NUMB3RS Activity: Logging Witnesses
Episode: "All's Fair"

Topic: Probability and logistic regression **Grade Level:** 11 - 12

Objective: Use logistic regression to determine the probability of an event.

Time: 30 minutes

Materials: TI-83/84 Plus calculator

Introduction

In the episode "All's Fair," Charlie uses a logistic regression to help determine which suspect might be most likely to strike next. In this activity, students create a logistic regression based on a set of data to determine the probability that a suspect is either male or female, given the suspect's height.

In this activity, students will determine, given a height, if a suspect is more likely to be a male (1) or female (0). Students will determine the experimental probability based on a set of data, use a TI-83/84 Plus calculator to determine a regression equation, and use a logistic regression equation to predict results. Specifically, the students will predict from the logistic regression equation the probability of a person being male based on the person's height.

The method suggested here is similar to a method used by statisticians. However, statisticians might use the logarithm of the odds of the outcome (called the log odds) to linearize the data, and then apply a statistical program to determine a linear regression based on the log odds. (Log odds are the result of taking the natural logarithm of the odds for an event happening.) The linear regression function is then used to predict the probability of being a male. This method is examined in an extension of the activity.

Discuss with Students

1. Suppose you were told that in a group of students, 3 males and 77 females, each has a height of 62 in. How would you determine the experimental probability that a person chosen at random from this group was male?

2. The following logistic regression equation finds the probability that a person is a male based on a given height in inches.

$$\text{probability of a male} = \frac{0.62}{1 + 6 \times 10^{30} e^{-1.1 \times \text{person's height}}},$$

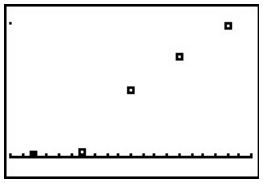
Using the equation, what is the probability that a person from the group in question 1 is a male?

3. If you were given a graph of a set of data and asked to find a function that modeled the data, describe the process you could use to find the function. [To the teacher: students should have some basic understanding of regression. If not, use this opportunity to explain regression.]

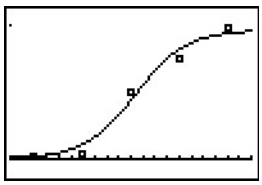
Terry Souhrada, University of Montana, Missoula, MT

Discuss with Students Answers:
1. *3/80 or approximately 0.038 = 3.8%* **2**. *approximately 0.04 = 4%* **3.** *Sample answer: You could examine the shape of the graph to determine what type of model would best fit the data. Then use a calculator to determine the function using the regression feature of the calculator.*

Student Page Answers:
1. *$P(60) \approx 0.0175$, $P(62) \approx 0.0357$, $P(68) = 0.5$, $P(72) = 0.75$, $P(76) \approx 0.9833$* **2.** *Sample graph is given below.*



3a. $P(x) = \dfrac{0.967}{1 + 1.342 \cdot 10^{14} \cdot e^{-0.476x}}$ **3b.** *$a \approx 1.342 \times 10^{14}$, $b \approx 0.476$, $c \approx 0.967$; Sample graph is given below.*



4a. *$P(63) \approx 0.069$ or 0.071, $P(67) \approx 0.330$ or 0.334, $P(73) \approx 0.870$ or 0.872 (actual logistic regression equation or rounded logistic regression equation).* **4b.** *The second witness is the most likely telling the truth since the probabilities would indicate that it is most likely that the suspects consisted of two women and one man.*

Terry Souhrada, University of Montana, Missoula, MT

Name: _____ Date: _____

NUMB3RS Activity: Logging Witnesses

Agent Don Eppes is attempting to solve a crime in which the testimony of some witnesses may be false. Several witnesses have all indicated the same approximate height for the three suspects, but their testimonies do not agree on whether the suspects were male or female. Don asks Charlie if there is some mathematical way to help determine which, if any, of the witness testimonies are accurate. Charlie suggests using probability and a regression model to help make the determination.

1. Charlie gathered the height information given in the table below. For each height given, find the experimental probability that a person of that height would be a male.

Height (inches)	# of Males	# of Females	Probability Person is Male at Each Given Height
60	1	56	
64	2	54	
68	31	31	
72	33	11	
76	59	1	

2. Create a scatterplot by plotting the heights on the horizontal axis and the probabilities for the person being a male on the vertical axis.

Charlie knows that he cannot use a linear regression to model the probabilities because as the height of the person increases, the linear model would ultimately predict that the probability would be greater than 1, which isn't possible. He knows that a logistic function has an upper and lower limit much like probabilities, so Charlie decides to use a **logistic regression equation** as his model. A logistic regression equation is based on the natural logarithm whose base is the irrational number *e* (approximately 2.71828). Because *e* is irrational, its decimal representation neither terminates nor repeats.

Terry Souhrada, University of Montana, Missoula, MT

3. The graph you created in question 2 has a shape that looks like the graph of a logistic function. The general equation of a logistic function, as given on a TI-83/84 Plus calculator, is

$$P(x) = \frac{c}{1 + a \cdot e^{-bx}}$$

a. Use a TI-83/84 Plus calculator to determine the logistic regression equation for the graph you created in question 2. Using the table in question 1, enter the values from the "Height" column into list L_1, and enter the values you found in the "Probability Person is Male at Each Given Height" column into list L_2. To find the logistic regression equation, use the command **Logistic L_1,L_2**. (To find **Logistic**, press STAT, go to the **CALC** menu, and select **B:Logistic**.)

b. Identify the values of *a*, *b*, and *c* in the logistic regression equation and graph this equation on the same coordinate system as the graph created in question 2.

4. The first witness stated that the heights of the three suspects were about 63 inches, 67 inches, and 73 inches. The first witness said the three suspects were definitely all women. The second witness gave approximately the same heights but said there were two women and one man. The third witness said that all of the suspects were men.

a. Using the logistic regression model, determine the probability that each of the suspects was male.

b. Based on these results, which witness do you think is telling the truth? Justify your answer.

Terry Souhrada, University of Montana, Missoula, MT

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.

Extensions

Extension 1

The values for *a*, *b*, and *c* that you found in question 3 of the activity transform the shape of the basic logistic equation $P(x) = \frac{1}{1 + e^{-x}}$.

1. Describe how each of the values for *a*, *b*, and *c* affected the graph of the equation.

2. A second form of the basic logistic equation is $P(x) = \frac{e^x}{1 + e^x}$. Explain why the two forms of the basic logistic equation are equivalent.

Extension 2

When statisticians use a logistic regression, they use a method somewhat different than the one outlined in the activity. They first determine the odds of the event happening rather than the probability. They then take the natural logarithm of these odds and from that determine a linear equation to model the probabilities. Using the Web sites below, investigate this process and report on the differences and similarities you find between what you did in this activity and what a statistician might actually do.

http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html

http://faculty.vassar.edu/lowry/logreg1.html

http://www.upa.pdx.edu/IOA/newsom/pa551/lectur21.htm

Terry Souhrada, University of Montana, Missoula, MT