NUMB3RS Activity: Fours and Sevens, Oh My!
Episode: "Structural Corruption"

Topic: Using histograms to analyze data   **Grade Level:** 9 - 12
Objective: Students will compare outcomes produced by chance to outcomes that could
  be caused by some other factor
Time: 20 minutes
Materials: TI-83 Plus/TI-84 Plus graphing calculator with Probability Simulation App
  installed

Introduction
While examining data from the contractor who poured the foundation for the Cole
Building, Charlie notices that the overtime hours and employee numbers contain an
inordinate number of 4s and 7s; it does not seem that this is a random occurrence. This
leads the FBI to determine that the data were fabricated and that there seemed to be
undocumented workers doing the welding in the foundation. These welds were then
determined to be substandard and resulted in the instability of the Cole building under
high winds.

This activity makes use of the Probability Simulation App to generate large sets of
random and biased data. This App can be downloaded from the Web site below:
**http://education.ti.com/educationportal/sites/US/productDetail/
us_prob_sim_83_84.html**

Discuss with Students
Individuals who spend considerable amounts of time looking at data are able to spot
irregularities as Charlie did in this episode. However, most people need computer
programs to detect them. Before using this activity, students should have a basic
understanding of what random numbers are and how they are generated. Calculators
generate random numbers using a random number list. The list begins from a seed
value (the default setting for all calculators). If students do not change the seed value,
they will all generate the same list of numbers. Have students generate random numbers
on their calculators by entering the **rand** command (press $\boxed{\text{MATH}}$, go to the **PRB** menu,
and select **1:rand**). If this is the first time the random number command has been used
with the calculator, each calculator will output the same random number. Students will
continue to generate identical random numbers until each calculator is set with a
different seed value. The first step in the activity reseeds the random number generator.
(Setting new seeds is like dropping a pencil on a random number table and reading off
random numbers from the point at which the pencil landed.)

Student Page Answers:

1a. *Around 31 times* $\left(\dfrac{250}{8}\right)$ **1b.** *based on the data, they do not seem equally likely, but because they are*

*random numbers – anything could happen. For example, 4 would be expected to occur close to 31 times, in
the simulated data it occurs more.* **1c.** *The histograms are different, because they are graphs of random
numbers (Note: in the long run all graphs would be approximately the same.)* **2a.** *While the unweighted
graph has no real pattern, the weighted graph contains a definite pattern of the two highest bars being 4 and
7.* **2b.** *Other histograms also have the highest bars as four and seven.* **3.** *The occurrence of each value
would be identical.* **4a.** *Theoretically*, *it would occur two times, or twice as often; in the simulated data it
might not be exactly twice.* **4b.** *Theoretically, they would occur with the same frequency, although this might
not be what exactly occurs in the simulated data.* **4c.** *Giving sections 4 and 7 weights that were twice those
of the other numbers caused them to appear approximately twice as often.* **5.** *Answers will vary, but there
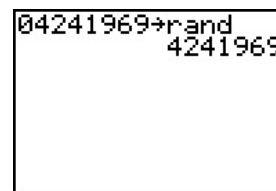will likely be bias.*

Name: _____ Date: _____

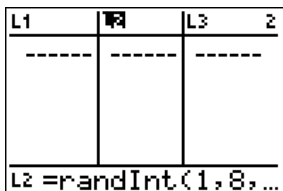# *NUMB3RS* Activity: Fours and Sevens, Oh My!

While examining data from the contractor who poured the foundation for the Cole Building, Charlie notices that the overtime hours and employee numbers seem to contain an inordinate number of 4s and 7s; it does not seem that this is a random occurrence. This leads the FBI to determine that the data were fabricated and that there were undocumented workers doing the welding in the foundation. These welds were then determined to be substandard, which resulted in the instability of the Cole building under high winds.

This activity will examine sets of random and biased data by generating a "random" list of values that contains more 4s and 7s than you might expect. This can be done efficiently on a calculator, but first, the random number generator on the calculator needs to be set to a random starting point. On your calculator, enter your date of birth (MMDDYEAR) and store it as **rand**. For example, to store April 24, 1969 enter **04241969** [STO▶] **rand**. (To find the **rand** command, press [MATH], go to the **PRB** menu, and select **1:rand**.) This will reset the random number generator on your calculator so that (barring any common birthdays), everyone in the class will generate a different set of random numbers.
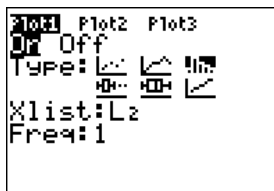
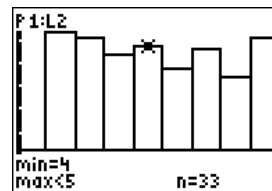Now, follow the steps below to generate your random set of numbers from 1 to 8.

1.

Press [STAT] and choose **Edit…**To generate the list of values from 1 to 8 where all values are equally likely to occur, move the cursor to highlight L₂, and enter the command **randInt(1,8,251)**. Then press [ENTER]. This will create a list of 251 random integers from 1 to 8.

2.

To generate a histogram for the random numbers, press [2nd] [STAT PLOT] and choose **Plot 1**. Then use the settings shown above. To view the Histogram, press [ZOOM] and select **9:ZoomStat**.

3.

To make the bars span units of 1, press [WINDOW] and set Xmin=0, Xmax=9, and Xscl=1.

1. Look at the data of numbers from 1 to 8 where all numbers are equally likely.

   a. About how many times would the value of 4 be expected to occur in 250 trials if all outcomes were equally likely?

   b. Make a list of the frequencies for each of the numbers 1 to 8. Do all of the outcomes 1 to 8 seem equally likely? Why or why not?

   c. Compare your frequency table and histogram to others around you. What do you notice and why?
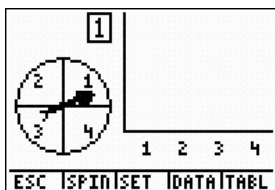
Use the steps below to generate a random set of numbers from 1 to 8, where you have weighted the choices so that some numbers (4s and 7s) will occur more often than the others.

1

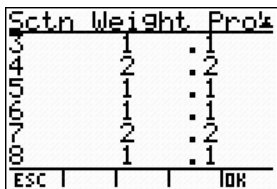Using the APPS button, select the **Prob Sim** App.

2

Select **4.Spin Spinner** and then press **SET** to select the settings.
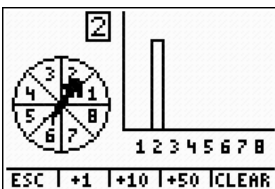
3

Use the settings shown above; this will use a spinner with eight sections and store all the data into a table.
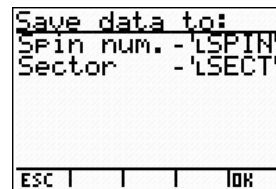
4

Select **ADV** and change the weight for sections 4 and 7 to 2. This will make each of these sections of the spinner larger, and will make the theoretical probability of landing on these sections twice that of the other sections. Select **OK** twice to return to the spinner.
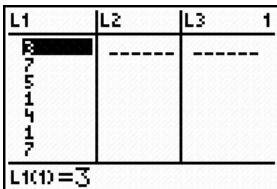
5

Press **SPIN** for the first trial. Notice that the app automatically shows the data in a histogram. To speed up the spinning, select **+1**, **+10**, or **+50** to do multiple spins. Use the **+50** option until you have a total of 251 spins.
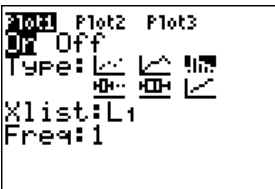
6

Press **ESC** and then select **DATA** to export the data into the displayed lists. Choose **OK** then press **ESC** to exit the Probability Simulation App (when asked to clear the trials from memory, choose **YES**).
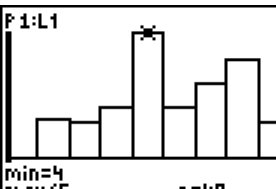
7

Press STAT and select **1:Edit...** to enter the list editor. Use the arrow keys to move the cursor to highlight L_1. Press 2nd STAT and select the list **SECT** to enter the spinner data. Press ENTER to complete the process. This stores the data from your histogram in list L_1.

8

To generate a histogram for the biased random numbers, press 2nd [STAT PLOT] and choose **Plot 1**. Then use the settings shown above. To view the Histogram, press ZOOM and select **9:ZoomStat**.

9

To make the bars span units of 1, press WINDOW and set Xmin=0, Xmax=9, and Xscl=1.

Using the trace feature you can see that the occurrence of 4s and 7s seems to be greater than that of the other numbers. To view the histogram of equally likely values, press 2nd [STAT PLOT] and choose **Plot 1**. Change the Xlist to L_2 and press GRAPH.

2. Look at the graph of numbers from 1 to 8 where 4 and 7 were set up to occur more frequently than the other numbers.

    a. How does this differ from the unweighted data's graph?

    b. How does this compare to other student's histograms of the data where the outcomes are not equally likely?

3. What do you think would happen to the histogram of the unweighted data if the number of trials gets very large?

4. **a.** If the number of trials gets very, very large for the weighted data, theoretically, how many times more often would the value of 4 occur when compared to the number of times 5 would occur? Explain your reasoning.

    b. About how many times more often would the value of 4 occur than 7? Explain your reasoning

    c. What caused the results that occurred in (a) and (b)?

5. Make your own list of "random" number list by entering 250 integers from 0 to 9 into list L_1. Do this without thinking about the values you are typing. Create a histogram of the data and determine which, if any, digits you unknowingly favored. Explain how you can tell from the histogram.

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

Introduction

A histogram is only one of the ways to display data for analysis. Graphs such as line plots can also be useful if the data are over time. For this extension, you will create a line plot of the average high and average low temperatures for Kansas City, Kansas.

The temperatures used can be found on the Web site
http://www.weather.com/outlook/travel/businesstraveler/wxclimatology/monthly/graph/USKS0298?from=search

1

L1	L2	L3	1
6			
7			
8			
9			
10			
11			
12			

L1(12) =12

Using your graphing calculator, press STAT and choose **Edit…** Then enter the values 1 to 12 to represent the months
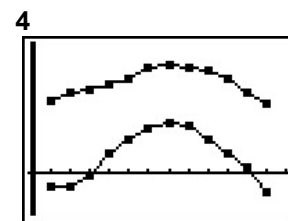
2

L1	L2	L3	3
7	112	52	
8	110	48	
9	106	34	
10	97	21	
11	82	7	
12	73	-19	
-------	-------	-------	

L3(13) =

Use the arrow keys and enter the average high temperature in L2 and the average low temperature in L3.

3

Plot1 Plot2 Plot3
On Off
Type: ⌐ ⊠ ⟋
      ⊞ ⊞ ⟋
Xlist:L1
Ylist:L2
Mark: □ + ·

Press the [2nd] [STAT PLOT] and choose **Plot 1**. Use the settings shown above; this will produce a line plot for the data in lists L_1 and L_2. Do the same for **Plot 2** but with the Ylist L_3.

4

Press [ZOOM] and select **9:ZoomStat** to display the two line plots.

For the Student

While using a histogram suggests one way to detect whether outcomes are equally likely, a theoretical way to detect them is by using confidence intervals. To learn more about confidence intervals, see the NUMB3RS activity "Thumb's Up!" This activity can be downloaded for free by going to **http://education.ti.com/exchange** and searching for "6559."

Additional Resources

- Chi Square is a statistical test used to determine if two groups are different enough in some characteristic by looking at samples from each group. This Web site contains a lesson on this topic: **http://www.mste.uiuc.edu/patel/chisquare/keyprob.html**.

- For information about other types of data plots on the TI-83 Plus/TI-84 Plus graphing calculators, consult your calculator's guidebook. These guidebooks can be downloaded for free from Texas Instruments at:
  http://education.ti.com/educationportal/appsdelivery/download/download_select_product.jsp?cid=us&displaymode=G&contentpaneid=17

Related Topic

- For another look at detecting fabricated data sets, see the NUMB3RS activity "We're Number 1!" This activity can be downloaded for free by going to **http://education.ti.com/exchange** and searching for "6573."