

NUMB3RS Activity: Sabermetrics Episode: "Sacrifice"

Topic: Linear Modeling

Grade Level: 9 - 10

Objective: In this activity, students will look at two different sabermetric formulas (sabermetrics is a term that is used to describe the analysis of objective baseball statistics to quantify the performance of individuals and teams). Using the TI-Navigator™ system, students will compare sabermetric measures to actual Major League Baseball outcomes with the goal of assessing the validity of these measures. In particular, students will find that the formulas studied in this activity produce strong linear correlation with actual results, which will allow them to be represented mathematically with linear equations.

Time: 25 - 30 minutes

Materials: TI-83/TI-84 Plus graphing calculators, TI-Navigator system, and the following activity settings files: *Sabermetrics1.act*, *Sabermetrics2.act*, *Sabermetrics3.act*
To download this file, go to <http://education.ti.com/exchange> and search for "7713."

Introduction

In "Sacrifice," Charlie is trying to "reverse engineer" a homicide victim's computer in an effort to reconstruct the data that was erased from the computer. His initial inquiry reveals "magnetic shadows underneath...and pieces of data left behind." Charlie found that these numbers were baseball statistics that were based on an application of sabermetrics, a statistical process that places numerical values on specific skills like home runs, hits, and walks. Many Major League baseball teams use sabermetrics to identify the best way to allocate their resources. Later in the episode, it is discovered that the victim was actually writing an algorithm to track human potential, rather a person's ability to play baseball. The sabermetric numbers were left on the computer to cover up the victim's true objectives.

In this activity, students will calculate the *Pythagorean Expectation*, a sabermetric statistic that is used to predict the winning percentage of a specific team. Students will use a comparison of these estimates with actual results to find a line of best fit that models the data.

This activity also includes an extension, which guides students through the analysis of another sabermetric formula, *on-base plus slugging* (denoted by OPS).

Part I: What is Pythagorean Expectation?

- Students should work individually or in pairs during this part of the activity.
 - Launch TI-Navigator on the computer and press **Begin Class** to start the session.
 - Have each student (or pair of students) log into NavNet on their calculators.
- Load the ***Sabermetrics1.act*** activity settings file into Activity Center and click the 'Graph' tab. This will display a background image showing the number of runs scored, runs allowed, and win percentage for several Major League teams during the regular 2005 baseball season. It will also display the sabermetric formula, *Pythagorean Expectation*, shown below.

$$\text{Winning Percentage} = \frac{\text{Runs Scored}^2}{\text{Runs Scored}^2 + \text{Runs Allowed}^2} = \frac{1}{1 + (\text{Runs Allowed}/\text{Runs Scored})^2}$$

- b. Ask students if they can guess why this formula is called *Pythagorean Expectation*. Make sure that they realize that this name was used because of its resemblance to the Pythagorean Theorem, $a^2 + b^2 = c^2$. Mathematically, there is no connection between the Pythagorean Expectation formula and the Pythagorean Theorem. In fact, the formula was found through empirical observation and experimentation rather than mathematical derivation.
- c. Also explain to students the meaning of each list used in this activity:
- L₁ (has 30 rows because there are 30 major league baseball teams): Contains the number of runs scored for each baseball team in the 2005 season.
 - L₂ (30 rows): Contains the number of runs allowed for each baseball team in the 2005 season.
 - L₄ (30 rows): Contains the winning percentage for each baseball team in the 2005 season.

Each row represents a different team. For example, the sixth element of each row shown below represents actual 2005 results for the Chicago White Sox (741 runs scored, 645 runs allowed, and a win percentage of 0.611).

L1	L2	L3	1
910	805	-----	
886	789		
775	705		
729	800		
750	936		
741	645		
790	642		
L1(6)=741			

L2	L3	L4	4
805	-----	.586	
789		.586	
705		.494	
800		.457	
936		.414	
645		.611	
642		.574	
L4(6) = .611			

- d. To show how the formula works, launch **Quick Poll** (with *Open Response*), making sure that the *Numeric box* is checked. Click **Start Poll** and have students use the numbers in one of the rows shown in the Activity Center window to calculate a specific team's *Pythagorean Expectation*. For example, students may choose to calculate the *Pythagorean Expectation* for the Boston Red Sox, which results in a value of 0.5609... Make sure that you compare these results with the actual winning percentages shown in Activity Center. For example, the Boston Red Sox had a winning percentage of 0.586 in 2005, which is very close to the value obtained with the *Pythagorean Expectation* formula. You can also uncheck the *Numeric box* to reveal each student's calculations.
- e. Press **Stop Poll** and close the **Quick Poll** window.

Part II: Transfer the Data to Student Calculators

3. a. Press **Start Activity** and instruct each student to press 1:ACTIVITY CENTER. Each student will receive the data contained in lists L₁, L₂, and L₄.
- b. Press **Stop Activity**. Students should then exit NavNet by pressing **2nd** [QUIT] to return to TI-Navigator Home and selecting 4:EXIT APP.

4. a. Tell students to calculate the *Pythagorean Expectation* for every team by executing the command below which places the results in L₃.

```
1/(1+(L2/L1)^2)→L
3
```

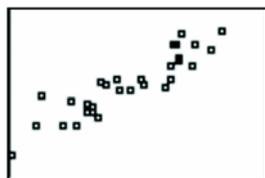
- b. Ask students to press **STAT** to view the lists on their calculators. Use **Screen Capture** to view the results, using the *Zoom* feature to focus on individual screens. For example, the highlighted number in the screen below relates to the Chicago White Sox, whose Pythagorean Expectation was 0.56893 (L₃) versus an actual winning percentage of 0.611 (L₄).

L2	L3	L4	3
805	.561	.586	
789	.55772	.586	
705	.54719	.494	
800	.45366	.457	
936	.39101	.414	
645	.63998	.611	
642	.60226	.574	
L3(6) = .568933360...			

- c. Have each student (or pair of students) log back into NavNet on their calculators.

Part III: Find a Linear Model

5. a. In this step, students will view a scatter plot in Activity Center of the *Pythagorean Expectation* versus actual results.
- b. Load the **Sabermetrics2.act** activity settings file into Activity Center and click the 'List-Graph' tab.
- c. Press **Start Activity**, have *one* student enter Activity Center, and **Send** lists L₃ and L₄ into Activity Center.
- d. A scatter plot of the data, similar to the one shown below, will appear in the Activity Center window showing a strong positive linear correlation between the *Pythagorean Expectation* (L₃) and actual results (L₄).



- e. Press **Stop Activity**.

6. a. Change *Contribute to Equations* and press **▶ Start Activity**. This will allow students to send a linear equation to model the data. Students will be able to view a scatter plot of the data and their equations on their calculators (by pressing **GRAPH**) before they **SEND** the equation to Activity Center. Students will also have the opportunity to resubmit their equation. Students may select two points from the data set shown in Activity Center to find their linear model or use a trial-and-error approach. (**Note:** The *least squares* equation for this data set is $y = 0.89x + 0.05$.)
- b. Press **■ Stop Activity** after all the equations have been sent into Activity Center.
- c. Click the 'Graph – Equation' tab and use the 'Show' and 'Hide' feature to analyze the results for individual students. Ask students several questions about their linear models including:
- What do you think the linear model would have been if the *Pythagorean Expectation* formula worked perfectly?
 - Do you think the *Pythagorean Expectation* formula is valid?
 - Even though the *Pythagorean Expectation* results are not exactly equal to actual winning percentages, can you still use it to rank order the teams?
 - Using your linear model, what do you expect a team's winning percentage will be with a *Pythagorean Expectation* of 0.575?
 - Using your linear model, what do you expect a team's *Pythagorean Expectation* will be with an actual winning percentage of 0.524?
 - The Arizona Diamondbacks have a *Pythagorean Expectation* value of 0.398 versus an actual winning percentage of 0.475. Can you explain why there is such a difference?
 - Suppose you want to predict next year's winning percentage for your favorite baseball team using the *Pythagorean Expectation* formula. What information will you need? How can you obtain this data?

The goal of this activity is to give your students a short and simple snapshot into a very extensive math 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

- Assign each student one major league baseball team and ask them to calculate the sabermetric statistic *on-base plus slugging* (denoted by OPS) for the nine batters on the team that had the most at-bats during the 2005 regular season. This statistic is found by adding a player's on base percentage (OBP) and slugging average (SLG). Students can also use the formula

$$OPS = \frac{H + BB + HBP}{AB + SF + BB + HBP} + \frac{TB}{AB}$$

where H = hits, BB = base on balls, HBP = hit by pitch, AB = at bats, SF = sacrifice fly, and TB = total bases.

(**Note:** Two good sources for this data are www.mlb.com and www.espn.com.)

- Students should find the *sum* of the OPS statistics found for the nine players.
- Students should also find the average number of runs per game for their team during the 2005 regular season.
- Load the ***Sabermetrics3.act*** activity settings file into Activity Center and click the 'List-Graph' tab. Students will submit the sum of the OPS statistics into Activity Center as L_1 and the average number of runs per game as L_2 .
- A graph of L_1 and L_2 should reveal strong positive correlation, that is, teams with players that have high OPS numbers will score more runs per game.
- Students will derive a linear model, either by hand or using linear regression.
- Ask a variety of questions, similar to those posed in Part III of this activity.

Additional Resources

- An additional NUMB3RS activity for "Sacrifice" can be downloaded for free by going to <http://education.ti.com/exchange> and searching for "7712."
- The NUMB3RS episode "Hardball" from season 3 also includes discussions about sabermetrics. Two NUMB3RS activities for this episode that involve sabermetrics can be downloaded for free by going to <http://education.ti.com/exchange> and searching for "7662" and "7664."
- If you would like to learn more about TI-Navigator™, visit <http://education.ti.com/navigator>.