NUMB3RS Activity: Is It for Real?
Episode: "Hardball"

Topic: Data analysis **Grade Level:** 9 - 10
Objective: Use formulas to generate data points. Produce line graphs of which inferences are made.
Time: 20 minutes
Materials: TI-83 Plus/TI-84 Plus graphing calculator

Introduction

In "Hardball," an amateur mathematician uncovers an equation that uses baseball statistics to identify which players are using performance enhancing substances. By analyzing player performance, he is able to determine when an athlete began using a performance enhancing substance.

Sabermetrics is the statistical study of baseball through analysis of data. Through this study, it can be possible to isolate and reveal human performance. While sabermetrics is not an exact science, many major league baseball teams use it when evaluating players and making decisions about team needs and offensive strategies.

This activity uses a sabermetric statistic in trying to determine a change point in a player's data. A change point is the point in time when incoming data (e.g., home runs, batting average, on-base percentage, etc.) is vastly different from previously collected data.

Discuss with Students

The sabermetric statistic described in the *NUMB3RS* episode is fictitious, so this activity will focus on a sabermetric statistic called the CHR (Cumulative Home run Ratio) to determine if a change point occurs in the player's data.

In this activity, students will use their calculators to create and analyze graphs. For the best results, make sure your students use the calculator window settings *x*: [17, 42], *y*: [0.0, 0.1]

Student Page Answers:
1. *0* **2.** *0.03922* **3.** *See next page* **4.** *4 years* **5.** *Babe Ruth y = 0.085, Hank Aaron y = 0.06* **6.** *Babe Ruth* **7.** *While Mark McGwire's CHR has a plateau in the middle of his career, both his and Barry Bonds' CHR continue to climb in the later half of their careers while Hank Aaron's and Babe Ruth's plateaued.* **8.** *Age 24* **9.** *Age 31* **10.** *Barry Bonds at age 35, Babe Ruth at age 24 and 28.*

Patrick Flynn, Olathe East High School, Olathe, KS

Babe Ruth

AGE	AB	HR	CHR
19	10	0	0.00000
20	92	4	0.03922
21	136	3	0.02941
22	123	2	0.02493
23	317	11	0.02950
24	432	29	0.04414
25	457	54	0.06573
26	540	59	0.07689
27	406	35	0.07839
28	522	41	0.07842
29	529	46	0.07969
30	359	25	0.07877
31	495	47	0.08058
32	540	60	0.08390
33	536	54	0.08555
34	499	46	0.08610
35	518	49	0.08678
36	534	46	0.08673
37	457	41	0.08691
38	459	34	0.08617
39	365	22	0.08503
40	72	6	0.08502

Hank Aaron

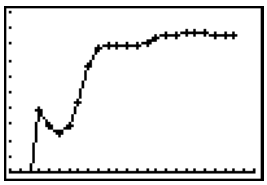
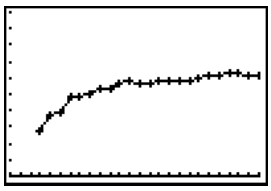
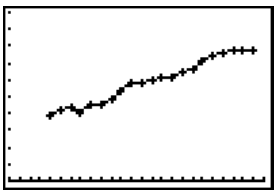
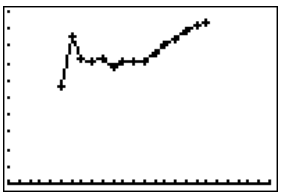
AGE	AB	HR	CHR
20	468	13	0.02778
21	602	27	0.03738
22	609	26	0.03931
23	615	44	0.04795
24	601	30	0.04836
25	629	39	0.05079
26	590	40	0.05323
27	603	34	0.05364
28	592	45	0.05613
29	631	44	0.05758
30	570	24	0.05622
31	570	32	0.05621
32	603	44	0.05753
33	600	39	0.05807
34	606	29	0.05737
35	547	44	0.05871
36	516	38	0.05949
37	495	47	0.06117
38	449	34	0.06177
39	392	40	0.06316
40	340	20	0.06304
41	465	12	0.06161
42	271	10	0.06106

Barry Bonds

AGE	AB	HR	CHR
21	413	16	0.03874
22	551	25	0.04253
23	538	24	0.04328
24	580	19	0.04035
25	519	33	0.04498
26	510	25	0.04564
27	473	34	0.04911
28	539	46	0.05384
29	391	37	0.05738
30	506	33	0.05817
31	517	42	0.06032
32	532	40	0.06162
33	552	37	0.06208
34	355	34	0.06379
35	480	49	0.06626
36	476	73	0.07148
37	403	46	0.07355
38	390	45	0.07542
39	373	45	0.07727
40	42	5	0.07746
41	367	26	0.07721

Mark McGwire

AGE	AB	HR	CHR
22	53	3	0.05660
23	557	49	0.08525
24	550	32	0.07241
25	490	33	0.07091
26	523	39	0.07179
27	483	22	0.06702
28	467	42	0.07045
29	84	9	0.07141
30	135	9	0.07121
31	317	39	0.07570
32	423	52	0.08060
33	540	58	0.08373
34	509	70	0.08907
35	521	65	0.09236
36	236	32	0.09409
37	299	29	0.09423

The data above is from **http://www.baseball-reference.com/players.shtml**.

Name: _____ Date: _____

NUMB3RS Activity: Is It for Real?

In "Hardball," an amateur mathematician uncovers a sabermetric equation that uses baseball statistics to identify which players are using performance enhancing substances. By analyzing player performance, he is able to determine when an athlete began using a performance enhancing substance. The point in time where the incoming data is vastly different from the previous data's trend is called a *change point*.

Sabermetrics is the study of baseball statistics and the effort of quantifying player and team strengths and weaknesses that go beyond the scoreboard. Some of the more interesting sabermetrics are: WHIP (Walks plus Hits per Inning Pitched), DIPS (Defense Independent Pitching Statistic), and VORP (Value Over Replacement player).

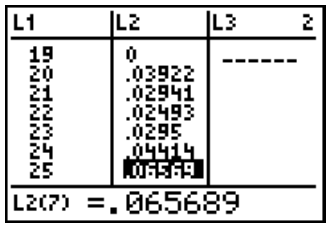
Consider one of the greatest home run hitters of all time, Babe Ruth. Why is he viewed as such a great home run hitter? To analyze his productivity, we will calculate his Cumulative Home run Ratio or CHR.

1. Using the attached data sheet, divide the number of home runs (HR) by the total number of at bats (AB) for Babe Ruth when he was 19. This is his CHR up to age 19.

2. Find Babe Ruth's CHR up to age 20 by dividing his total HR by his total AB for ages 19 and 20.

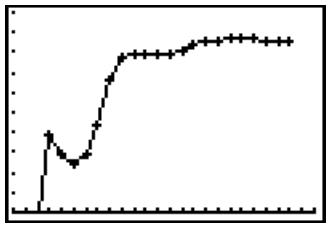
3. Compute Babe Ruth's CHR for ages 21 – 40.

To better analyze this data look at a line graph of the data.



Press [STAT] and select **1:Edit…** Enter Age in LB_1 and CHR in L_2.

Press [2nd] [STAT PLOT] and select **Plot 1**. Create a line graph using the settings shown above.



Press [GRAPH] to view the line graph.

4. Use the [TRACE] key to verify that Babe Ruth's CHR began to plateau at age 26. How long did this last?

Consider another great baseball player, Hank Aaron. Using the attached data table, compute his CHR along with the connected line graph using the same window settings as before.

Patrick Flynn, Olathe East High School, Olathe, KS

5. Did the graphs for both Hank Aaron and Babe Ruth reach a horizontal asymptote? If so, estimate their values.

6. Which player maintained a higher CHR?

Using the attached data table, compute the CHR, and generate line graphs for two modern-day baseball players, Barry Bonds and Mark McGwire.

7. Compare the CHRs for Barry Bonds and Mark McGwire to the CHRs for Hank Aaron and Babe Ruth.

8. The graph of Mark McGwire's CHR appears to have a plateau. When did it start?

A *change point* is a point in time where the incoming data is vastly different from the previous data's trend. Because the *NUMB3RS* episode focused on looking for performance enhancing substances, we will only consider change points that increased a player's CHR.

9. Does Mark McGwire's CHR graph have a change point? If so, when?

10. Did any of the other batters have a change point?

Patrick Flynn, Olathe East High School, Olathe, KS

Player Data sheet

Babe Ruth

Age	AB	HR
19	10	0
20	92	4
21	136	3
22	123	2
23	317	11
24	432	29
25	457	54
26	540	59
27	406	35
28	522	41
29	529	46
30	359	25
31	495	47
32	540	60
33	536	54
34	499	46
35	518	49
36	534	46
37	457	41
38	459	34
39	365	22
40	72	6

Hank Aaron

Age	AB	HR
20	468	13
21	602	27
22	609	26
23	615	44
24	601	30
25	629	39
26	590	40
27	603	34
28	592	45
29	631	44
30	570	24
31	570	32
32	603	44
33	600	39
34	606	29
35	547	44
36	516	38
37	495	47
38	449	34
39	392	40
40	340	20
41	465	12
42	271	10

Barry Bonds

Age	AB	HR
21	413	16
22	551	25
23	538	24
24	580	19
25	519	33
26	510	25
27	473	34
28	539	46
29	391	37
30	506	33
31	517	42
32	532	40
33	552	37
34	355	34
35	480	49
36	476	73
37	403	46
38	390	45
39	373	45
40	42	5
41	367	26

Mark McGwire

Age	AB	HR
22	53	3
23	557	49
24	550	32
25	490	33
26	523	39
27	483	22
28	467	42
29	84	9
30	135	9
31	317	39
32	423	52
33	540	58
34	509	70
35	521	65
36	236	32
37	299	29

The data above is from **http://www.baseball-reference.com/players.shtml**.

Patrick Flynn, Olathe East High School, Olathe, KS

The goal of this activity is to give your students a short and simple snapshot into a very extensive mathematical topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extension
Box Plots and Outliers

Introduction

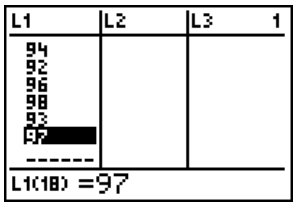
While change points identify when incoming data is vastly different from others in a data stream, how would one detect data that is vastly different from others after the data has been grouped (i.e., class scores on an exam)? In this type of situation, the data values that are very different from the rest are called outliers. An outlier is defined as a data point that is two standard deviations away from the norm, or 1.5 times the interquartile range above the upper quartile or below the lower quartile.

For example, suppose your class had the following exam scores:

90, 82, 92, 94, 83, 82, 95, 94, 96, 91, 86, 72, 94, 92, 96, 98, 93, 97

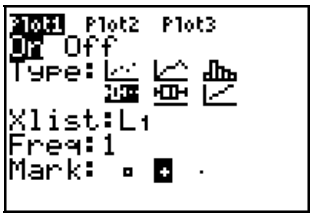
To detect the presence of an outlier, the calculator can be quickly used.

1.



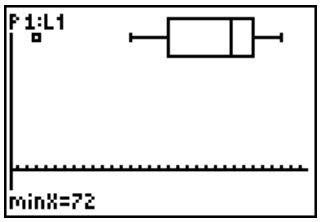
Press [STAT] and select **1:Edit…** Enter the test scores in L_1.

2.



Press [2nd] [STAT PLOT] and select **Plot 1**. Create a box plot with outliers using the settings shown above.

3.



Press [ZOOM] and select **9:ZoomStat** to view the box plot. The data point to the left is the outlier. Press [TRACE] to find its value.

To further explore the outlier with regards to the data, change some values and see what happens.

Additional Resources

- To explore other sabermetric statistics, see the *NUMB3RS* activity "The Pythagorean Expectation." This activity can be downloaded for free by going to **http://education.ti.com/exchange** and searching for "7662."

- Look up your own baseball data to analyze at the Web site **http://www.baseball-reference.com/players.shtml**

Patrick Flynn, Olathe East High School, Olathe, KS