

## **How Many Wins are Needed to Make the MLB Playoffs?**

### **Part 1. Background**

I read Sayar Banerjee's recent article on the statistical basis at the heart of the book and movie Moneyball as well as the entire baseball sabermetrics movement (Ref. 1). The main point of moneyball (i.e., baseball sabermetrics) is to determine ways to win Major League Baseball (MLB) games and get into the playoffs with a limited budget. And how to select undervalued players based upon the needs of the team.

Mr. Banerjee's article gives some background of the moneyball push and explains the basis of the movie. He then gives his own statistical checks and Python code snippets to back up the numbers tossed around in the movie.

Recently I wrote some articles (Ref 2, for example) using neural networks to predict the outcomes of National Football League (NFL) games. I performed that effort because I just picked up Python starting in January and wanted to learn the language, have some fun, and tie it in to sports as well.

So let me turn over to baseball for a while at Mr. Banerjee's push.

### **Part 2. Re-Creation**

For other efforts I had previously scraped a lot of baseball, football, and ice hockey statistics from the web. Thus I grabbed my baseball statistics (Ref. 3) and started to reproduce Mr. Banerjee's results. (And also reproducing the basis of Moneyball.) I removed any season where less than 162 games were played due to strikes or other matters.

For baseball seasons between 1962 and 2001, I was able to reproduce the Moneyball and Banerjee results as shown in Figure 1. Namely, approximately 99 wins in a 162 game baseball season would "guarantee" a team of making the playoffs.

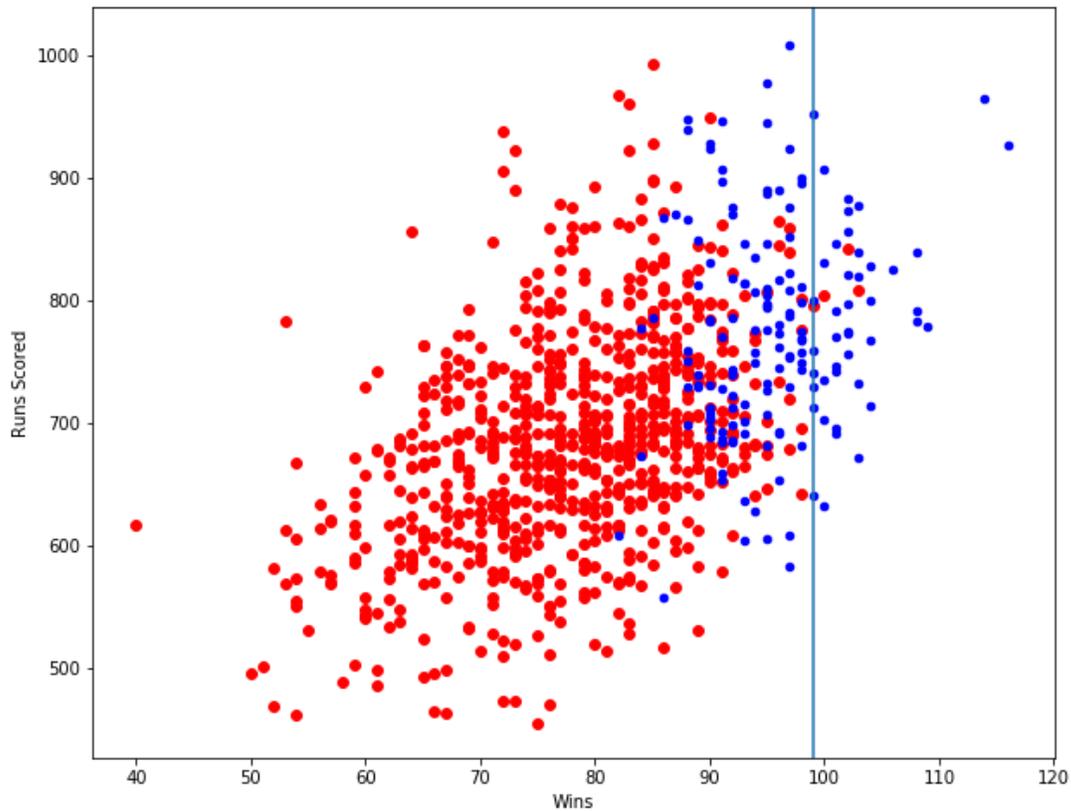


Figure 1. Estimate of Wins Needed to Make Playoffs (Year < 2002)

Figure 2 shows the number of wins accomplished for playoff teams starting from 1962 to 2001. The blue line is a pure least squares fit of the data.

Does it bother anyone that the linear regression line is sloping down indicating less wins are needed to make the playoffs now than twenty years ago? It shouldn't. The baseball rules have changed to allow more teams into the post-season play. This rule inherently means that it will take less wins to make it into the playoffs. Before 1969, only two teams made the playoffs and went directly to the World Series, i.e., the winner of the American League and the winner of the National League. Between 1969 and 1993 four teams made the playoffs. This was increased to eight teams making the playoffs from 1994 to 2011 as more teams and divisions were added. And finally in 2012, not yet shown, ten teams made the playoffs as MLB rules added a second wild card team. Thus the number of wins needed to make post-season play would certainly drop.

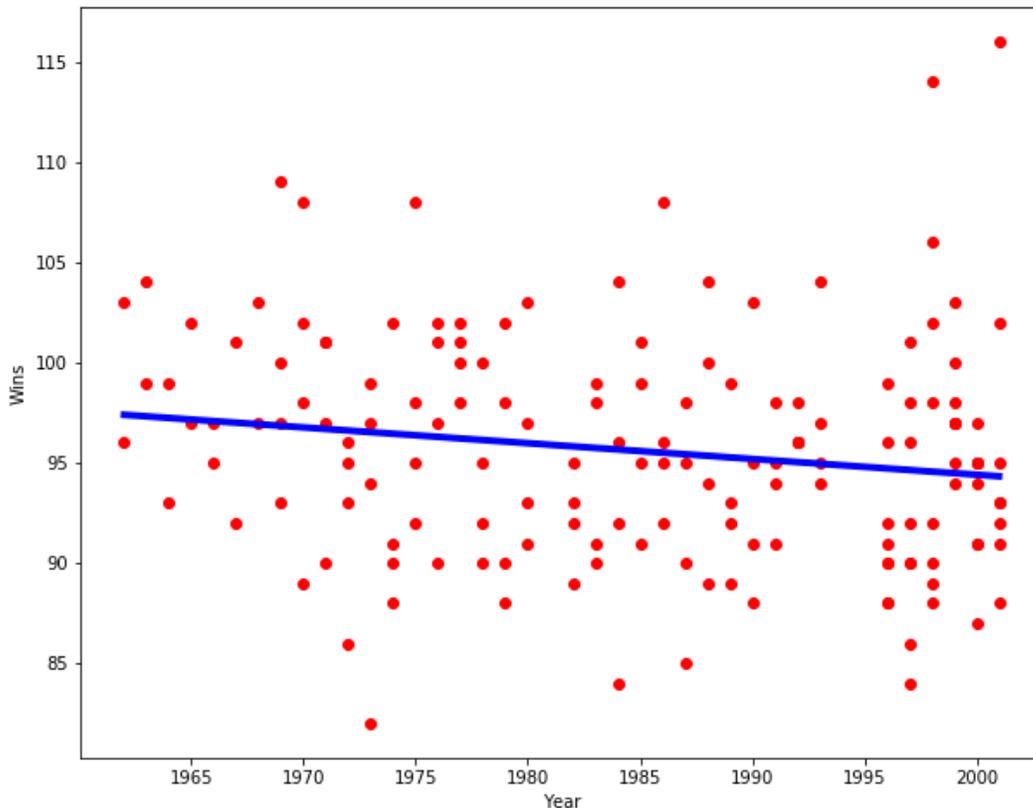


Figure 2. Historical Playoff Team Wins Per Season (Overall, Year < 2002)

Figure 3 shows the number of wins accomplished for playoff teams along with a piecewise least squares curve fit (in green for each period). The piecewise least squares curve fits separate the years based upon the number of teams allowed into the playoffs.

Prior to 1969 there is a generic drop in the number of wins needed to win the league and make it directly to the World Series. What is the cause of this drop? I don't know yet. This was before the era of free agency and the ability of players to refuse trades.

Between 1969 and 1993 there is a drop in the number of wins needed to make the playoffs. It's only about a two game drop in that period, but it points out how important every single game is for a team even when they've blown a four run lead and trail by one run going into the last two innings in a June game. Every game counts.

The post 1993 time period shown in Figure 3 shows a significant uptrend in the least squares number of games required to make the playoffs. This is probably skewed by the incredible 1998 New York Yankees with 114 wins and the 2001 Seattle Mariners

116 wins. But the “bottom end” of this period also show an increasing number of wins required to make post-season play from 84 games in 1997 to 88 games in 2001.

It's interesting to note for teams making the playoffs, the mean win values for each period of time are 98, 96, and 95 while the stand deviations are 4, 5, and 6 games. If one wanted to make the claim that you had to play better than 1 standard deviation above the mean to make the playoffs, it would take 102, 101, and 101 wins for each period of time.

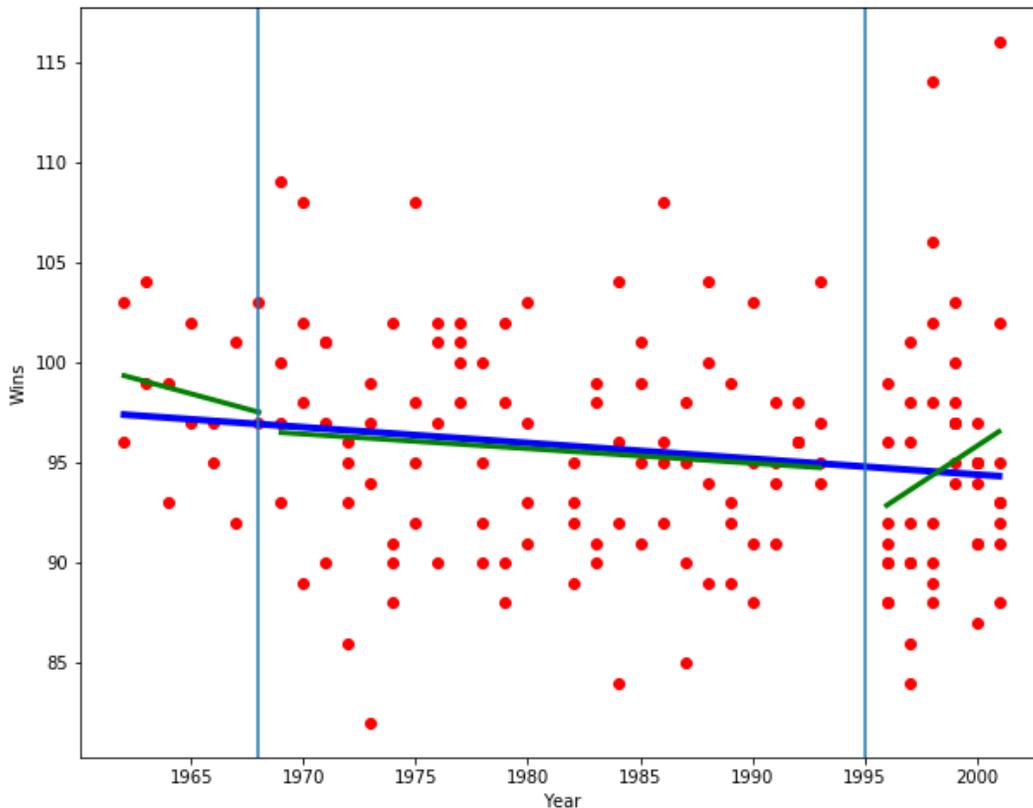


Figure 3. Historical Playoff Team Wins Per Season (Piecewise, Year < 2002)

### Part 3: Into Modern Times

I had the same set of team statistics available up through the completion of the 2017 season and I wanted to see what the trend was for this most recent period of time was. We should note that there was another change in rules starting with the 2012 season that allowed for a second wild card team to make the playoffs in each league.

Figure 4 shows the Wins against Runs Scored for all non-strike seasons from 1962 through 2017 inclusive. Again the 99 win vertical line is shown and the same three teams as before are the only ones with greater than 99 wins that did not make post-season play.

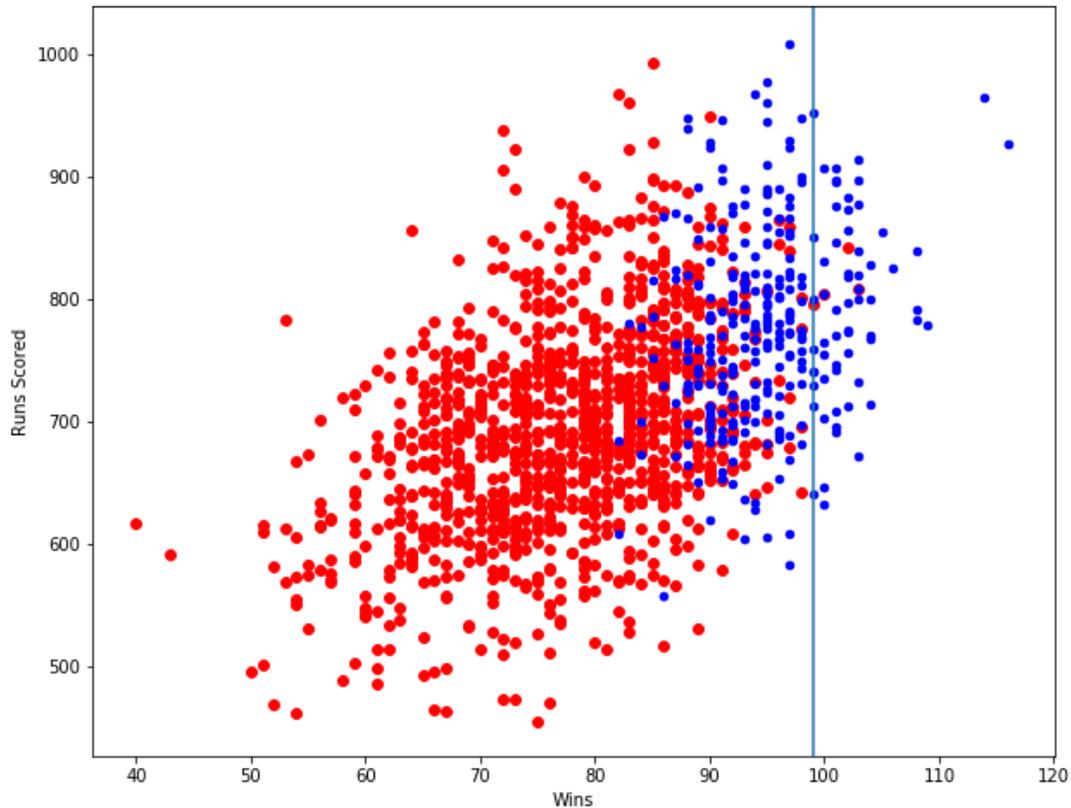


Figure 4. Estimate of Wins Needed to Make Playoffs (Year < 2018)

Figure 5 shows the piecewise least squares curve fits for each period of time in green (along with the overall least squares curve fit in blue). It's interesting to note that for the third time period, between the years 1994 and 2011, the green least squares curve fit has gone from a sharp positive slope (see Figure 3) to a slightly decreasing slope. This is a result of more seasons being played and the concept of “resorting to the norm” of the overall downward trend.

The mean number of wins for post-season teams are 98, 96, 94, and 93 for the four different playoff format time periods with standard deviations of 4, 5, 5, and 4 wins. Thus the mean plus one standard deviation to make the playoffs would be 102, 101, 99, and 97 wins.

Further examination of Figure 4 indicates that the 99 win vertical line could be moved down to 97 wins and not invalidate the findings.

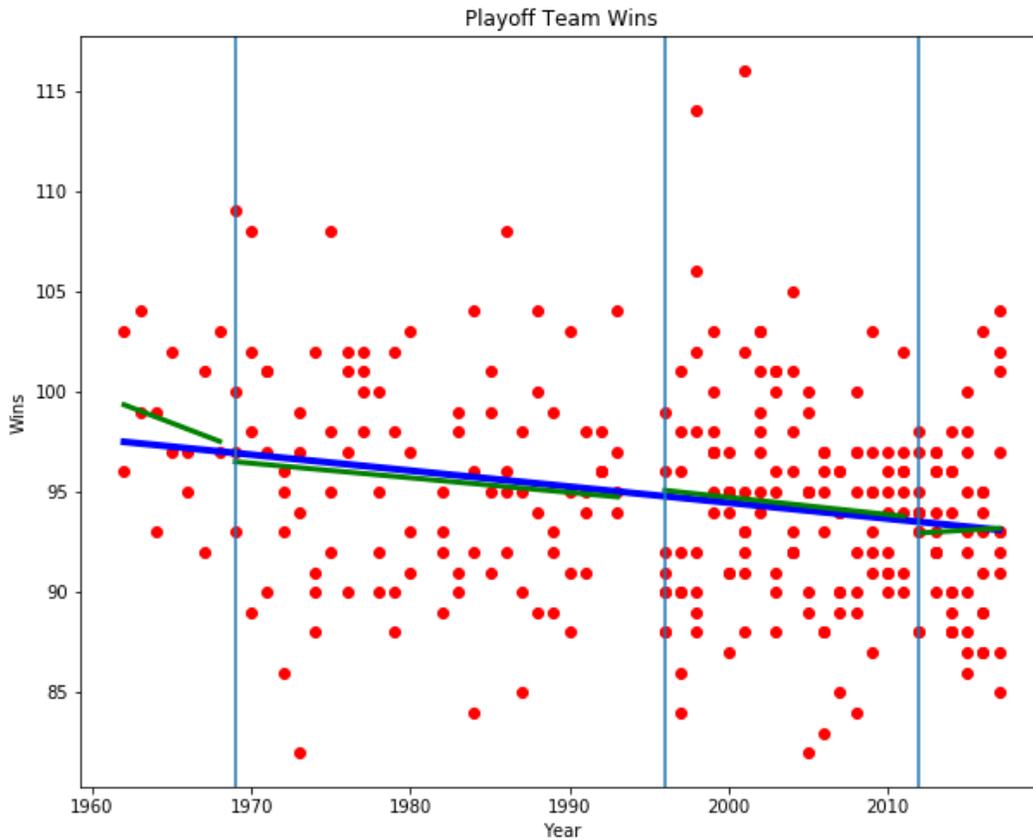


Figure 5. Historical Playoff Team Wins Per Season (Piecewise, Year < 2018)

#### Part 4. Next Steps

I hope to continue on and use the team statistics that I have available to derive a “model” of the factors that go into producing the 99 wins (or 97 wins) needed to make the playoffs.

#### References

1. Banerjee, Sayar, “Linear Regression: Moneyball – Part 1”, 15 April 2018, <https://towardsdatascience.com/linear-regression-moneyball-part-1-b93b3b9f5b53>.
2. Manning, Ray, “Predicting the Outcome of NFL Games – Success!”, 25 May 2018, <https://medium.com/@ray.90807/predicting-the-outcome-of-nfl-games-success-75a6ad13bfee>.
3. Lahman, Sean, <http://www.seanlahman.com/baseball-archive/statistics/>

