NUMB3RS Activity: The Eyes Have It
Episode: "The Running Man"

Topic: Probability **Grade Level:** 10 - 12

Objective: Students use probabilities to compare the effectiveness of two identification methods: iris scanning and facial recognition.

Time: 20 - 30 minutes

Introduction

In "The Running Man," Charlie is asked to help the FBI track criminals who have successfully bypassed an iris scanning security system. In this activity, students will use probability to examine why an iris scan is a more effective method for identification than a simple "mug shot."

In the episode, a security camera has taken a photograph showing a partial front view of the face of one of the criminals. Don wants to create a sketch of the criminal's face to help identify the suspect. Charlie suggests that iris recognition using the partial photograph might be a better choice. He uses probabilities to persuade Don that iris recognition greatly reduces the chance of misidentifying the suspect.

Iris recognition is based on identifying patterns on the iris (colored part) of a person's eye. The most widely used iris recognition algorithm creates coordinates on a complex plane based on these patterns. It collects and plots 2,048 "phase bits" (256 bytes) of coded information, creating a plot that is virtually unique for any recorded iris. This creates a setting in which the comparison of two iris patterns has an extremely high level of reliability in the identification process.

Iris patterns found in a suspect's eye are compared to those found in a database, much like a fingerprint analysis. The two are statistically checked for independence. If the patterns are found to be independent, then the irises do not match. In this activity students compare the probability of falsely identifying a suspect based on a "mug shot" versus iris recognition. In the final part of the activity, students determine the necessary probability to guarantee a given acceptable percentage of false identification.

False identification is similar to a false positive on a drug test. False positive test results were the basis for the activity "How Reliable is the Test," which appeared with the *NUMB3RS* episode "Calculated Risk."

Students should have experience with probability, exponents, and equations containing unknown exponents. Students should realize that if the probability of an event occurring is p, then the probability of the event *not* occurring is $(1 - p)$. Students will also need to know that when computing the probability of multiple events, the events "none" and "at least one" are complementary events; that is if the probability of "none" is p, then the probability of "at least one" is $(1 - p)$. Students should be able to evaluate expressions involving exponents and scientific notation. In the final portion of the activity, students should be able to solve equations involving unknown exponents.

If students do not have experience with solving equations with unknown exponents, you might wish to have them skip the last three questions in the activity.

© 2006 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

Discuss with Students

In "The Running Man," iris recognition is used to identify a possible criminal. Using iris recognition gives an extremely high degree of certainty in correctly identifying a person.

1. Suppose the iris of the eye can be used to identify a person correctly 92 times out of 100.

a. What is the probability that the person would be correctly identified?
b. What is the probability the person would be incorrectly identified?

2. In general, if the probability of correctly identifying someone by an iris scan is *n*, what is the probability that a mistake would be made?

3. If the probability of making a mistake is 10^{-3}, the probability of making two such errors in a row is $(10^{-3})^2$. What is the probability of making a mistake two times in a row, written as a percent?

4. If $10 = 5^x$, how could you find an approximate value for *x*?

Discuss with Students Answers: 1a. $\frac{92}{100}$ **1b.** $\frac{8}{100}$ **2.** $(1 - n)$ **3.** 0.0001% **4.** *Sample answer: You could use logarithms or guess and check using a calculator. In this case,* $x = \log_5 10 = \frac{\log 10}{\log 5}$.

Student Page Answers: 1. $(1 - 0.999) = 0.001 = 0.1\%$ **2.** $(0.999)^{100} \approx 0.905 = 90.5\%$
3. $1 - (0.999)^{100} \approx 1 - 0.905 = 9.5\%$ **4.** $1 - (0.999)^{3,000} \approx 0.950 = 95.0\%$

5. $\frac{1}{690,000} \approx 1.45 \times 10^{-6}$ **6.** $1 - \left(\frac{689,999}{690,000}\right)^{3000} \approx 0.004 = 0.4\%$ **7.** $P = 1 - (S)^n$ **8.** *If the HD value is 0.35, the equation* $0.001 = 1 - \left(\frac{132,999}{133,000}\right)^n$ *results in* $n \approx 133$ *people.* **9.** *Solving the equation* $0.0001 = 1 - (S)^{10,000}$ *for S results in S = 0.99999999. This means an HD value of 0.31 or less is acceptable.*

© 2006 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

Name: _____Date: _____

NUMB3RS Activity: The Eyes Have It

The FBI is investigating a robbery at the research facility of a large company. The FBI believes that the robbers were employees of the company, because they were able to bypass the iris recognition security scan and gain entrance. However, a security camera was able to capture a partial image of one of the suspect's faces. Don wants to use the partial photograph to create a "mug shot" sketch. The sketch could then be compared to a database of pictures of all the company's employees.

Charlie relates research that "mug shots" taken at least one year apart can create error rates of 43–50% in correctly matching photos of the same person. To further explain his concern, Charlie asks Don to suppose that the probability of correctly identifying the suspect using a "mug shot" was as high as 99.9%.

1. Given Charlie's suggested 99.9% probability, what is the probability of making a false match (incorrect comparison) of the sketch with *one* employee's company photograph?

2. If this sketch is compared to *n* employee photographs, the probability of making a correct comparison is the probability of making a correct comparison raised to the *n*th power. What is the probability of making a correct match if there are 100 employees in the company?

3. Based on your answer to Question 2, what is the probability of making at least one false match using the sketch?

4. Charlie points out that the company has roughly 3,000 photographed employees. What is the probability that there will be at least one incorrect match in comparing the sketch to all of the employees' photographs?

Charlie suggests that an analysis of the iris (colored part) of the suspect's eye be compared to the iris scans the company requires for each of its employees. He explains that an iris scan creates a graph based on over 2,048 pieces of data collected in the scan. This graph is used to generate a pattern for each iris. Once the graph of an iris scan is complete, it is then possible to measure the dissimilarity of any two irises and to compute a probability of correct identification, using a measure known as the Hamming Distance (HD). The HD is the measure of dissimilarity in any two irises. The HD varies by the number of pixels involved in the analysis of the iris. Once a given Hamming Distance has been computed, the probability of correct identification can be found from the chart shown on the next page.

© 2006 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

For example, if the HD value is 0.34, the probability that the person will be correctly identified is $\frac{689,999}{690,000}$. Use an HD value of 0.34 to answer questions 5 and 6.

5. What is the probability of making a false match when comparing the suspect's iris pattern to one randomly selected iris scan of the employees?

6. What is the probability of making at least one false match when comparing the suspect's iris pattern to all 3,000 employee iris scans?

7. Write an equation that could be used to determine the probability *P* of making at least one false match when comparing *n* employee scans to the suspect's iris patterns given that the probability of making a single correct comparison is *S*.

Charlie further explains that as the HD value is reduced, the probability of making a correct identification in a single comparison increases. This is shown in the table below.

Probability for Various HD Values

HD value	Probability of Correct Identification
0.31	$\frac{184,999,999}{185,000,000}$
0.32	$\frac{25,999,999}{26,000,000}$
0.33	$\frac{3,999,999}{4,000,000}$
0.34	$\frac{689,999}{690,000}$
0.35	$\frac{132,999}{133,000}$

Source: Adapted from *How Iris Recognition Works*, John Daugman, p. 27

8. Assume the HD value is 0.35. Use your equation from Question 7 to determine the number of employees n you could compare to the suspect's data if you wanted to guarantee a probability of less than 0.1% of making at least one false match.

9. Suppose the company has 10,000 employees. From the table, what is the greatest HD value you could use and still guarantee at most a 0.01% probability of making a false match when comparing the suspect's data to all the iris scans of the employees? (Hint: One method for finding the solution using a TI-84 graphing calculator is to write an appropriate equation and graph each side of the equation on the same coordinate system. Then locate the *x*-coordinate of the intersection.)

© 2006 Texas Instruments Incorporated
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

- Incorrectly identifying a person based on their iris scan is much like a false positive for a drug test. A false positive would incorrectly identify the presence of a drug in someone's system when that person had not taken the drug. An earlier *NUMB3RS* activity entitled "How Reliable is the Test," which appeared with the episode "Calculated Risk," examines drug tests and false positives. This activity shows how conditional probability and tree diagrams can be used to help determine the chances of a false positive occurring. This activity can be downloaded from the Web site below.

  http://education.ti.com/educationportal/activityexchange/activity_detail.do?cid=us&activityid=6574

  10-14_Act1_HowReliable_CalculatedRisk_final.pdf

- There are many other ways that biological information is used to identify people. One of the most familiar identification methods is fingerprinting. Use the Web sites below to research fingerprinting methods. Then write a report comparing the process of fingerprint recognition to other forms of biological recognition methods.

  http://biometrics.cse.msu.edu/fingerprint.html

  http://www.policensw.com/info/fingerprints/finger08.html

  http://www.szabist.edu.pk/NCET2004/Docs/Session%20VII%20Paper%20No%202%20(P%20141-146).pdf

  http://eprint.iacr.org/2004/021.pdf

Reference

Daugman, John. "How Iris Recognition Works." *IEEE Transactions on Circuits and Systems for Video Technology*. 14 (January 2004): 21-30.
http://www.cl.cam.ac.uk/users/jgd1000/csvt.pdf

© 2006 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT