USING MATHEMATICS AND STATISTICS TO ANALYZE WHO ARE THE GREAT SLUGGERS IN BASEBALL

Randy Taylor
Las Positas College, United States
Steve Krevisky
Middlesex Community College, United States
skrevisky@mxcc.commnet.edu

In this presentation, we share the results of statistical work that we have done over several years, in order to determine who are the best sluggers in the game of baseball in the US. Using z scores, we examined yearly home run and slugging average figures, so as to analyze which batters were the most SD's above the mean. We used cutoffs of 200 at bats or 250 plate appearances before expansion, and increased this by about 5%, to account for the increased number of games played after the expansion in the early 1960s. Since real data are involved, we feel that this would be a very good application for students in a basic statistics class, and we will present various charts in the following discussion, and summarize our findings. This is a shortened version of what we have presented at numerous conferences in the US. Please let us know if you wish to see the full version of our paper!

INDIVIDUAL YEAR HOME RUN RESULTS

Gathering the statistics was much more time consuming than imagined. Data was obtained for all years 1920-2003 from *The Sports Encyclopedia: Baseball 2004* by David S. Neft, Richard M. Cohen, and Michael L. Neft. This book contained all the information required, but the print was very small. It also used numbers to indicate players who were traded, which was important in counting players totals in both leagues if they played in both leagues. We now look at *Z*-Scores for various Home Run Champions. The *Z*-Score is the number of standard deviation above the mean a player's total is and is given by:

Z-Score = (individual HR total – year's mean HR)/year's standard deviation

Next we calculate some American League results:

Home Run Champions 1920 – 2003					
American League					
Year	HR Mean	HR Standard Deviation	Name	HR	*Z*-Score
1920	4.85	7.27	Babe Ruth	54	6.76
1921	6.05	8.87	Babe Ruth	59	5.97
1927	5.42	10.05	Babe Ruth	60	5.43
1932	8.59	10.32	Jimmie Foxx	58	4.79
1938	11.60	11.88	Hank Greenberg	58	3.91
1956	13.34	9.39	Mickey Mantle	52	4.12
1961	15.01	12.34	Roger Maris	61	3.73
1967	11.87	8.99	Carl Yastrzemski Harmon Killebrew	44	3.57
1990	11.12	8.74	Cecil Fielder	51	4.56
2002	15.83	10.51	Alex Rodriguez	57	3.92

The top five *Z*-Scores are: (1) Babe Ruth, 1920, $z = 6.76$, (2) Babe Ruth, 1921, $z = 5.97$, (3) Babe Ruth, 1927, $z = 5.43$, (4) Jimmy Foxx, 1932, $z = 4.79$, and (5) Cecil Fielder, 1990, $z = 4.56$.

Next we calculate some National League results:

Home Run Champions 1920 – 2003					
National League					
Year	HR Mean	HR Standard Deviation	Name	HR	*Z*-Score
1922	6.31	7.18	Rogers Hornsby	42	4.97
1930	10.86	11.20	Hack Wilson	56	4.03
1949	10.87	9.78	Ralph Kiner	54	4.41
1954	14.01	11.70	Ted Kluszewski	49	2.99
1998	15.44	12.48	Mark McGwire	70	4.37
2001	18.03	13.37	Barry Bonds	73	4.11

The top five *Z*-Scores are: (1) Rogers Hornsby, 1922, $z = 4.97$, (2) Ralph Kiner, 1949, $z = 4.41$, (3) Mark McGwire, 1998, $z = 4.37$, (4) Barry Bonds, 2001, $z = 4.11$, and (5) Hack Wilson, 1930, $z = 4.03$.

INDIVIDUAL SLUGGING AVERAGES

Most batters who hit many Home Runs normally have high Slugging Averages, but some don't, like Harmon Killebrew. Also, many Slugging Average leaders were good all-around hitters and never lead in Home Runs, like Stan Musial. Using this category will give us another way to analyze who the top sluggers/power hitters are. We will be using *Z* scores so that each hitter is compared to the time that he played in.

- *Z* SCORE = (PLAYER'S SLA – MEAN SLA OF THE SEASON) / YEAR'S SD

First of all, we define Slugging Average, and show how it is computed.

- Slugging Average = Total Bases divided by Total at Bats.

SLA = TB / AB.

Example: You get 10 hits in 30 at bats and you have 3 doubles, 2 triples and 2 home runs. This is the data listed would be listed in *The Sports Encyclopedia: Baseball 2004*. First, you need the number of singles. How do we do this? There is a formula for this: We take the following approach:

- # Singles = # Hits – (# 2B + # 3B + # HR)

In this case, # S = 10 – (3 + 2 + 2) = 10 – 7 = 3.

Next, you get the # of totals bases as follows:

- # Total Bases = 1(#S) + 2(#2B) + 3(#3B) + 4(#HR)

This can be viewed as the dot product of 2 vectors, as follows:

- #TB = [1, 2, 3, 4] *[3, 3, 2, 2]

#TB = (1*3) + (2*3) + (3*2) + (4*2)
#TB = 3 + 6 + 6 + 8 = 23

Then, SLA = 23/30 = .767, which is quite good? It should be noted that this decimal number is often just call 767, just as a hitter batting .310 is frequently said to be hitting 310.

Now we calculate some American League results:

Slugging Average Champions 1920 - 2003					
American League					
Year	SLA Mean	SLA Standard Deviation	Name	SLA	Z-Score
1920	400.76	95.07	Babe Ruth	847	4.69
1921	431.29	84.98	Babe Ruth	846	4.88
1932	422.46	82.80	Jimmie Foxx	749	3.94
1937	437.91	77.45	Joe DiMaggio	673	3.04
1941	414.57	84.30	Ted Williams	735	3.80
1956	422.51	81.45	Mickey Mantle	705	3.47

The top five *Z*-Scores are: (2) Babe Ruth, 1921, $z = 4.88$, (2) Babe Ruth, 1920, $z = 4.69$, (3) Jimmy Foxx, 1932, $z = 3.94$, (4) Ted Williams, 1941, $z = 3.47$, and (5) Mickey Mantle, 1956, $z = 3.47$

Next we calculate some National League results:

Slugging Average Champions 1920 - 2003					
National League					
Year	SLA Mean	SLA Standard Deviation	Name	SLA	Z-Score
1922	437.68	80.84	Rogers Hornsby	756	3.94
1930	467.46	116.51	Hack Wilson	723	2.19
1948	405.03	76.71	Stan Musial	702	3.87
1949	416.11	75.31	Ralph Kiner	658	3.21
1954	431.84	86.72	Willie Mays	667	2.71
1998	432.25	98.95	Mark McGwire	752	3.23
2001	451.64	98.42	Barry Bonds	863	4.18

The top five *Z*-Scores are: (1) Barry Bonds, 2001, $z = 4.18$, (2) Rogers Hornsby, 1922, $z = 3.94$, (3) Stan Musial, 1948, $z = 3.87$, (4) Mark McGwire, 1998, $z = 3.23$, and (5) Ralph Kiner, 1949, $z = 3.21$.

CONCLUSIONS AND IMPLICATIONS FOR FURTHER STUDY

We hope to eventually have a complete data base for all the years from 1920-2006. This will take time, but we think that this work will shed more light on who the great sluggers in baseball are. We think that students and professors alike would enjoy this work, and many a good project could stem from this activity. We could also do this for individual players' careers! Finally, we note the following:

> Even using lower cuts offs for At Bats of 250 before 1961/1962 and 265 afterward, some significant Home Runs totals were still not used in the calculations. For example:
>
> 1960 Johnny Blanchard, New York Yankees, 243 AB, 21 HR

1963 Johnny Blanchard, New York Yankees, 218 AB, 16 HR
1966 Joe Adcock, California Angels, 231 AB, 18 HR
1996 Gary Sheffield, Florida Marlins, 218 AB, 16 HR
2000 Mark McGwire, St. Louis Cardinals, 236 AB, 32 HR

WE'D LOVE TO GET FEEDBACK FROM YOU ON WHAT WE HAVE DONE HERE!

REFERENCES

Taylor, R. and Krevisky, S. (2005). Using mathematics to catch the wave of great baseball sluggers. Paper presented at annual meeting of the American Mathematical Association of Two Year Colleges (AMATYC), San Diego.

Cohen, R., Neft, D. and Neft, M. (2004). *The Sports Encyclopedia Baseball.* New York: St. Martin's.