NUMB3RS Activity: We're Number 1!
Episode: "The Running Man"

Topic: Benford's Law **Grade Level:** 9 - 12
Objective: Explore Benford's Law
Time: 20 - 30 minutes

Introduction

In "The Running Man," Charlie discusses how Benford's Law can be used to identify a fabricated set of data. Benford's Law describes the surprising fact that in many naturally occurring data sets described by a geometric progression, the leading digit is more likely to be 1 or 2 than any other number. (The leading digit of a number is the leftmost digit of a number, and cannot be zero.) In fact, approximately 30% of the time, the leading digit will be 1.

This phenomenon was first discovered in 1881 by astronomer Simon Newcomb, when he discovered that the beginning pages of his book of logarithms were more worn than the others. Dr. Frank Benford, a physicist at the General Electric Company rediscovered this in 1938 and hypothesized that numbers with low leading digits are more prevalent than others. Over a period of six years, he analyzed over 20,000 values, including such diverse data as scientific constants, stock values, lake depths and numbers occurring in newsprint. When his research supported his hypothesis, he developed the following formula for the probability that the leading digit of a number is *d*:

$$p(d) = \log\left(1 + \frac{1}{d}\right)$$

Some restrictions apply on the use of Benford's law. The set of numbers cannot be assigned values (such as a social security or phone numbers), cannot have a uniform distribution of digits (such as lottery numbers), and should not have a preset maximum or minimum. Because Benford's work assumes that the data can be described by a geometric progression, two common applications are approximating population growth and financial information. However, other types of data sets such as house numbers and Fibonacci numbers have shown the same surprising results.

Many people assume that numbers will have a uniform and random distribution, and that all digits are equally as likely to be the leading digit of a number. That is, most people would not consider Benford's Law when falsifying documents. Because of this, the IRS uses Benford's Law to help determine if the numbers on a tax form need further review.

© 2006 Texas Instruments Incorporated
Kathy Erickson, Monument Mountain Regional High School, Great Barrington, MA

Discuss with Students

1. You are assigned to compile a list of 100 numbers that have been printed in newspapers during the week. Do you think that some numbers will appear more often than others as the leading digit?

2. Suppose half of the students in your class were told to make up a list of 100 values. If you were one of the students assigned to make up values, but don't want to be found out, how could you make your list believable?

3. In this activity, you will generate a hypothetical set of data for the cost of gas. If gas costs $1.00 at time $t = 0$ and we assume 5% yearly inflation, write a function $c(t)$ that can be used to find the cost of gas in t years.

Discuss with Students Answers:
1. *Answers will vary. Students will examine this question in the activity.* **2.** *Answers will vary. Discuss with students the idea that the leading digits of numbers are not equally distributed.*
3. $c(t)=e^{0.05t}$

Student Page Answers:
1a.*$1.11; $1.28* **1b**. $t = 14$
1c.

Cost (in dollars per gallon)	$2.00	$3.00	$4.00	$5.00	$6.00	$7.00	$8.00	$9.00	$10.00
***t* (in years)**	14	22	28	33	36	39	42	44	47

1d–1e.

Cost Range	$1.00–$1.99	$2.00–$2.99	$3.00–$3.99	$4.00–$4.99	$5.00–$5.99	$6.00–$6.99	$7.00–$7.99	$8.00–$8.99	$9.00–$9.99
Number of years	14	8	6	5	3	3	3	2	3
% of time	30%	17%	13%	11%	6%	6%	6%	4%	6%

1f. *For the cost to increase from $1.00 to $2.00, the cost must increase by 100%. However, to increase from $2.00 to $3.00, the cost only increases by 50%. Similarly, the percentage increase continues to go down as the cost in dollars increases. So, the prices spend the longest amount of time in the $1.00–$1.99 range.*

2a.

d	1	2	3	4	5	6	7	8	9
***p*(*d*)**	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

2b. *The values in the two charts are roughly the same. The differences arise because time values were rounded to the closest year in Question 1. If that time period were smaller, the results would be closer to those from Question 2.* **3**. *Larry's data set does not follow Benford's Law, so it may have been fabricated.*

 © 2006 Texas Instruments Incorporated
Kathy Erickson, Monument Mountain Regional High School, Great Barrington, MA

Name: ___________________________________ Date: _________________

NUMB3RS Activity: We're Number 1!

In "The Running Man," Charlie discusses a law called Benford's Law, which can be used to identify falsified information by analyzing the leading digits of numbers in a data set. The leading digit of a number is the leftmost digit; for example, the leading digit of 392 is 3 and the leading digit of 7,684 is 7. Many people think that in a large set of numbers, the leading digits are fairly evenly distributed (that is, there will be roughly the same number of 1s, 2s, 3s, etc. as leading digits). However, Benford's Law describes the surprising fact that in many naturally occurring data sets, the leading digit is more likely to be 1 or 2 than any other number.

To introduce Benford's Law, we will look at an example involving the price of gasoline.

1. Assume that at some point in time (defined as $t = 0$), gas was priced at $1.00 per gallon. For simplicity, use a yearly inflation rate of 5%. The cost per gallon of gas in t years is given by the formula $c(t) = e^{0.05t}$.

a. What will the cost of gasoline per gallon be in 2 years? In 5 years?

b. Determine the first year that the cost per gallon will be at least $2.00.

c. Complete the chart below to find the number of years t it will take for the cost per gallon to exceed each dollar amount.

Cost (in dollars per gallon)	$2.00	$3.00	$4.00	$5.00	$6.00	$7.00	$8.00	$9.00	$10.00
***t* (in years)**	14	22							

d. Determine the number of years that the price per gallon will be within the ranges shown in the table. Use this data to find the total number of years that the cost per gallon will be less than $10.00.

Cost Range	$1.00–$1.99	$2.00–$2.99	$3.00–$3.99	$4.00–$4.99	$5.00–$5.99	$6.00–$6.99	$7.00–$7.99	$8.00–$8.99	$9.00–$9.99
Number of years	14	8							
% of time	30%								

e. The table now shows information about gasoline costs for 47 years. Find the percent of time, rounded to the nearest whole number, that the cost of gasoline was in each cost range.

f. Why is the percent of the time for the $1.00–$1.99 range more than any other range? (Hint: Think of the dollar amount of the cost increase.)

© 2006 Texas Instruments Incorporated
Kathy Erickson, Monument Mountain Regional High School, Great Barrington, MA

2. Benford developed the following formula for the probability that the leading digit of a number is *d*:

$$p(d) = \log\left(1 + \frac{1}{d}\right)$$

a. Use the formula to complete the chart below. Round to the nearest thousandth.

d	1	2	3	4	5	6	7	8	9
***p*(*d*)**									

b. How do these probabilities compare to your answers to Question 1d? Explain why there are differences in your answers.

3. Suppose that Charlie has data sets submitted by three different people, and he suspects that one person has fabricated his data. After counting the occurrences of leading digits in each number of each person's data set, Charlie records the results in the table below. Use Benford's Law to determine which set of data is most likely to have been fabricated.

Leading Digit	**Number of Occurrences**		
	Larry	**Carl**	**Mark**
1	320	390	452
2	192	230	264
3	176	164	185
4	160	120	147
5	320	103	117
6	102	86	102
7	100	75	87
8	110	66	78
9	120	60	68

© 2006 Texas Instruments Incorporated
Kathy Erickson, Monument Mountain Regional High School, Great Barrington, MA

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

For the Student

- Benford's Law applies to Fibonacci numbers. Write a program or use a spreadsheet to analyze the first digit frequencies for the first *n* Fibonacci numbers.

- Just for fun, and to help convince yourself that Benford's Law is true, collect your own set of data and analyze the frequencies of the leading digits. Remember that there are restrictions to the types of data you should use – the set of numbers cannot be assigned values (such as a social security or phone numbers), cannot by definition have a uniform distribution of digits (such as lottery numbers), and should not have a preset maximum or minimum.

- Benford's work assumes that the data can be described by a geometric progression. Imagine a graph of such data, where the *x* is the occurrence and *y* is the observed value. A set of data forms a geometric curve from the points (1, 10) to (1000, 100). Write an equation for the function and develop the first digit frequencies, which should agree with Benford's Law.

- Use your calculator to draw a graph of $p(d) = \log\left(1 + \frac{1}{d}\right)$. Explain how the shape of the graph describes what you have already learned about the probabilities of the leading digits of numbers.

Additional Resources

Mark J. Nigrini, PhD, of Saint Michael's College, has done much work with Benford's Law. His Web site includes links to many papers and lectures.
http://www.nigrini.com/Benford's_law.htm

A description of the law and an introduction to the calculus and statistics behind the formula can be found at **http://mathworld.wolfram.com/BenfordsLaw.html**.

Another resource for the mathematics behind the formula can be found at **http://www.mathpages.com/home/kmath302/kmath302.htm**.

This Web site contains an article about Benford's Law that appeared in the New York Times in August 1998.
http://www.math.yorku.ca/Who/Faculty/Brettler/bc_98/benford.html

© 2006 Texas Instruments Incorporated
Kathy Erickson, Monument Mountain Regional High School, Great Barrington, MA