

The Best Baseball Season Ever? A Triple Crown Perspective

Gregory A. Colston Jr.

Faculty Mentors: Dr. Joseph Nolan and Dr. David Agard



INTRODUCTION

In Major League Baseball, the Triple Crown is considered by most to be the greatest achievement a hitter can accomplish in a single season. To win a triple crown, a hitter must lead his league in the following three categories at the end of the season, with a minimum number of plate appearances¹ (PA):

- Batting average (BA) – Computed as the number of safe hits divided by the number of at-bats (plate appearances that result in a hit or out)
- Runs Batted In (RBI) – Runs scored by a team as a direct result of the batter's successful plate appearance
- Home Runs (HR) – The batter's successful hit results directly in his own advancement to score a run.

Recognized triple crown winners from 1901 to present are noted below:

Year	Player	Lg	Team	BA	HR	RBI
1967	Carl Yastrzemski	AL	BOS	.326	44	121
1966	Frank Robinson	AL	BAL	.316	49	122
1956	Mickey Mantle	AL	NYG	.353	52	130
1947	Ted Williams	AL	BOS	.343	32	114
1942	Ted Williams	AL	BOS	.356	36	137
1937	Joe Medwick	NL	STL	.374	31	154
1934	Lou Gehrig	AL	NYG	.363	49	165
1933	Jimmie Foxx	AL	PHA	.356	48	163
1933	Chuck Klein	NL	PHI	.368	28	120
1932	Jimmie Foxx ²	AL	PHA	.364	58	169
1925	Rogers Hornsby	NL	STL	.403	39	143
1922	Rogers Hornsby	NL	STL	.401	42	152
1909	Ty Cobb	AL	DET	.377	9	107
1901	Nap Lajoie	AL	PHA	.426	14	125

In the course of MLB history there have been many changes related to qualification for the batting title. In particular, if current requirements were used in 1932, Jimmie Fox would have won the batting title (and hence the triple crown since he also led the other two categories). We choose to recognize that (and use the current requirements throughout the analysis) in an effort to maintain comparability across seasons.

Our data³ consisted of the all hitting records from 1901 – 2010 seasons in both the American and National leagues. Our research goals included answering the following questions:

- The triple crown winners were all certainly great players. Which of these players had the “greatest” triple crown season?
- There have been a number of great players who did not win triple crowns? Which of these players were “closest” to winning?
- Which player(s) had the best overall careers in regards to the Triple Crown categories?

Because many rules were very different prior to 1901, we do not include data from that era in our analysis. However, for completeness we note that the following players did win triple crowns prior to 1901:

Year	Player	Lg	Team	BA	HR's	RBI's
1894	Hugh Duffy	NL	BOS	.440	18	145
1887	Tip O'neal	AA	STL	.435	14	123
1878	Paul Hines	NL	PRO	.358	4	50

¹ A plate appearance is any turn for a hitter to bat. Unlike at-bats, plate appearances also include walks, hit-by-pitch, etc. Under the current rule (implemented in 1967) a player must have at least 3.1 PAs per team game to qualify for the batting title. Players who fall short may meet this requirement by adding the necessary number of hitless at-bats.

² Not recognized by MLB as a triple crown.

³ Data obtained from <http://baseball1.com/statistics/>

STATISTICAL MEASURES FOR BEST SEASONS

Rankings For Best Season

Using the specified Triple Crown categories we employed several different statistical metrics to examine major league hitters in an effort to identify the best overall seasons. The following comparative measures represent the most interesting measures of “best” that we found:

Sum of Z-scores – summing a players z-score from each category to identify the highest summed Z-scores throughout all seasons.

- A Z-score is calculated by: $Z = (x - \mu) / \sigma$ where x is the players actual number (BA, RBI, or HR), μ is the league average, and σ is the league standard deviation. The table below shows the best 10 seasons based on the sum of Z-scores from all three categories. Of the triple crown winners in the first table, **Nap Lajoie** is identified as the strongest via this metric.

Using this metric allows for us to determine the relative amount by which players were better than the “average” player from their season. We can use these to compare players seasons since seasonal effects have been removed.

- One disadvantage of this metric is that the player with the highest summed z-score achieved it mostly with a single category. Babe Ruth was so much better at hitting home runs than every other player of the 1919 season that he earned a ridiculously high Z-score of 9.45 for HR's. On the other hand he wasn't remotely close to winning a triple crown.

Player	Year	AVG	HR	RBI	Z_AVG	Z_HR	Z_RBI	Z_Sum
Babe Ruth	1919	.322	29	114	1.41	9.45	2.64	13.49
Gavvy Cravath	1915	.285	24	115	2.55	3.37	6.79	12.71
Ernie Banks	1958	.313	47	129	3.50	2.92	5.38	11.80
Ty Cobb	1917	.383	6	102	7.02	1.63	2.81	11.46
Jeff Bagwell	1994	.368	39	116	3.28	4.48	3.31	11.07
Babe Ruth	1924	.378	46	121	4.11	4.50	2.07	10.67
Barry Bonds	2002	.370	46	110	5.14	3.84	1.26	10.23
Nap Lajoie	1901	.426	14	125	3.84	3.46	2.88	10.18
Frank Baker	1912	.347	10	130	2.30	4.42	3.00	9.72
Chuck Klein	1933	.368	28	120	3.09	3.89	2.74	9.72

Minimum of Z-scores – rather than the sum, only considering the player's 3rd best category means that they must be a well-rounded player to make the list.

- A “high” minimum Z-score indicates a great season in regard to all three triple crown categories. We believe that this method may be an ideal choice in comparing triple crown winners. The chart below contains those rankings.
- The 1919 season of Babe Ruth falls completely off the list due to the less impressive Z-score for average. On the other hand, Jeff Bagwell looks more impressive with a minimum Z-score of 3.28 in 1994.

Player	Year	Z_AVG	Z_HR	Z_RBI	Min_Z
Nap Lajoie	1901	3.839	3.464	2.876	2.876
Chuck Klein	1933	3.085	3.893	2.740	2.740
Joe Medwick	1937	2.440	2.399	2.450	2.399
Lou Gehrig	1934	2.512	2.829	2.123	2.123
Rogers Hornsby	1922	2.631	3.166	1.967	1.967
Ty Cobb	1909	2.606	3.069	1.841	1.841
Ted Williams	1942	3.523	2.249	1.705	1.705
Carl Yastrzemski	1967	3.041	2.691	1.701	1.701
Jimmie Foxx	1933	2.656	2.569	1.690	1.690
Rogers Hornsby	1925	3.480	2.798	1.522	1.522
Frank Robinson	1966	2.854	2.256	1.492	1.492
Jimmie Foxx	1932	1.803	2.537	1.160	1.160
Ted Williams	1947	2.649	2.086	1.083	1.083
Mickey Mantle	1956	2.085	2.407	0.961	0.961

Triple Crown “Closeness”

Our second goal was to measure the “closeness” of a hitter's season to winning the Triple Crown. We examined several statistical measures of “closeness”, the most interesting of which are presented below.

Sum of Ranks – this represents the sum of each player's ranks in each of the three triple crown categories.

- Triple Crown winners each have a rank-sum of three since they were 1st in each category. (Note: they cannot be distinguished by their rank-sum.)
- The table below identifies the ten best hitter-seasons for which a hitter did not win the triple crown. Not surprisingly, each of these players had a rank-sum of four (i.e. finished first in two categories and second in one category). Bold entries indicate their winning categories.

Player	Year	Lg	Team	AVG	HR	RBI
Al Rosen	1953	AL	CLE	.336	43	145
Johnny Mize	1940	NL	SLN	.314	43	137
Babe Ruth	1926	AL	NYA	.372	47	150
Gavvy Cravath	1913	NL	PHI	.341	19	128
Babe Ruth	1923	AL	NYA	.393	41	131
Babe Ruth	1924	AL	NYA	.378	46	121
Rogers Hornsby	1921	NL	SLN	.397	21	126
Cy Seymour	1905	NL	CIN	.377	8	121
Jimmie Foxx	1938	AL	BOS	.349	50	175
Honus Wagner	1908	NL	PIT	.354	10	109

Euclidean Distance (D) - This metric allows for us to measure the “distance” of a hitter to a “fictitious” Triple Crown winner:



Noting that triple crown winners have D = 0, the next best seasons according to this metric are Ted Williams (1949), Johnny Mize (1940), Al Rosen (1953), and Stan Musial (1948).

Necessary Additional Hits (NAH) - This metric allows for us to estimate the number of additional hits a player might have needed to win the Triple Crown. It is calculated via the maximum of the following NAH values:

- Smaller NAH means the player was “closer” to winning the Triple Crown; Triple crown winners have NAH = 0.
- Sixteen players throughout MLB history achieved NAH < 10 and seven of these (listed below) were within five hits of the triple crown.

Player	Year	AVG	HR	RBI	AVG_NAH	HR_NAH	RBI_NAH	NAH
Ted Williams	1949	0.343	43	159	1	0	0	1
Al Rosen	1953	0.336	43	145	1	0	0	1
Johnny Mize	1940	0.314	43	137	2	0	0	2
Babe Ruth	1926	0.372	47	150	3	0	0	3
Gavvy Cravath	1913	0.341	19	128	5	0	0	5
Dick Allen	1972	0.308	37	113	5	0	0	5
Hank Aaron	1963	0.319	44	130	5	0	0	5

LONGEVITY METRIC

Our final objective is to identify the best hitters throughout MLB history from a Triple Crown perspective. Using the concept of NAH, we developed a similar Longevity Metric (LNAH) that allows us to rank and determine the greatest all-time career hitters based upon the Triple Crown categories.

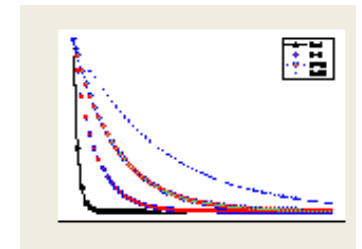
$$LNAH = \sum_{seasons} e^{-\frac{NAH}{k}}$$

This function was based on the following ideas:

- Seasonal contributions should be summed for each player. The maximum individual season contribution will be 1 and this can only be obtained by winning the Triple Crown.
- A player should neither be rewarded nor penalized for seasons in which they had a large NAH (e.g. rookie season, injury-shortened seasons, late career decline, etc.)

Additional technical considerations include:

- Euler's constant ($e \approx 2.72$) is used to achieve exponential decay.
- The constant ($k > 0$) adjusts the rate of decay and is somewhat subjectively chosen.
 - If the chosen k is too small, this would only credit players who were extremely “close” to winning the Triple Crown.
 - A value of k too large means that players who were “far” from winning the Triple Crown still receive substantial credit.
 - The graph below illustrates the affect of the constant k on the seasonal contribution to LNAH.



Based on LNAH ($k = 10$), our final rankings for the current top 20 all-time greatest hitters with respect to the Triple Crown are as follows:

Player	LNAH
Ted Williams	3.672
Babe Ruth	2.863
Rogers Hornsby	2.233
Jimmie Foxx	2.093
Chuck Klein	1.766
Nap Lajoie	1.302
Mickey Mantle	1.281
Lou Gehrig	1.209
Hank Aaron	1.176
Frank Robinson	1.120
Gavvy Cravath	1.105
Carl Yastrzemski	1.023
Al Rosen	1.020
Ty Cobb	1.002
Joe Medwick	1.000
Johnny Mize	0.956
Albert Pujols	0.902
Hank Greenburg	0.852
Barry Bonds	0.825
Mike Schmidt	0.724

Acknowledgement: This research was supported in part by a grant from NKU's Center for Integrative Science and Mathematics.