

Activity 7

Collecting Pens

In many real world problems involving probability and randomness, it is difficult, if not impossible, to determine answers through direct computation. In such cases, solutions can often be approximated using simulation strategies. In a simulation, an experiment is designed to resemble the important characteristics of the situation we are examining. Many trials of this experiment are conducted and analyzed to estimate a desired answer. When the simulations involve random outcomes they are often referred to as *Monte Carlo* simulations. In their simplest form, Monte Carlo simulations may involve flipping coins, rolling dice, or spinning spinners. These simulations can also be carried out using random numbers produced by a calculator or computer. The following problem is an example of a situation that can be solved using a Monte Carlo procedure.

The Problem

The Kellogg Company (the cereal makers) once placed a free felt-tipped marker in each box of *Kellogg's® Raisin Bran* they distributed. The back of the box proclaimed:

FREE INSIDE — BRUSH MARKER ...

You'll find one of the 6 washable colors in each specially marked package of *Kellogg's® Raisin Bran* cereal.

Start collecting all 6 NOW!

- ✍ To help you get started analyzing this problem, answer #1 through #3 in the **Questions** section.

This activity will most likely require two 50-minute class periods. If your students have not studied simulations before this time, you may want to have them first work through Activities 5 and 6 in this book.

Students have most likely participated in such collection activities themselves. It is not unusual to find some advertising scheme like this. You may want to search the grocery shelves for some cereal with a similar enticement.

If this promotion were still being conducted today, you could solve this problem by actually going out and buying boxes of cereal. This shopping trip would involve purchasing boxes of cereal and seeing which of the six markers was enclosed in each box. Your spending spree would end when you had a complete set of six pens. Since this is not practical, you need to design a simulation model that accurately reflects the essential characteristics of the shopping trip.

Discuss this problem further in your group and decide on a strategy for obtaining an approximate solution using the calculator's random numbers. If you have completed the *Heads Up!* activities in this book, you should recall that the command `iPart 6rand+1` generates random whole number values from the set {1, 2, 3, 4, 5, 6}.

- ✍ As you analyze the problem and design an experiment to investigate it, think about and answer #4 through #8 in the **Questions** section.

Using the Calculator

Opening a box of cereal and determining which color pen is inside is an experiment with six equally likely outcomes (Red, Tan, Yellow, Blue, Green, and Violet, for example). Entering the `iPart 6rand+1` command is also an experiment with six equally likely outcomes. The outcomes from these two experiments can be matched as shown in the following table.

Table 7.1. Pen Colors

| Pen Color | Red | Tan | Yellow | Blue | Green | Violet |
|-----------------------------|-----|-----|--------|------|-------|--------|
| iPart 6rand+1 Result | 1 | 2 | 3 | 4 | 5 | 6 |

As your group may have suggested, you can simulate purchasing a box of cereal by entering `iPart 6rand+1` on your calculator with the resulting number determining which color of pen is in the box. The calculator command can be repeated until each of the six possible outcomes has been obtained. The total number of times the command was entered gives an estimate of the answer for the original problem.

Use your calculator to simulate one customer purchasing cereal and collecting a set of pens.

Calculating the Results

1. Press **[MATH]** **[>]** (highlights NUM) **2:iPart** 6. Press **[MATH]** **[4]** (highlights PRB) **1:rand** **[+]** 1 **[ENTER]** to get the first random number.
2. Press **[ENTER]** repeatedly to get additional numbers.
 - ✍ Keep a tally of the number of times each number comes up using tables 7.2, 7.3, and 7.4 in the **Questions** section. When all six numbers have appeared at least once, move to the next table and repeat the simulation for two additional customers.

On the TI-83, press 3 to select iPart.

The results from your three “customers” do not necessarily represent the expectations for the average customer. It is best to compile the results of many trials of a simulation and to use these combined data to arrive at an estimate of the desired answer.

The data might best be collected using a show of hands with one person selected to record the number of customers who were able to get all six pen colors by purchasing the given number of boxes of cereal.

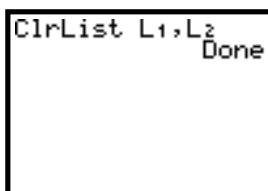
- ✍ Use Table 7.5, **Frequency Table**, in the **Questions** section to summarize the results of all of the trials conducted in your class. Then answer #9 through #11 in the **Questions** section.

If your class schedule requires it, this might be a good point to stop the first part of the lesson.

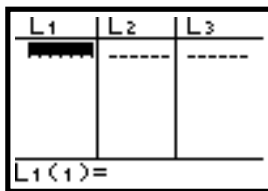
Summary Statistics on the Calculator

You can use your calculator to get other views and summary statistics for the data. The data must be entered into the calculator list storage before statistics and graphs can be produced.

1. Press **[STAT]** **4:ClrList** **[2nd]** **[L1]** **[,]** **[2nd]** **[L2]** **[ENTER]** to clear the needed lists. The display should look like the one at the right.



2. Press **[STAT]** **1:Edit** to gain access to the calculator's list storage. The display should look something like the one at the right.



3. Type the numbers from the first column of the table (representing the number of boxes purchased) into the first list (L1) on your calculator. Type a number and press **[ENTER]**, repeating until all numbers have been entered.

4. Press \blacktriangleright to move to the second list. Type the numbers from the second column (representing number of customers) into the second list (L2) on your calculator. Make certain to enter the numbers in this second column on the corresponding row as found in your table.

Once the paired data has been entered, you can obtain descriptive statistics on the data in Table 7.5, **Frequency Table**.

5. Press STAT \blacktriangleright 1:1-Var Stats 2nd [L1] , 2nd [L2] ENTER to produce a list of statistics summarizing information on the number of boxes purchased by all customers represented in your class simulations.

The calculator produces a variety of statistics, many of which are not needed for our current analysis. Note the arrow pointing down \downarrow in front of the last line. This is an indication that more information is available from the last set of calculations. If you press \blacktriangledown several times, you can scroll to a second screen of statistics. The following graphics give a brief explanation of the statistics presented on the two screens.

The command 1-Var Stats L1,L2 informs the calculator that the data has been stored as a frequency table with the data in L1 and the frequencies in L2.

Many of these statistics are beyond the consideration of middle school students and are used in more advanced settings. In particular, the two measures of standard deviation are very difficult to explain without much more exposure to the concepts of variability and deviation scores, topics often introduced in high school level statistics classes.

The statistics shown in the second screen will be familiar to those who have studied box-and-whisker plots. If students have completed quartile analyses before, it would be appropriate to examine the frequency table and determine the "correctness" of the calculator values from the tabled data.

| 1-Var Stats | |
|------------------|--|
| \bar{x} = | Arithmetic mean of the data. |
| Σx = | Sum of all of the data values. |
| Σx^2 = | Sum of the squares of the data values. |
| Sx = | Standard deviations: measures the amount of variability in the data. |
| σx = | |
| $\downarrow n$ = | Number of data values. |

| 1-Var Stats | |
|----------------|---|
| $\uparrow n$ = | Number of data values. |
| minX= | Minimum: smallest value in the data. |
| Q1= | Lower quartile: cut-off for the bottom 25% of the data. |
| Med= | Median: cut-off point for the bottom 50% of the data. |
| Q3= | Upper quartile: cut-off point for top 25% of the data. |
| maxX= | Maximum: largest value in the data. |

- ✍ Based upon the statistics generated by your calculator, answer #12 in the **Questions** section.

Displaying a Statistical Graph

Your calculator has statistical graphing, or plotting, features that allow the construction and display of a histogram and a box-and-whisker plot of the six-pens data. This can be done by defining the types of graphs desired and by setting appropriate window limits for viewing the graphs. You must also make certain that any function-graphing capabilities are cleared or turned off. The usual order is:

- a. Clear or disable any functions stored in the $\boxed{Y=}$ menu.
 - b. Define an appropriate viewing window.
 - c. Define and activate the appropriate [STAT PLOT] features.
1. To clear the functions stored in the $\boxed{Y=}$ menu, press $\boxed{Y=}$, move the cursor over the first character of any function you find, and press the $\boxed{\text{CLEAR}}$ key. Repeat until all functions are cleared.
 2. Return to the home screen by pressing $\boxed{2\text{nd}}$ [QUIT].

To set an appropriate viewing window for the plots associated with the data, you need to define an interval that contains all of the values in the list. Look at the set of statistics you calculated earlier. Find and record the minimum and the maximum values for the data.

Minimum value: _____ Maximum value: _____

For a histogram you also need to know the maximum *frequency* for any of the values in order to determine the “height” of the graph. This information can be found through a visual search of the second list L2 containing the frequency counts associated with the data values.

3. Press $\boxed{\text{STAT}}$ $\boxed{\text{ENTER}}$ to get to the list storage. Use $\boxed{\downarrow}$ and $\boxed{\rightleftarrows}$ to scroll through and locate the maximum value in L2. Record your finding below.

Largest frequency : _____

These minimum and maximum values are needed in determining the limits along the x-axis. Make certain students can explain why this is so.

Similarly, this largest frequency is needed to determine the limits along the y-axis.

Students should be asked to explain its importance.

Knowing these three values, you are ready to define the limits of the viewing window for the statistical plots.

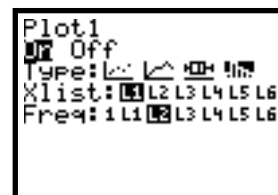
4. Press **WINDOW** and use **▼** to highlight the **Xmin** setting. Set **Xmin** at 5 (one less than the minimum) and press **ENTER**.
5. Set **Xmax** to two more than the maximum, **Xscl** to 1, **Ymin** to -5, **Ymax** to one more than the largest frequency, and **Yscl** to 1. (These settings produce a small border around the actual graph.) Press **2nd** **[QUIT]** to return to the home screen.

It is important to note that the **Xscl** value serves a useful purpose in the plot of a histogram. It is used to define the width of the interval used for each bar in the graph. By defining **Xscl=1**, you are asking for a histogram that displays the frequency counts for each data value in the set. (If **Xscl** were set equal to 2, you would be asking for a frequency count on the 6's and 7's grouped together, the 8's and 9's grouped together, and so on.)

The histogram is just a graphical view of the frequency table. Students should arrive at similar conclusions after viewing both the table and the histogram.

The last step in obtaining a statistical plot is to define the type of plot desired and the data lists to be used in constructing the plot. This can be done by accessing the **Stat Plot** menu.

6. Press **2nd** **[STAT PLOT]** to gain access to the statistical plotting menus. Press **1:Plot1** to access the options for **Plot1**.
7. Using the blue arrow keys to highlight an option and the **ENTER** key to select that option, set the options to match the window at the right.



You now have given the calculator all the information needed to plot the histogram.

8. To see the graph displayed, press the **GRAPH** key.

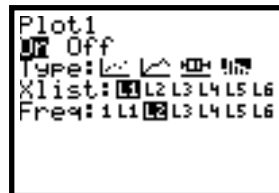
Note the distribution of the data. Do the values occur with the same frequency throughout the entire interval? If not, where do most of the values lie?

Note: You can use the **TRACE** key followed by the blue arrow keys to read coordinates from the histogram. To remove the coordinates from the display, press the **CLEAR** key.

A box-and-whisker plot is also an appropriate display for the six-pens data. You have completed most of the work needed to obtain this plot. To display this plot, return to the statistical plotting menu and redefine the type of graph.

It is not unusual for a box-and-whisker plot of data for this problem to have an extremely long upper whisker (usually outliers are involved).

- To define the box-and-whisker plot, press **2nd** [STAT PLOT] **ENTER**. Select the options shown on the window at the right.



Expect to see the lower three quartiles (the lower whisker and the box) show much less variability.

- To display the plot, press the **GRAPH** key.
- ✍ Note the distribution of the data and answer #13 and #14 in the **Questions** section.

Questions

Answers can vary considerably. A popular answer is 36 (just sounds good?).

The minimum is six, but only for a very lucky person. Students will be asked later to estimate the probability of actually getting all of the pens in six purchases.

Theoretically, the maximum is 5/6 of the number of boxes distributed plus 1, but again this is very unlikely to occur. Some students may suggest 6^6 since this value is so large.

The pen color received in a box of opened cereal is assumed to be random.

There are 6 outcomes for this event (the six pen colors).

These outcomes are equally likely (again our assumption).

1. Assuming that equal numbers of each of the six pen colors were randomly distributed in the packaging of the cereal by the Kellogg company, about how many boxes of cereal do you think a customer would have to buy in order to collect all six of the pens?

2. What do you think is the minimum number of boxes a person might have to buy to get all of the pens? Explain.

3. What do you think is the maximum number of boxes a person might have to buy before getting all of the pens? Explain.

◆ Return to page 62.

4. What random event exists in this situation?

5. What are the possible outcomes for this random event?

6. Are the outcomes equally likely?

7. How can the possible outcomes be modeled using the command `iPart 6rand+1`?

Each of the six outcomes can be matched to one of the 6 integers produced by the command `iPart 6rand+1`. The calculator could be used to repeatedly produce the random integers until each number has been displayed.

8. How could the calculator be used to determine the number of boxes of cereal purchased by a single customer?

The number of times the command needed to be issued gives an estimate to the number of boxes a customer would need to purchase.

◆ *Return to page 62, **Using the Calculator.***

Table 7.2. Simulation for Customer 1

| Pen Color | Tally |
|------------|-------|
| Red (1) | |
| Tan (2) | |
| Yellow (3) | |
| Blue (4) | |
| Green (5) | |
| Violet (6) | |

Total boxes purchased by this customer _____

Table 7.3. Simulation for Customer 2

| Pen Color | Tally |
|------------|-------|
| Red (1) | |
| Tan (2) | |
| Yellow (3) | |
| Blue (4) | |
| Green (5) | |
| Violet (6) | |

Total boxes purchased by this customer _____

Table 7.4. Simulation for Customer 3

| Pen Color | Tally |
|------------|-------|
| Red (1) | |
| Tan (2) | |
| Yellow (3) | |
| Blue (4) | |
| Green (5) | |
| Violet (6) | |

Total boxes purchased by this customer _____

◆ Return to page 63, **Calculating the Results.**

**Table 7.5. Frequency Table:
Class Results for SIXPENS Collection Problem**

| Number of Boxes Purchased | Number of Customers |
|---------------------------|---------------------|
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

The data is being recorded in a frequency table with one column representing the actual data values and the other the frequency counts corresponding to each data value.

Answers will vary but it is expected that students will find that most customers can expect to buy between ten and twenty boxes of cereal.

The **Frequency Table** provides one way to look at the data your class has collected. Your responses to #9 through #11 should be based upon the distribution of the data in the table, not upon any further calculations.

9. About how many boxes of cereal do you believe a customer would have to buy in order to collect all six of the pens?

10. What are the minimum and maximum numbers of boxes customers had to purchase? Were these values close to those you predicted earlier?

11. The difference between the maximum and minimum data values in a set is called the *range*. What is the range for your class data?

◆ Return to page 63, **Summary Statistics on the Calculator**.

12. Based upon the experiments you have just conducted and the statistics produced by your calculator, how many boxes of cereal can the *average* customer expect to buy in order to collect all six pen colors? Explain your answer. For example, did you use the mean? The median? Some other summary statistic? Why? How close is your answer to your original guess?

There is no one correct way to answer this question. Accept any reasonable answer for which students are able to give support.

◆ *Return to page 65, **Displaying a Statistical Graph.***

13. Look at the distribution of the data on the box-and-whisker plot. Where does most of the data lie? How does the plot convey this information?

14. Using the TRACE key to see the coordinates of each point on the plot, answer the following questions:

a. A typical customer would expect to get all six pen colors 50% of the time by purchasing _____ or fewer boxes of cereal.

Use the median for this 50% figure.

A typical customer would expect to get all six pen colors 75% of the time by purchasing _____ or fewer boxes of cereal.

Use the upper quartile for this 75% figure.

b. Describe how the histogram and the box-and-whisker plot provide information about the six pens collection problem. How does the summary information provided by these graphical views differ from the tabular view used earlier?

The histogram and box plot are visual display of the data that provide at a quick glance a sense of the "average" data value (mean from the histogram and median from the box plot). The tabular view of the data provides more specific information on the complete data set; however, interpretation of the results may require a more lengthy analysis.

Problems for Additional Exploration

1. Suppose that a customer buys several boxes of *Kellogg's® Raisin Bran* hoping to get all six of the pen colors. Use your class simulation results to answer each of the following.

Examine the frequency table. The ratio of the number of customers getting all pens in just 6 boxes to the total number of customers provides an estimate for this probability. (The estimate could end up being zero.)

To estimate the probability of getting all pens in 15 or fewer boxes, first determine the number of customers getting all pens with 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 purchases. Then divide this sum by the total number of customers.

Students could use *iPart 8rand+1* and conduct the simulation again including the generation of the graphs using eight colors instead of six.

- a. If a customer buys exactly six boxes of cereal, what is the probability that he or she will get all six different pen colors?

- b. If the customer buys fifteen boxes of cereal, what is the probability that he or she will get all six pen colors?

Note: The customer may not have to open all fifteen boxes to get all six colors.

2. How would you simulate the pen collection problem if the Kellogg company decided to distribute eight different colors of pens in their cereal instead of six?

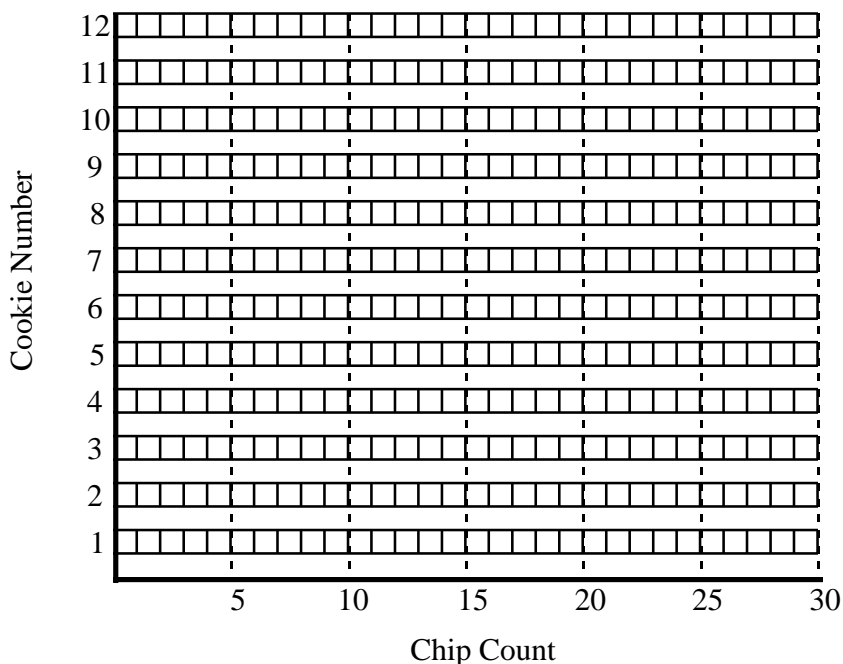
3. The Chocolate Chip Mixing problem presented below provides another example of a rather complex problem that can be solved using Monte Carlo simulation techniques..

The Cookie Bin is a small shop that bakes and sells cookies at an indoor shopping mall. The owner plans to introduce a new bite-sized chocolate chip cookie to her line of products. She would like to have each cookie contain at least five chocolate chips. To minimize the cost of ingredients, she wants to determine the least number of chips that need to be mixed into a batch of dough for a dozen cookies so as to meet the five-chip-per-cookie minimum. How many chips should she use for each batch? (Note that she will need *at least* 60 chips.)

- a.** Assume that each chip placed into the dough for one dozen cookies has an equally likely chance of ending up in any one of the twelve cookies made from that batch. Simulate the random placement of chocolate chips into the individual cookies in the batch. The calculator command `iPart 12rand+1` can be used to determine which of the 12 cookies is to get a given chocolate chip. Since each cookie is to contain at least five chips, repeatedly use the command and keep a frequency count of the number of chips in each cookie until the minimum number of chips needed to meet the five-chip-per-cookie requirement is reached. You can fill in the following bar graph to record your progress.

You may want to ask students to “place” the first 60 chips and then examine the distribution to see how “uneven” the chip placement is once the minimum number of chips have been mixed in.

Expect a wide variation in answers to this simulation. In sample trials, we have experienced chip counts ranging from the mid-seventies to over 160 with the median across all trials somewhere between 100 to 110.



- b.** How many chocolate chips were needed for the batch of cookies you “baked”?
- _____
- _____
- c.** Collect the data from all batches of cookies “baked” by members of your class and enter them into a list on your calculator.
- d.** Construct a box-and-whisker plot of the class data.

If the baker were to mix the median number of chips into several batches of cookies, he should expect that in 50% of the batches, each cookie would contain 5 or more chips.

- e.** What is the median number of chocolate chips used to meet the five-chip-per-cookie minimum? How would you interpret this number (the median) for the Cookie Bin owner?

Use the upper quartile.

- f.** What number of chips will result in each cookie having five or more chips 75% of the time?

Answers will vary but should be supported.

- g.** How many chocolate chips do you think the Cookie Bin owner should put into the dough for each batch of one dozen cookies? Explain.
