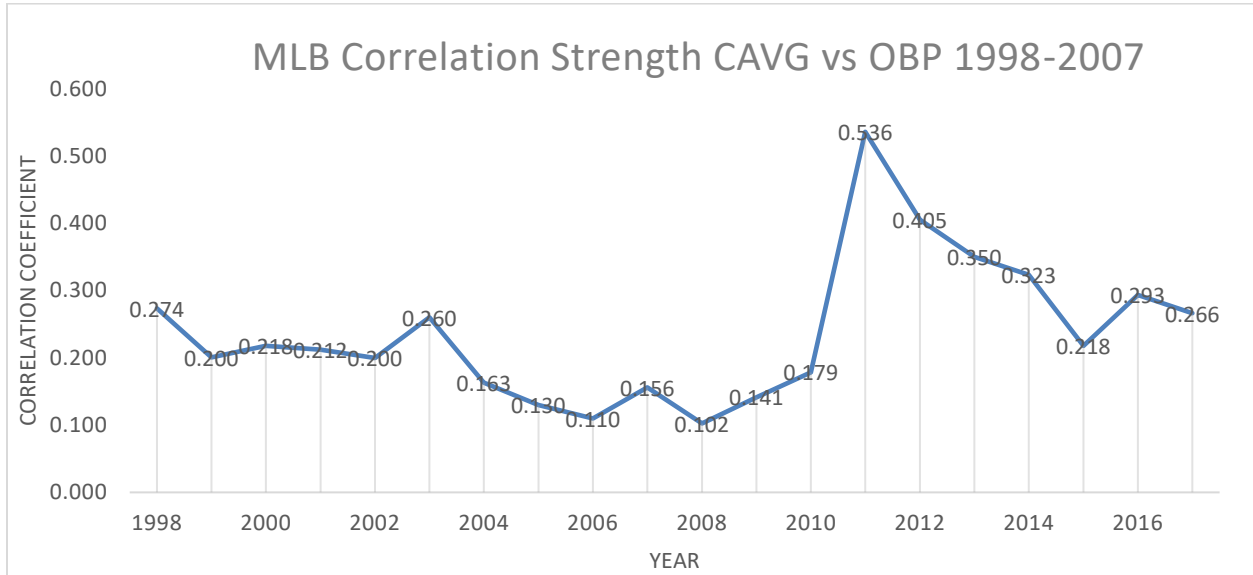


Figure 5: Correlation Coefficient of CAVG and OBP in Major League Baseball for past 20 years

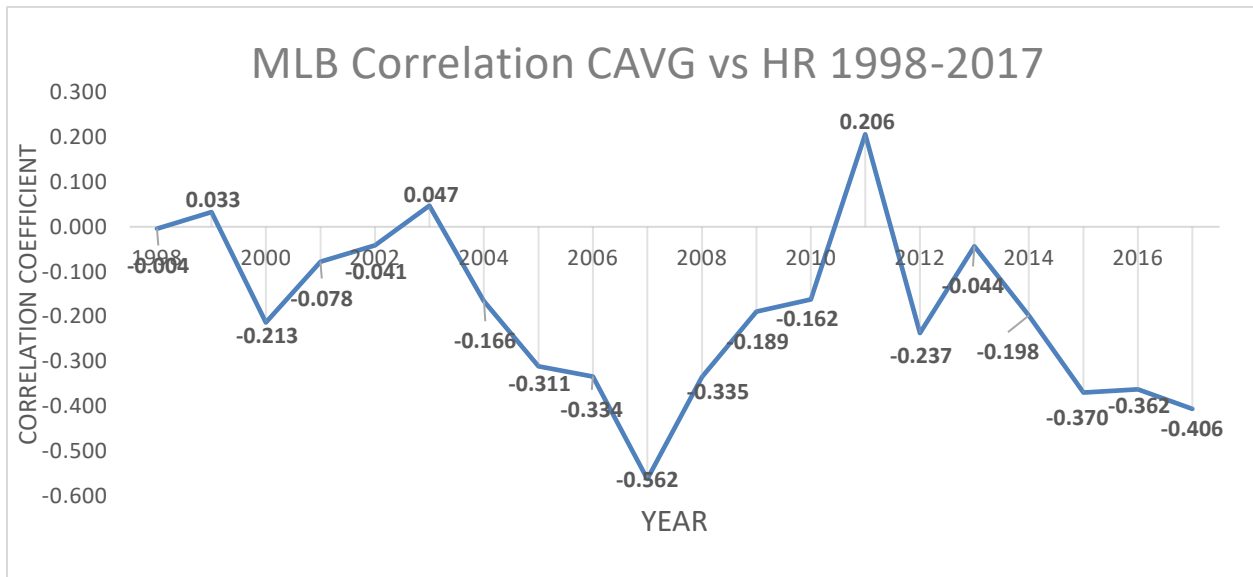


Figure 6 Correlation Coefficient of CAVG and HR in Major League Baseball for past 20 years



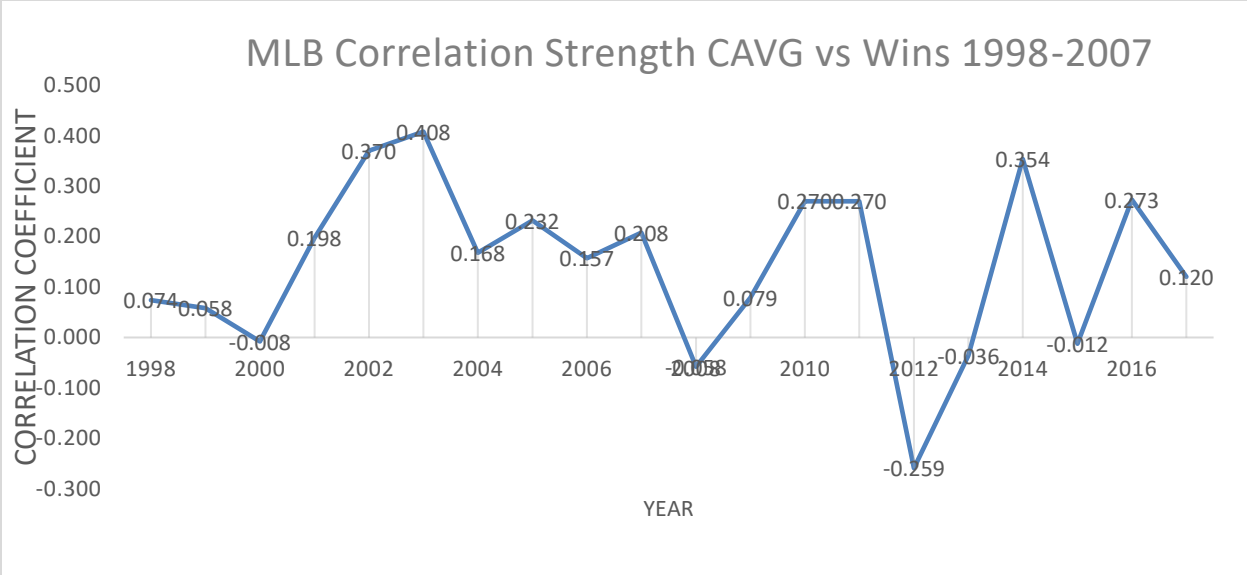


Figure 7: Correlation Coefficient of CAVG and Wins in Major League Baseball for past 20 years

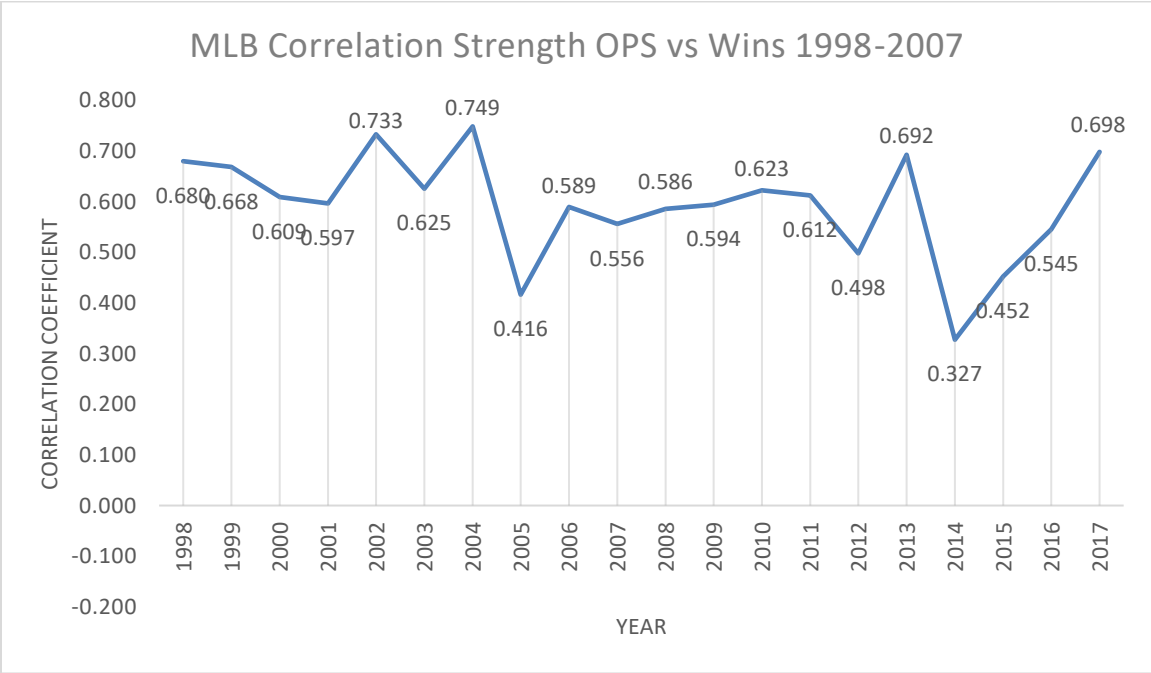


Figure 8: Correlation Strength of OPS and Wins

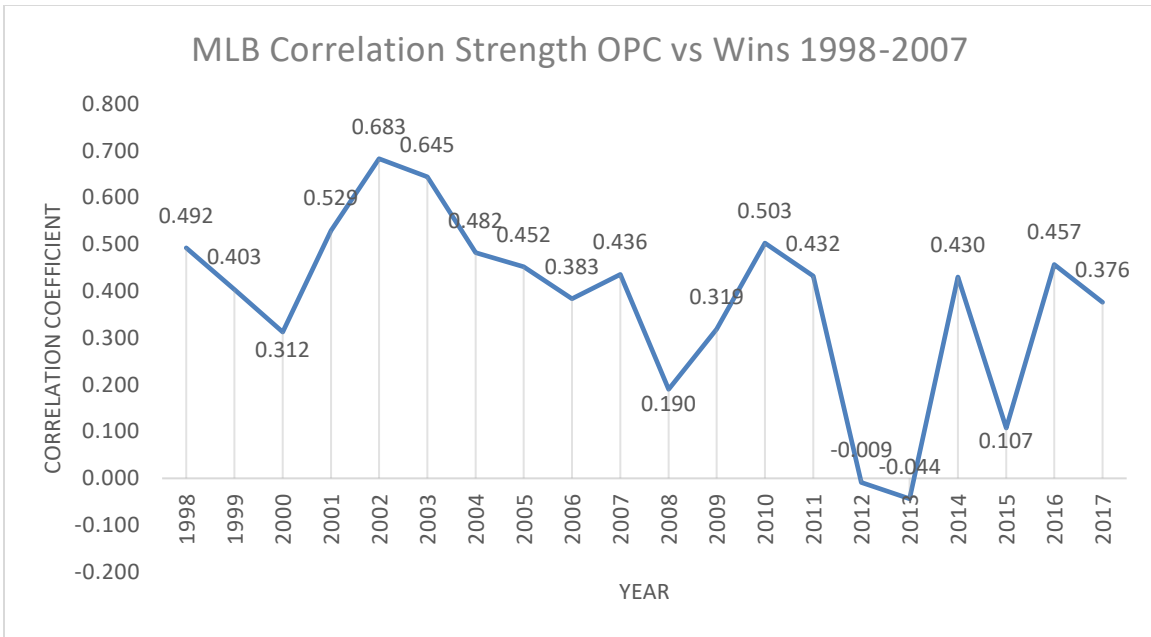


Figure 9: Correlation Strength of OBP Plus CAVG and Wins

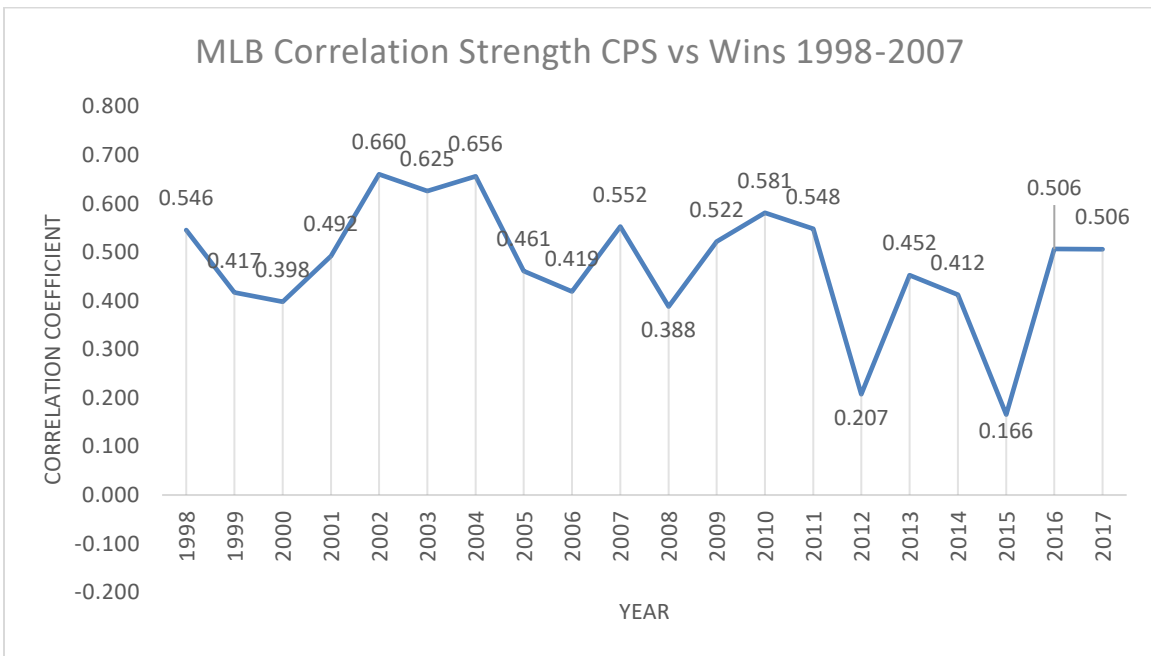


Figure 10: Correlation Strength of CAVG Plus Slugging and Wins

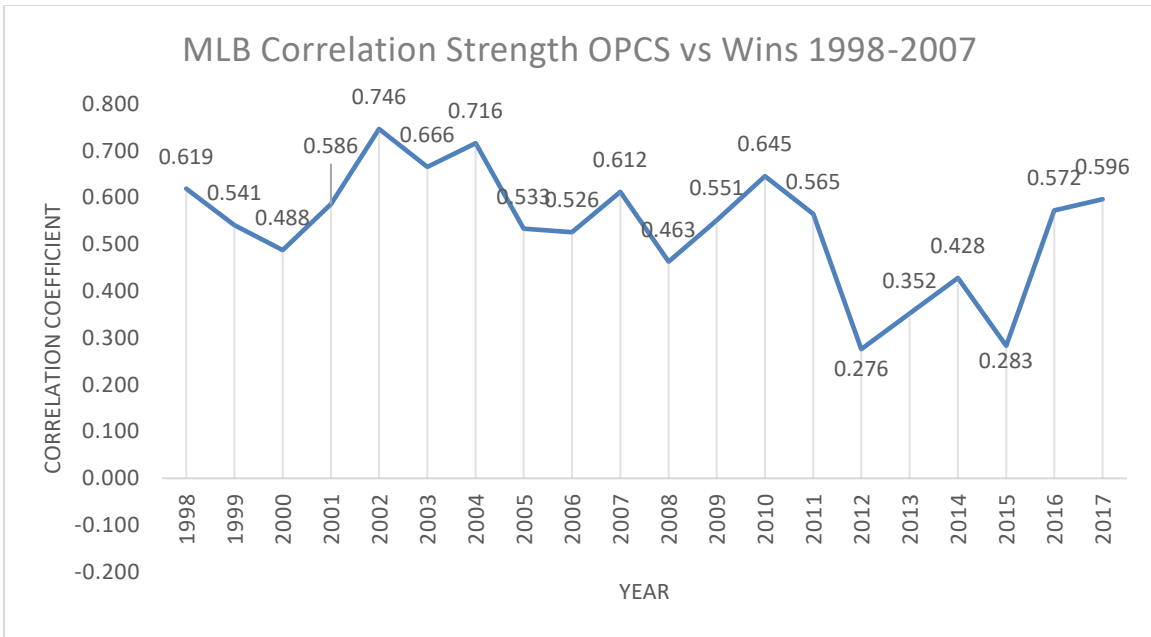


Figure 11: Correlation Strength of CAVG Plus Slugging and Wins

Career Contact Average	
<b>0.978068</b>	Lloyd Waner
<b>0.977239</b>	Nellie Fox
<b>0.953918</b>	Tony Gwynn
<b>0.892998</b>	Dustin Pedroia
<b>0.892617</b>	Ichiro Suzuki
<b>0.889422</b>	Hank Aaron
<b>0.887176</b>	Paul Molitor
<b>0.883477</b>	Albert Pujols
<b>0.814945</b>	Derek Jeter
<b>0.803638</b>	Prince Fielder
<b>0.797812</b>	Ryan Braun
<b>0.700129</b>	Jim Thome
<b>0.656462</b>	Adam Dunn
<b>0.648159</b>	Mark Reynolds

Figure 12: Notable Career Contact Averages of Past and Present Players