

# Standardization of Winning Streaks in Sports

Patrick R. McMullen

School of Business, Wake Forest University, Winston-Salem, USA

Email: mcmullpr@wfu.edu

**How to cite this paper:** McMullen, P.R. (2017) Standardization of Winning Streaks in Sports. *Applied Mathematics*, 8, 344-357. <https://doi.org/10.4236/am.2017.83029>

**Received:** March 1, 2017

**Accepted:** March 24, 2017

**Published:** March 27, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This research provides some metrics to better summarize streaks in sporting events with binary outcomes. In sporting events, information is often lost when “statistics” are presented regarding “streaks,” and whether or not certain teams or players have been recently been successful or unsuccessful. This usually leads to the presentation of metrics with no common baseline. This particular research effort provides statistics to capture the information regarding recent success or lack thereof, in a more standardized manner. To illustrate the presented metrics, data from the 2016 seasons for the American sports leagues National Basketball Association and Major League Baseball are used in an attempt to standardize streaks.

## Keywords

Streak, Statistic, Variation

---

## 1. Introduction

Sports are rife with statistics—some of them are useful, while others are not. In baseball, a player’s batting average is an important statistic. It tells us the number of base hits per plate appearances. A batting average of 0.300 (or 30%) is considered a very good batting average. A batting average of about 0.200 or less is not considered good. An Earned Run Average (ERA), which is the number of earned runs forfeited by a pitcher per every nine innings is another important statistic, which gives us an idea of the pitcher’s ability to prevent the opponent from scoring. An earned run average of 3.00 is considered good, whereas an ERA of 2.00 or less is considered exceptional. In basketball, a shooting percentage of 50% or better is typically considered good.

All of the above example statistics are informative, because the baseline is firmly established. That is, we know what the minimum possible values are, and with the exception of the earned run average in baseball, we also know the possible maximum value. The maximum possible value is, in theory, infinity, but such

a value will not exist, as a pitcher ineffective to that degree will not be permitted to continue pitching. Typically, an ERA of 4.50 or above is considered ineffective.

Unfortunately, not all sports statistics fall into the “tidy” category of having firmly established baselines. Streaks fall into this unfortunate category. Articulation of a streak, such as a winning streak or losing streak, is intended to inform the fan of recent success or lack thereof. While this can be informative, it can also be inconsistent. Consider, for example, a baseball team that has won its last nine consecutive games. Clearly, this is an impressive run. Also consider that this same team has won 12 out of its last fifteen games—an 80% winning percentage. Which measure is more important and/or informative? There is no clear answer to this question. Let us further assume that before this team won 12 out of fifteen games, they lost five consecutive games. Therefore, this team has won 60% of its last 20 games—slightly better than average. When we “stretch” the chronology of the measurement, the success becomes less impressive. Let us review this via a table:

Recent Time Frame (Games)	Winning Percentage
9	100%
15	80%
20	60%

When sports media talk about streaks, and/or recent runs of success or lack of success, there is a tendency to “package” the information in such a way that maximizes or enhances the success or lack of success. While the general point of this is understandable, such statistics are usually biased due to a small sample size.

The above is not intended to be critical of studying streaks. Streaks are important to show that binary outcomes in sports can on occasion defy expectation, and this is worth study. The intent here is to standardize streaks across a larger time frame. In short, it is intended here, to study streaks across an entire season, and isolate the teams that tend to show more “streakiness” than other teams. This standardization consists of a few new metrics to study consistency of winning and losing across an entire season. These metrics also consider the winning percentage of teams across the season. In other words, these analyses of streaks are adjusted for the team’s success. After the metrics are presented, they are used to assess the performance of all teams in American Major League Baseball, and all teams in the American National Basketball Association for their respective 2016 seasons.

## 2. Literature Review

Much work has been done to study streaks in sports. Perhaps one reason for this is due to Joe DiMaggio’s 56-game hitting streak in 1941. Many consider it the most impressive streak in sports history. Much effort has been put forth in an

attempt to better understand the forces at work during the streak, which has subsequently led to deeper understanding to streaks in general, and the entities that are related to streaks, which could be considered possible “causes” or contributing factors to the existence of streaks [1]. Effort has gone into deciding whether or not a resultant set of data actually qualifies as an actual “streak” [2], and much work has been done at trying to predict streaks [3] [4] [5]. Vallone and Tversky [6] demonstrated that single outcomes in sports are not related to prior outcomes. Streaks have even been studied so that gamblers can improve their chances of successful sports betting [7].

Streaks are difficult to measure, because there are no rules defining what constitutes a meaningful streak. Because of this, there are many opportunities to research what is considered a streak, and how meaningful a streak is [8]. In certain ways, a streak can be related to a degree of variation that exists in the data—the “streakier” the data is, the more variation the data will show. Conversely, the less streaky the data, the smaller the variation. This problem has been addressed [9]. The work in this paper attempts to extend this work by first studying descriptive statistics associated with win/loss streak performance of two sports leagues, and secondly understanding the relationship between expected wins and actual wins throughout a single season. This second motivation has exploited previous work associated with production scheduling [10] to generalize actual win/loss performance with expected win/lost performance.

### 3. Methodology

Here, we first describe descriptive statistics associated with win/loss streaks of sports teams. Next, we describe a “Gap” measure which compares a team’s actual win/loss performance to their expected performance though each game of the season. Finally, a “runs” test is described. The section concludes by illustrating the methodology via a simple, simulated data set.

#### 3.1. Streak Analysis

The first part of our methodology pertains to actual streaks: winning streaks and losing streaks. Here, we compute all descriptive statistics relevant to winning and losing streaks. Prior to delving to the mathematics of these metrics, a table of definitions is provided (see **Table 1**).

Let us assume that there are  $n$  games in a season,  $m$  teams, and  $w_{ij}$  represents team  $j$  winning game  $i$ , shown via the following:

$$w_{ij} = \begin{cases} 1 & \text{if team } j \text{ wins game } i, \forall i, j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The total number of wins for team  $j$  ( $Wins_j$ ) is computed as follows:

$$Wins_j = \sum_{i=1}^n w_{ij}, \forall j \tag{2}$$

Similarly, the total number of losses for team  $j$  ( $Losses_j$ ) is computed as follows:

**Table 1.** Terms used for analysis.

Value	Description
$w_{ij}$	Binary variable denoting win or loss for team $j$ in game $i$
$Wins_j$	Number of wins for team $j$
$Losses_j$	Number of losses for team $j$
$pct_j$	Winning percentage for team $j$
$WSL_{Ij}$	Length of the $I^{th}$ winning streak for team $j$
$WSC_j$	Number of winning streaks for team $j$
$MaxW_j$	Maximum winning Streak for team $j$
$LSL_{Ij}$	Length of the $I^{th}$ losing streak for team $j$
$LSC_j$	Number of losing streaks for team $j$
$MaxL_j$	Maximum losing streak for team $j$
$\bar{x}_{wj}$	Average winning streak length for team $j$
$\bar{x}_{lj}$	Average losing streak length for team $j$
$s_{wj}$	Standard deviation of winning streak length for team $j$
$s_{lj}$	Standard deviation of losing streak length for team $j$
$Gap_j$	Gap measure for team $j$

$$Losses_j = n - \sum_{i=1}^n w_{ij}, \forall j \tag{3}$$

The winning percentage for team  $j$  ( $Pct_j$ ) is computed as follows:

$$Pct_j = \frac{1}{n} \sum_{i=1}^n w_{ij}, \forall j \tag{4}$$

The above is trivial-we are simply comparing wins and losses for each team. More importantly, we wish to glean information from winning and losing streaks. In order to do this, we need to use the  $w_{ij}$  values to construct a list of winning and losing streaks for each team. In order to do this, we define the  $I^{th}$  winning streak as follows:

$$w_{aj} = 1, w_{a+1,j} = 1, \dots, w_{b-1,j} = 1, w_{bj} = 1, \forall j \tag{5}$$

In other words, team  $j$  wins all games, starting with game  $a$ , and ending with game  $b$ . The values of both  $w_{a-1,j}$  and  $w_{b+1,j}$  are zero. This results in the  $I^{th}$  winning streak of the following length:

$$WSL_{Ij} = 1 + (b - a), \forall j \tag{6}$$

The count of winning streaks for team  $j$  is incremented by one via the following:

$$WSC_j = WSC_j + 1, \forall j \tag{7}$$

It should be noted that for all  $j$ ,  $WSC_j$  is initialized to zero prior to analysis. The longest winning streak for team  $j$  is determined as follows:

$$MaxW_j = \max(WSL_{Ij}), \forall I, j \tag{8}$$

Calculating losing streak characteristics is done in similar fashion to winning

streaks. First of all, the  $J^{th}$  losing streak is defined as follows:

$$w_{aj} = 0, w_{a+1,j} = 0, \dots, w_{b-1,j} = 0, w_{bj} = 0, \forall j \tag{9}$$

Analogous to the case for winning streaks, the losing streak above has  $w_{a-1,j}$  and  $w_{b+1,j}$  values equal to 1. The length of this  $J^{th}$  losing streak for team  $j$  is computed as follows:

$$LSL_{ij} = 1 + (b - a), \forall j \tag{10}$$

For team  $j$ , the count of the losing streaks is incremented as follows:

$$LSC_j = LSC_j + 1, \forall j \tag{11}$$

As was the case with the winning streak counts, all  $j$  teams have their  $LSC_j$  values initialized to zero prior to analysis. The longest losing streak length for team  $j$  is as follows:

$$MaxL_j = \max(LSL_{ij}), \forall J, j \tag{12}$$

### 3.2. Gap Analysis

There is another measure of importance that is not directly related to streaks. That is the “smoothness” of a team’s success throughout the season. For example, if we assume a team wins 66.67% of their games, and they won two games then lost one, with this pattern repeating itself throughout the season, their winning pattern would map exactly to their winning percentage, and the “smoothness” of their winning would be optimal. In reality, of course, this does not happen, so it’s important to quantify the smoothness of teams’ winning patterns. We can quantify this via a variation of the “smoothness index” that has been used to study many scheduling algorithms [10]. Given the above definitions, we call this smoothness index the “Gap” measure, and each team has such a measure. It is calculated as follows:

$$Gap_j = \sqrt{\sum_{i=1}^n \left( \left( \sum_{h=1}^i w_{hj} \right) - i \cdot Pct_j \right)^2}, \forall j \tag{13}$$

This metric essentially tells us how many games team  $j$  has won through game  $i$  compared to how many games they are expected to win through game  $i$ . This difference is then squared, summed for all  $n$  games, and then the square root of this quantity is taken for standardization purposes. In layman’s terms, this metric tells us the smoothness of a team’s winning pattern. Lower quantities suggest more consistency in the winning patterns, while higher quantities suggest less consistency in winning patterns.

### 3.3. Runs Test

A “runs test” is a popular way to determine if a sequence of binary outcomes is truly random [11]. The runs test essentially has three properties that can be gathered from the sequence of binary outcomes:  $n_1$  is the total number of one type of binary outcome (Wins, regarding this effort),  $n_2$  is the total number of the other type of binary outcome (Losses, regarding this effort), while “ $r$ ” is the

number of streaks (or “runs”), analogous to  $WSC_j + LSC_j$  regarding this effort. We compute the mean, standard deviation, and associated  $z$ -score according to the following:

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1 \tag{14}$$

$$\sigma_r = \frac{(\mu_r - 1)(\mu_r - 2)}{(n_1 + n_2 - 1)} \tag{15}$$

$$z = \frac{r - \mu_r}{\sigma_r} \tag{16}$$

Given the standardized normal deviate, we can determine the two-tailed  $p$ -value is follows:

$$p = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{-|z|} e^{-x^2/2} dx \tag{17}$$

If the  $p$ -value associated with the test is less than a pre-specified critical value, we reject the null hypothesis and claim that the values comprising the sequence is not random. Otherwise, we fail to reject the null hypothesis and conclude that the values comprising the sequence are in fact random.

This research effort employs the runs test to see if the win-loss distribution is random or not.

### 3.4. Example Problem

A “toy” data set is presented to provide an illustration as to how the presented metrics work. The data set is binary data on the passing success of an American football quarterback. A “1” means an attempted pass was completed, a “0” means the attempted pass was not completed. The data set is simulated such that the percentage of completed passes is 57%. One hundred simulated passes were generated. The simulated data is as follows:

```
1 0 0 0 1 1 1 1 1 1 1 0 1 1 0 1 0 0 0 1 1 1 0 1 1 1 1 0 0 1 0 0 1 1 0 0 1 1 0 1 1 0 0 0
1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 0 1 0 1 1 1 1 0 1 0 1 0 1 1 0 1 0 1 0 0 1 0 1 0 1 1 0 0 1 1
0 1 0 1 0 1 0 0 1 1 0
```

This data set was used for the presented formulae, and the statistics are summarized accordingly. For the winning streak statistics, “ $n$ ” is used to represent  $WSC_p$ , “ $\bar{x}$ ” is used to represent  $\bar{x}_{W_j}$ , “ $s$ ” is used to represent  $s_{W_p}$  and “max” is used to represent  $MaxW_j$ . Similarly for losing streaks, “ $n$ ” is used to represent  $LSC_p$ , “ $\bar{x}$ ” is used to represent  $\bar{x}_{L_j}$ , “ $s$ ” is used to represent  $s_{L_p}$  and “max” is used to represent  $MaxL_j$ . Since there is only a single entity (one “team,” so to speak) for this example, the subscript is not recognized for convenience. The winning streak data and the losing streak data are segregated for the presentation below (see **Table 2**):

In the context of this example, a completion is considered a success (analogous to a “win”) while an incompleteness is considered a failure (analogous to a “loss”).

**Table 2.** Results for “Toy” data set.

Comp.	Incomp.	Pct	Completion Streaks				Incompletion Streaks				Gap	Runs	
			n	$\bar{x}$	s	max	n	$\bar{x}$	s	max		z	p
57	43	0.57	29	1.97	1.43	7	30	1.47	0.73	3	12.92	0.29	0.7691

The summary tells us that there were 57 completed passes and 43 incomplete passes for a 57% completion percentage. There were 29 completion streaks, each having a mean length of 1.97 completions and a standard deviation of 1.43 completions. The maximum streak of completed passes was 7. There were 30 incompletion tasks, each having a mean length of 1.47 incompletions and a standard deviation of 0.73 incompletions, with a maximum streak of incomplete passes of 3. The “Gap” measure, or “smoothness” measure here is 12.92. This value is not particularly informative here because there is no basis for comparison. This measure will be more informative when there are several entities used for comparison. Given the high p-value associated with the runs test, we cannot reject the null hypothesis of randomness.

### 3.5. Experimentation

The simple example above is used to merely illustrate the use of the presented statistics. It is important to apply the presented methodology to real data to better understand the “streakiness,” consistency and/or winning patterns for real sports teams. As such, the 2016 performances of all National Basketball Association (NBA) teams and all Major League Baseball (MLB) teams are studied and compared. Both the NBA and MLB are very popular sports leagues in the United States and abroad. In particular, the NBA has gained much international interest in recent years, while MLB has gained interest in the Caribbean region, along with Japan and South Korea.

It is of particular interest to understand which teams are most “streaky,” or inconsistent. The presented methodology is intended to shed some light on this fundamental question.

## 4. Results

This section presents the results for the aforementioned leagues-Major League Baseball and the National Basketball Association. The results are then discussed in some detail.

### 4.1. Major League Baseball

MLB has 162 games on their regular season schedule. However, not all of the thirty teams play this number of games, due to rain postponements and the like. If a team affected by a postponement is eligible for a post-season playoff berth, any missed games will be made up if the result of the makeup game impacts the post-season scenario. Any other postponed games will not be made up unless a makeup game is convenient for both affected teams. Because of this, not all

teams play the full 162 games. In 2016, most teams did in fact play 162 games, but a few played 161 games due to weather postponements<sup>1</sup>.

With this said, the Chicago Cubs were the best team in the regular season with a winning percentage of 0.6398, while the Minnesota Twins were the worst team in the league with a winning percentage of 0.3642. Full results are shown in **Table 3**.

**Table 3.** MLB results.

Team	W	L	Pct	Winning Streaks				Losing Streaks				Gap	Runs	
				n	$\bar{x}$	s	max	n	$\bar{x}$	s	max		z	p
ARI	69	93	0.4259	41	1.68	1.23	5	41	2.27	1.53	6	17.35	-0.20	0.8389
ATL	68	93	0.4224	36	1.83	1.59	7	38	2.53	2.17	9	76.3	0.61	0.5428
BAL	89	73	0.5494	40	2.25	1.75	7	38	1.92	1.22	5	27.15	0.35	0.7238
BOS	93	69	0.5741	39	2.38	1.95	11	39	1.77	0.87	4	20.9	0.25	0.8043
CHC	103	58	0.6398	38	2.68	2.17	11	39	1.54	0.97	5	31.18	-0.21	0.8324
CHW	78	84	0.4815	39	2	1.32	6	39	2.15	1.53	7	33.91	0.42	0.6711
CIN	68	94	0.4198	38	1.79	1.09	5	37	2.54	1.99	11	30.56	0.54	0.5898
CLE	94	67	0.5839	39	2.33	2.39	14	41	1.73	0.84	3	18.22	-0.09	0.9303
COL	75	87	0.463	40	1.88	1.2	5	40	2.18	1.36	6	21.56	0.17	0.8628
DET	86	75	0.5342	34	2.53	1.85	8	34	2.24	1.37	7	19.62	1.35	0.1772
HOU	84	78	0.5185	41	2.05	1.41	7	42	1.88	1.06	5	35.01	-0.13	0.9003
KC	81	81	0.5	37	2.19	1.68	9	37	2.19	1.66	8	23.09	0.85	0.3958
LAA	74	88	0.4568	38	1.97	1.35	6	37	2.38	2.05	11	29.41	0.69	0.4910
LAD	91	71	0.5617	42	2.17	1.41	6	42	1.69	1.05	6	18.6	-0.37	0.7100
MIA	79	82	0.4907	43	1.84	1.19	7	44	1.89	1.06	5	34.09	-0.64	0.5214
MIL	73	89	0.4506	41	1.78	0.96	4	41	2.17	1.28	6	15.46	-0.09	0.9289
MIN	59	103	0.3642	33	1.79	0.93	4	33	3.12	2.72	13	36.33	1.08	0.2781
NYM	87	75	0.537	41	2.12	1.54	8	43	1.77	1.09	4	27.02	-0.28	0.7808
NYN	84	78	0.5185	40	2.1	1.5	7	42	1.88	1.27	6	24.58	-0.01	0.9900
OAK	69	93	0.4259	34	2.03	1.27	6	34	2.74	1.69	7	16.58	1.27	0.2037
PHI	71	91	0.4383	44	1.64	1.14	6	43	2.12	1.82	9	32.83	-0.73	0.4657
PIT	78	83	0.4845	36	2.17	1.54	7	36	2.33	1.67	8	31.81	0.99	0.3205
SD	68	94	0.4198	43	1.58	0.73	3	44	2.14	1.34	5	13.45	-0.84	0.4019
SEA	86	76	0.5309	40	2.15	1.49	8	41	1.85	1.35	6	22.45	0.08	0.9385
SF	87	75	0.537	41	2.12	1.69	8	40	1.88	1.26	6	46.7	0.06	0.9505
STL	86	76	0.5309	44	1.95	1.24	5	44	1.73	0.95	5	10.86	-0.73	0.4630
TB	68	94	0.4198	36	1.89	1.12	4	36	2.61	2.14	11	22.91	0.85	0.3951
TEX	95	67	0.5864	38	2.5	1.64	7	38	1.76	0.97	4	21.68	0.40	0.6912
TOR	89	73	0.5494	43	2.07	1.32	7	42	1.74	0.99	5	21.59	-0.44	0.6629
WAS	95	67	0.5864	38	2.5	1.54	7	37	1.81	1.27	7	17.61	0.50	0.6138

<sup>1</sup>The Miami Marlins and the Atlanta Braves cancelled a game in observance of the unexpected death of a Miami player. This game was not made up because neither team was playoff eligible, and a make-up game was not convenient for either team.



In terms of winning streaks, the Chicago Cubs had the longest average winning streak of 2.68 wins/streak. They were closely followed by the Washington Nationals (2.5), Texas Rangers (2.5) and the Cleveland Indians (2.33). It is also worth noting that the Cleveland Indians had a (14) game winning streak in 2016, the longest of the season. For losing streaks, the Minnesota Twins had the longest average losing streak of 3.3 losses/streak, while the Tampa Bay Rays and Atlanta Braves also had long average losing streaks (2.61 and 2.53 respectively).

Clearly, better teams will have longer average winning streaks, while lesser teams will have longer average losing streaks. In fact, the correlation between the two entities is  $-0.5235$ , which is statistically significant ( $p = 0.0030$ ).

In terms of consistency, or lack of “streakiness,” the following teams did well: St. Louis, San Diego, Milwaukee, Oakland, Arizona, Washington, Cleveland, Los Angeles Dodgers and Detroit, all having “Gap” measures of less than 20. These teams, in effect, consistently won and lost games at a rate in accordance with their overall winning percentage. They had relatively small maximum winning streaks, and relatively small maximum losing streaks. The exception to this was Cleveland, whose long 14 game winning streak was basically offset by a very short maximum losing streak of 3 games.

In terms of inconsistency, or “streakiness,” the Atlanta Braves are seen to be the most salient, which a “Gap” measure of 76.3. This high number can be understood by taking note of their maximum winning streak of (7) games and their maximum losing streak of (9) games—both long streaks, which detract from their consistency. San Francisco was also “streaky,” with a maximum winning streak of (8) games, and a maximum losing streak of (6) games.

It should be noted that there is no significant correlation between a team’s winning percentage and their “Gap” measure—the correlation is  $-0.2175$ , with a  $p$ -value of 0.2483. As such, the “Gap” measure does provide information beyond a team’s winning percentage.

The runs test for this data set is not informative, because in all instances, the conclusion is random sequences of wins and losses.

## 4.2. National Basketball Association

In the NBA, there are (82) regular-season games scheduled. Unlike MLB, these games are not postponed due to weather. As such, all teams play 82 games in a regular season.

For the 2016 regular NBA season, the Golden State Warriors had the best winning percentage (0.8902), which was the best record in league history—no team had ever won 73 games in a season before Golden State accomplished this. Conversely, the Philadelphia 76 ers had the worst winning percentage in the league (0.1220). All findings are shown in **Table 4**.

In terms of winning streaks, Golden State had the longest average winning streak of 7.3 wins/streak, followed by San Antonio, with 5.15 average wins/streak. It is also worth noting that Golden State started the season with a record-breaking 24-game win streak. For losing streaks, Philadelphia had the longest

**Table 4.** NBA results.

Team	W	L	Pct	Winning Streaks				Losing Streaks				Gap	Runs	
				n	$\bar{x}$	s	max	n	$\bar{x}$	s	max		z	p
ATL	48	34	0.5854	19	2.53	1.84	7	20	1.7	0.8	3	10.82	0.28	0.7771
BOS	48	34	0.5854	21	2.33	1.46	5	19	1.79	1.03	4	11.14	0.13	0.8982
BRK	21	61	0.2561	17	1.24	0.44	2	18	3.39	2.4	10	10.68	-0.52	0.6011
CHA	48	34	0.5854	20	2.4	1.57	7	20	1.7	1.42	7	24.6	0.13	0.8982
CHI	42	40	0.5122	21	2	1.38	6	20	2	1.26	5	18.06	0.15	0.8788
CLE	57	25	0.6951	17	3.35	2.32	8	19	1.37	0.6	3	11.36	-0.04	0.9664
DAL	42	40	0.5122	23	1.83	1.53	6	24	1.71	1	5	12.73	-0.84	0.3997
DEN	33	49	0.4024	22	1.5	0.86	4	22	2.23	1.74	8	8.6	-0.60	0.5486
DET	44	38	0.5366	23	1.96	1.19	5	21	1.81	1.08	5	10.44	-0.36	0.7178
GS	73	9	0.8902	10	7.3	6.58	24	9	1	0	1	12.69	-0.54	0.5890
HOU	41	41	0.5	21	1.95	1.2	5	21	1.95	1.07	4	9.49	0.00	1.0000
IND	45	37	0.5488	22	2.05	1.21	6	22	1.68	0.89	3	11.45	-0.39	0.6959
LAC	53	29	0.6463	20	2.65	2.28	10	21	1.43	0.75	3	11.4	-0.43	0.6674
LAL	17	65	0.2073	14	1.29	0.61	3	13	5	2.31	10	7.94	0.18	0.8544
MEM	42	40	0.5122	21	2	1.14	5	22	1.82	1.4	6	24.92	-0.16	0.8697
MIA	48	34	0.5854	24	2	1.1	5	25	1.4	0.82	4	9.17	-1.4	0.1485
MIL	33	49	0.4024	20	1.65	0.88	4	21	2.33	1.15	5	7.68	-0.09	0.9273
MIN	29	53	0.3537	21	1.43	0.68	3	19	2.79	1.96	9	17.15	-0.26	0.7984
NO	30	52	0.3659	20	1.5	0.69	3	21	2.48	1.6	6	14.53	-0.33	0.7423
NY	32	50	0.3902	19	1.68	1	4	19	2.63	1.54	7	19.36	0.32	0.7492
OKC	55	27	0.6707	18	3.06	2.21	8	19	1.47	0.61	3	14.25	0.04	0.9705
ORL	35	47	0.4268	22	1.59	1.01	5	24	2	1.79	8	19.18	-0.83	0.4088
PHI	10	72	0.122	10	1	0	1	11	6.55	6.04	18	11.87	-0.64	0.5224
PHX	23	59	0.2805	18	1.33	0.59	3	17	3.47	3.45	13	18.95	-0.16	0.8717
POR	44	38	0.5366	22	2.05	1.46	6	20	1.9	1.59	7	26.81	-0.03	0.9721
SA	67	15	0.8171	13	5.15	3.16	13	13	1.15	0.55	3	11.49	-0.10	0.9190
SAC	33	49	0.4024	20	1.65	1.09	5	22	2.27	1.49	6	9.43	-0.26	0.7974
TOR	56	26	0.6829	18	3.11	2.35	11	17	1.53	0.72	3	13.23	0.25	0.8011
UT	40	42	0.4878	21	1.9	1.37	7	22	1.91	1.06	5	10.3	-0.16	0.8697
WAS	41	41	0.5	22	1.86	1.32	5	21	1.95	1.12	5	13.38	-0.16	0.8728

average losing streak of 6.55 losses/streak, followed by the Los Angeles Lakers, who averaged 5 losses/streak. Also, it is clear that there is a correlation between average lengths of winning streaks and losing streaks (0.5342), which is statistically significant ( $p = 0.0024$ ). It is also worth noting that for Golden State, their average losing streak length was (1) loss/streak, and for Philadelphia, their average winning streak was (1) win/streak. In other words, Golden State never had a losing streak exceed (1) game, and Philadelphia never had a winning streak exceed (1) game.

The most consistent or least “streaky” teams were the Milwaukee Bucks, Los Angeles Lakers and the Denver Nuggets, all with “Gap” measures under (9). None of these teams had winning streaks in excess of (4) wins, and their maximum losing streaks, while as high as (10) losses for Los Angeles, were nevertheless consistent with their dismal winning percentages. In short, these three teams won their games fairly proportionally to winning percentages.

There are three teams that are very streaky, or inconsistent throughout the season: the Portland Trail Blazers, the Memphis Grizzlies, and the Charlotte Hornets. Despite all of these teams being good, with winning percentages above 0.500, and making the playoffs, inconsistency throughout the regular season is prevalent. Portland had a maximum winning streak of (6) games and a maximum losing streak of (7) games. Memphis had a maximum winning streak of (5) games and a maximum losing streak of (6) games. Charlotte had maximum winning and losing streaks of (7) games each. These lengthy winning and losing streaks increase the “Gap” measure to the “top of the list.”

As is the case with MLB, there is no correlation between a team’s winning percentage and their “Gap” measure. The correlation is 0.0197, with an associated p-value of 0.9178. Given this lack of relationship, the “Gap” measure does provide information beyond winning percentage.

As was the case for the runs test used for MLB, the result is not informative, as the runs tests informs us that the sequence of wins and losses is random.

### 4.3. Comparison of MLB and NBA

There is a vast difference in the win/loss dynamic between the NBA and MLB. There is much more disparity in the NBA as compared to MLB. **Table 5** shows a general breakdown of winning percentage statistics.

The average winning percentage for any league will always be 50%, because every team’s win is offset by another team’s loss. The standard deviation in winning percentage (Albert, 2012) through the league is a different story, however. In MLB this value is 6.62%, but in the NBA it is 16.92%—the NBA has much more performance parity as compared to MLB. The same applies to the winning percentage gap between the best and worst teams in the league. The difference in winning percentage between the Chicago Cubs and the Minnesota Twins (the best and worst teams in the league) is 27.56%. Similarly, the difference in winning percentage between the Golden State Warriors and the Philadelphia 76 ers is 76.82%—an immense difference in success.

**Figure 1** and **Figure 2** also show a difference in the win/loss dynamic when MLB is compared against the NBA. These two plots are organized such that the horizontal axis represents the winning percentage, while the vertical axis is the

**Table 5.** Winning percentage comparison.

	MLB	NBA
Std. Dev. in Winning Pct.	6.62%	16.92%
Winning Pct. Span	27.56%	76.82%

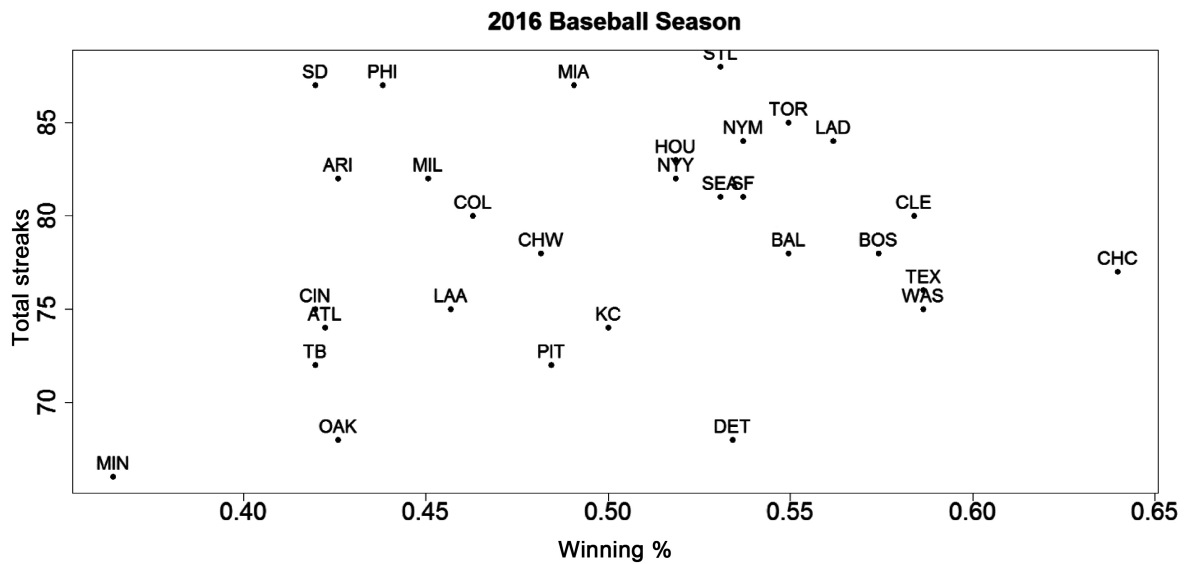


Figure 1. Winning percentage vs number of streaks for MLB.

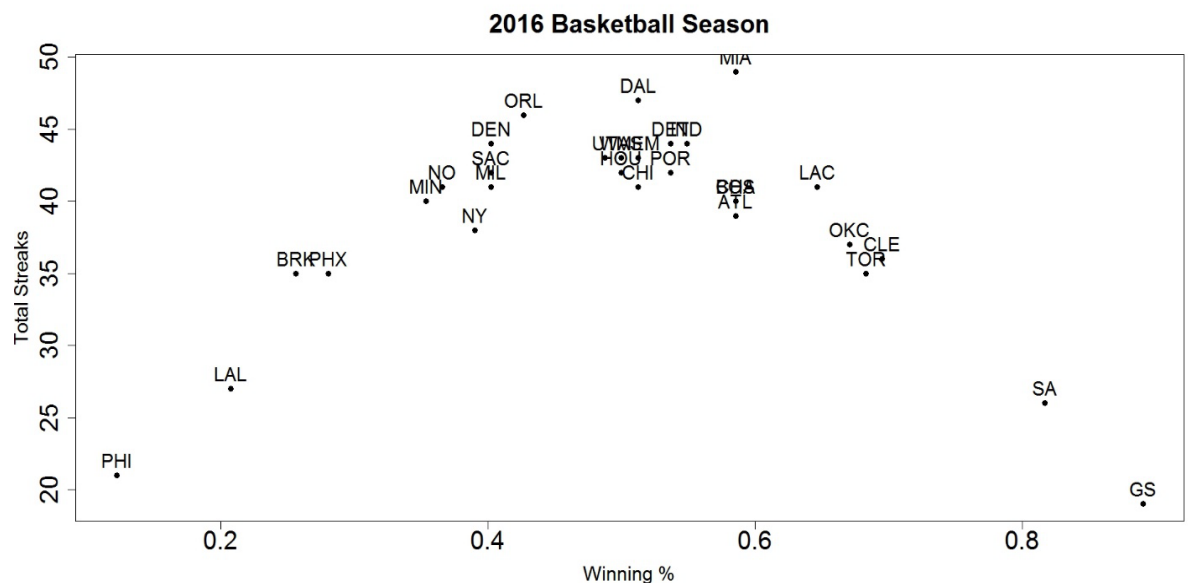


Figure 2. Winning percentage vs number of streaks for NBA.

sum of the number of winning streaks and losing streaks ( $WSC_j + LSC_j$ ) for each team.

Figure 1 shows no relationship between winning percentage and number of streaks-there is seemingly nothing interesting to report. Figure 2, however, shows a nonlinear relationship between winning percentage and number of streaks. The mediocre teams have more streaks than as compared to the teams that perform very poorly and/or very well.

Figure 3 shows the same as Figure 2, but with teams removed whose winning percentages are confined to match the range of MLB winning percentages. In other words, this filtered data set has omitted teams whose winning percentages are more extreme than those from the MLB data set.

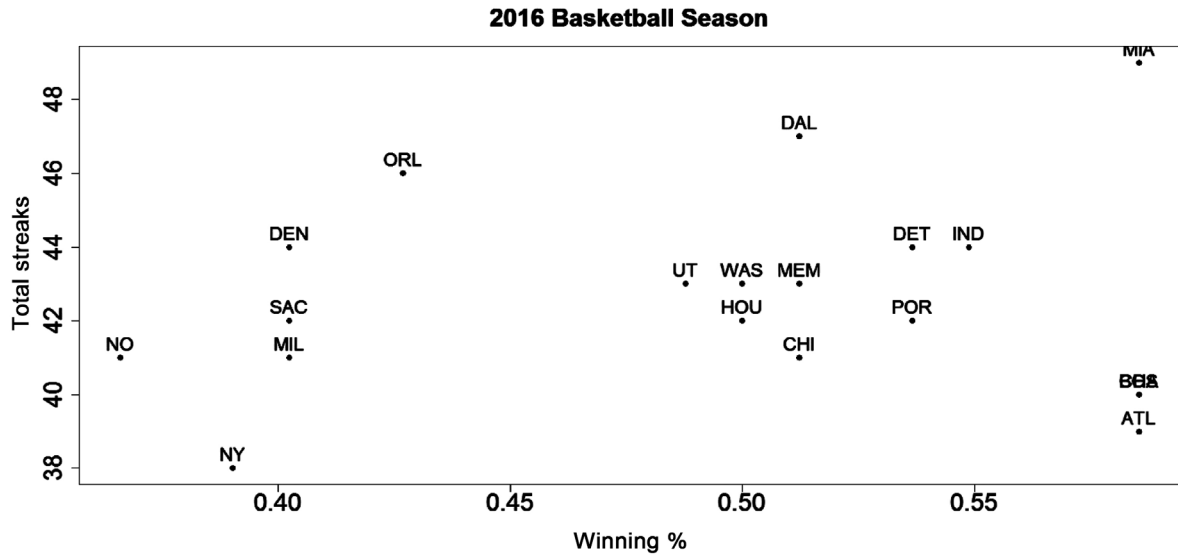


Figure 3. Winning percentage vs number of streaks for filtered NBA data set.

Figure 3 is more resembling of Figure 1—there is no relationship between winning percentage and the number of streaks when teams with extreme records are filtered out of the analysis.

The high disparity in success for the NBA as compared to MLB is a surprise finding. Nevertheless, this enables us to basically conclude that the NBA team performance is more “streaky” or less consistent as compared to MLB. Our streak statistics and “Gap” measure have demonstrated that.

### 5. Concluding Comments

Methodology has been presented in an attempt to standardize streaks in sports. We have presented these metrics in two different forms: studying streaks via descriptive statistics associated with the streaks themselves, and studying how a team performs throughout the season as compared to how they should perform according to their season winning percentage. The methodology was applied to the 2016 NBA and 2016 MLB seasons. Our findings have shown us that NBA performance involves much more disparity as compared to MLB. The reason for this, beyond statistical analysis, is beyond the scope of the paper.

This type of binary analysis involves winning or losing. Our “toy” problem data set used completions vs. incompletions for an American football quarterback. The binary nature of our data can be used for many other sporting applications: a baseball player’s batting average for all at bats (hit vs. no hit), a soccer player’s success regarding penalty kicks (goal vs no goal), a hockey player’s success regarding penalty shots (goal vs. no goal), etc.

This type of analysis can also be used for other binary outcomes outside the world of sports. For example, we can study market streaks at the close of some stock exchange-market increase vs. market decrease. We can study streaks regarding the success of salespeople-successful sales call (customer places an order) vs. an unsuccessful sales call. In short, the applications for studying streaks

with binary outcomes are only constrained by one's imagination.

## References

- [1] Albert, J. (2008) Streaky Hitting in Baseball. *Journal of Quantitative Analysis in Sports*, **4**. <https://doi.org/10.2202/1559-0410.1085>
- [2] Albright, C. (1993) A Statistical Analysis of Hitting Streaks in Baseball, *Journal of the American Statistical Association*, **88**, 1175-1183. <https://doi.org/10.1080/01621459.1993.10476395>
- [3] Rockoff, D.M. and Yates, P.A. (2009) Chasing DiMaggio: Streaks in Simulated Season Using Non-Constant At-Bats. *Journal of Quantitative Analysis in Sports*, **5**. <https://doi.org/10.2202/1559-0410.1167>
- [4] Arkes, J. (2010) Revisiting the Hot Hand Theory with Free Throw Data in a Multivariate Framework. *Journal of Quantitative Analysis in Sports*, **6**.
- [5] Thomas, A.C. (2010) That's the Second-Biggest Hitting Streak I've Ever Seen! Verifying Simulated Historical Extremes in Baseball. *Journal of Quantitative Analysis in Sports*, **6**. <https://doi.org/10.2202/1559-0410.1266>
- [6] Vallone, R. and Tversky, A. (1985) The Hot Hand in Basketball: On the Misperception of Random Sequences. *Cognitive Psychology*, **17**, 295-314. [https://doi.org/10.1016/0010-0285\(85\)90010-6](https://doi.org/10.1016/0010-0285(85)90010-6)
- [7] Bowman, R.A., Ashman, T. and Lambrinos, J. (2015) Using Sports Wagering Markets to Evaluate and Compare Team Winning Streaks in Sports. *American Journal of Operations Research*, **5**, 357-366. <https://doi.org/10.4236/ajor.2015.55029>
- [8] Albert, J. (2012) Streakiness in Team Performance. *Chance*, **17**, 37-43. <https://doi.org/10.1080/09332480.2004.10554913>
- [9] Albert, J. (2013) Looking at Spacings to Assess Streakiness. *Journal of Quantitative Analysis in Sports*, **9**, 151-163. <https://doi.org/10.1515/jqas-2012-0015>
- [10] Miltenburg, J. (1989) Level Schedules for Mixed Model Assembly Lines in Just in Time Production Systems. *Management Science*, **92**, 192-207. <https://doi.org/10.1287/mnsc.35.2.192>
- [11] Groeber, D.F., Shannon, P.W., Fry, P.C. and Smith, K.D. (2001) Business Statistics: A Decision-Making Process. 5th Edition, Prentice-Hall, New Jersey.



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [am@scirp.org](mailto:am@scirp.org)