

# The Gumball Machine

Alexander White, M. Alejandra Sorto, Rini Oktavia  
Texas State University – San Marcos  
[sorto@txstate.edu](mailto:sorto@txstate.edu)

**Published: March 2014**

## Overview of Lesson

This lesson gives students the opportunity to explore and discuss the variation that occurs in sampling. Students are asked to imagine that a gumball machine contains 1 red, 2 green, 3 yellow, and 4 blue gumballs that were thoroughly mixed before they were put into the machine. Then students are asked several questions about the probabilities of certain events that could occur.

This lesson has four goals: (1) to understand that events associated with small probability can and do occur; (2) to realize that the distribution of data from small samples often does not reflect the parent distribution; (3) to take variation into account when describing possible sampling outcomes; and (4) to recognize an effect of sample size on sampling variation. In order to lead students to focus more on variation than point estimates, students are asked to imagine that 5 students select 10 gumballs each and to predict the number of blue gumballs the students get (see Figure 1).

Name: \_\_\_\_\_

Imagine that 5 students select 10 gumballs with replacement from the magic gumball machine with 4 blue, 3 yellow, 2 green and 1 red gumballs. Predict the numbers of blue gumballs the students get.

| Student A   | Student B   | Student C   | Student D   | Student E   |
|---|---|---|---|---|
|  |  |  |  |  |
| Select 10   | Select 10   | Select 10   | Select 10   | Select 10   |
| ↓   | ↓   | ↓   | ↓   | ↓   |
| Number of Blues _____   | _____   | _____   | _____   | _____   |

Figure 1. Visualizing the gumball machine.

To see the effect of sample size, students are asked to find and plot the experimental probability for both samples of  $n = 10$  gumballs and  $n = 20$  gumballs; and to help visualize variability across many samples, students are given circular colored stickers and asked to plot their answers for each color and sample size on large number lines in the front of the class. After a discussion about the prediction of the outcomes, students do the experiment. After students record their outcomes, they compare with their previous predictions. Based on previous use of the lesson conducted, we find that most students are surprised to see that their predictions were, in past students' own words, "way off".

### **GAISE Components**

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level A activity.

### **Common Core State Standards for Mathematical Practice**

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.

### **Common Core State Standards Grade Level Content (Grades 3 through 5)**

3. MD. - 5. MD. Represent and interpret data.
6. SP. 1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
6. SP. 4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
6. SP. 5. Summarize numerical data sets in relation to their context by: reporting the number of observations.

### **NCTM Principles and Standards for School Mathematics**

#### **Data Analysis and Probability Standard for Grades 3-5**

**Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:**

- design investigations to address a question and consider how data-collection methods affect the nature of the data set;
- collect data using observations, surveys, and experiments;
- represent data using tables and graphs such as line plots, bar graphs, and line graphs;
- recognize the differences in representing categorical and numerical data.

**Select and use appropriate statistical methods to analyze data**

- describe parts of the data and the set of data as a whole to determine what the data show.

**Develop and evaluate inferences and predictions that are based on data:**

- use proportionality and a basic understanding of probability to make and test conjectures about the results of experiments and simulations.

## Prerequisites

Students should know how to place fractions (in particular tenths and twentieths) on the number line.

## Learning Targets

Students will be able to graphically display data and interpret data.

## Time Required

1 class period.

## Materials Required

Color sticky notes, a cup, color chips, pencil and paper, and a copy of the Activity Sheet (pages 7 through 9).

## Instructional Lesson Plan

### The GAISE Statistical Problem-Solving Procedure

#### I. Formulate a Question

Begin the investigation by asking students to imagine that a gumball machine contains 1 red, 2 green, 3 yellow, and 4 blue gumballs that were thoroughly mixed before they were put into the machine. Then ask students to imagine that 5 students select 10 gumballs each with replacement (i.e. select one gumball, observe its color, replace the gumball in the machine, repeat this process 10 times). To encourage students to develop their abilities in formulating questions, ask students to write some questions that they would be interested in investigating about the number of red, green, yellow, and blue gumballs that the 5 students will get.

The class generates a list of questions, which the teacher records on the board. Some possible questions might be:

- What is the list of possible numbers of blue gumballs that the five students get?
- What number of blue gumballs showed up the most?
- What color of gumballs is chosen the most?

Have the students think about how the number of red, green, yellow, and blue gumballs that each student gets will vary. Then ask students to answer the following questions:

1. If you got one gumball from the machine, which color do you think would come out?
2. If we sampled 10 gumballs from the magic gumball machine, how many gumballs of each color will come out?
3. Imagine that 5 students pick 10 gumballs each, predict the number of blue gumballs student A will get? How about student B? Student C, D and E?

#### II. Design and Implement a Plan to Collect the Data

Give students a copy of the "Gumball Machine" activity worksheet.

**First Task.** (Problem 1) Ask one of the students to read out loud the first question ("If you got one gumball from the machine, which color do you think would come out?"). Demonstrate the gumball machine and the gumballs using a cup and color chips. Then, ask students to

individually answer the first question and explain their reasoning in writing on their worksheet. After a few minutes, asks students to share their responses. It is more likely that students will say that a blue gumball will come out because the machine has more blue gumballs than any other color. Even though "blue" is the best prediction for a single outcome, it is important that students are aware of the complement of this event. Therefore, as a follow-up question, ask students: "What are the chances of getting a color different than blue?" and "Are the chances of getting a color different than blue the same as getting blue?". Ask students to record their answers.

**Second Task.** (Problem 2) Ask students to answer the second question, "Suppose that after each gumball comes out, the gumball wizard magically puts another one of the same color into the machine so that the gumball machine always has the same number of each color of gumball. If you took 10 gumballs out of this magic machine, with the gumball wizard replacing your gumball each time, how many times do you think each color would come out?" Since the wording of the question may lead students to think that all colors occur with equal frequency instead of in the 4:3:2:1 proportion, act out the selection and replacement process in the front of the class. The answers to this question vary much more than the answers to the first question. Students may show a large variety of reasoning. For example, a few students may say things such as "you don't know" or "you never know" because "it will be restored each time" or "because you take one out and you are putting them back in the machine". They may show an "equiprobability bias" (Lecoutre 1992) that is the belief that since we are selecting randomly: anything can happen and everything has the same chance of occurring. Other students may say "at least all once or twice" or "at least twice" because "there is ten and it's almost like pulling each one time" or "each gumball can come out equally". Other students may give a different distribution of colors such as "3 blue, 4 yellow, 1 green, 2 red," or "6 blue, 2 yellow, 1 green, 1 red" or outcomes that contain less than 10 total gumballs. Other students may predict "4 blue, 3 yellow, 2 green, 1 red" and their explanations are based mainly on the original distribution of the colors in the gumball machine.

**Third Task.** (Problems 3 - 5) This task focuses only on the number of blue gumballs. The question addressed is "Imagine that 5 students select 10 gumballs each, what is your prediction of the number of blue gumballs the students get?" Phrasing the prediction question in this fashion will give the students the opportunity to explore and discuss the variability we would expect in sampling. Students' predictions for the numbers of blue gumballs will vary. Most students may provide a sequence of numbers varying from 1 to 9. Using statistical theory as a guide, the predictions can be characterized as "low", "high", "wide" and "narrow". Low predictions are a sequence of numbers with means less than 3 and high predictions are those with means greater than or equal to 6. A "narrow" prediction is a sequence of repeated numbers (e.g. "6, 6, 6, 6, 6") that suggest no variability in the sampling results. A sequence is "wide" if its range is greater than or equal to 6 (Shaughnessy, Ciancetta, and Canada 2004).

Give each pair of students a cup with 10 colored chips and ask them to simulate the gumball machine. Have the students follow the directions given on the activity sheet. Make sure the students sample with replacement and shake the cup between each selection. Students will have outcomes for 10 trials. After the students record their outcomes, compare the outcomes to their previous predictions.

### III. Analyze the Data

Ask students to repeat the experiment one more time (Problem 6). Then have students combine their data from both experiments (Problem 7). So, students will have outcomes for samples of size 10 and size 20. In problems 8 and 10 on the worksheet, students will compute the fractions of blue, yellow, green, and red that they get for both 10 and 20 trials. This allows for a discussion of experimental probability and the possibility to compare the results from the two different sample sizes by placing the values on a number line (Problems 9 and 11). To show the variability of the experimental probabilities, have each student place a blue dot sticker representing their observed results for  $n = 10$  trials on a pre-made large number line (see Figure 2).

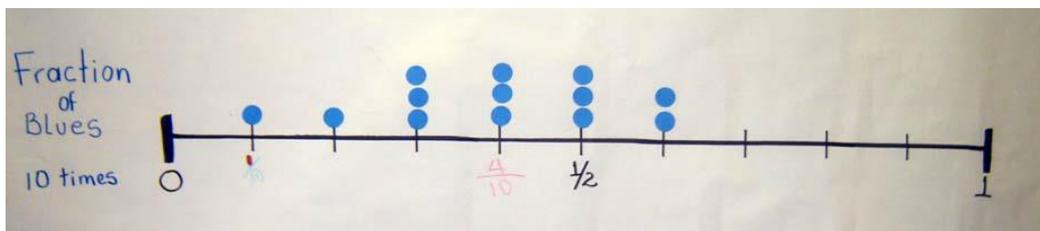


Figure 2. Dot plot of observed experimental probabilities

Use a separate number line to plot the results for  $n = 20$ . Compare the results from the two dot plots.

### IV. Interpret the Results

Use the dot plots as the spring board for a discussion about what occurred when selecting the samples of gumballs. For example, you may ask the students “Did anyone get 4 blues?” and “Where do you see this on the graph?” Compare the results to their predictions about the number of blues for the 5 imaginary students. Discuss why it is not likely to get 4 every time, but it is not likely to get 0 or 10. The discussion should not focus only on the center of the resulting distribution but the spread and extreme values as well. This will set the basis for more formal reasoning about variation down the road. Have the students compare the results for  $n = 10$  and  $n = 20$ . By comparing the results, students should see that with a larger sample the experimental probabilities are less spread out.

## Assessment

Imagine that a gumball machine contains 1 red, 2 green, 3 yellow and 4 blue gumballs that were thoroughly mixed before they were put into the machine.

1. If each student selected 10 gumballs. Which is more likely to be true?
  - a. All of the students will have exactly 1 red ball.
  - b. Most of the students will have close to 1 red ball.
2. If each student selected 10 gumballs. Which is more likely to be true?
  - a. No students will have 2 red balls.
  - b. Some, but not many students will have 2 red gumballs.

## Answers

1. (b)
2. (a)

## Possible Extensions

1. After the dot plots are constructed, ask students to describe the shape, center and spread of the distribution of the number of blue gumballs and to examine the distribution for clusters, gaps, and potential outliers.
2. Ask students to construct a boxplot to display the data distribution and to determine if there are any outliers using the  $1.5 \times \text{IQR}$  rule.
3. Ask students to calculate and discuss measures of center (mean, median).
4. Ask students to discuss the spread of the distribution using common measures of spread: range, interquartile range, mean absolute deviation.

## References

1. Lecoutre, M. P. (1992), "Cognitive models and problem spaces in 'purely random' situations," *Educational Studies in Mathematics*, 23, 557- 568.
2. Shaughnessy, J.M., Ciancetta, M., and Canada, D. (2004), "Types of Students Reasoning on Sampling Tasks," *Proceedings of the 28<sup>th</sup> Conference of the International Group for the Psychology of Mathematics Education*, 2004 Vol. 4, pp. 177 – 184.
3. Sorto, M. A. & White, A. (2008). The Gumball Machine: Linking Research and Practice about the concept of variability. *Journal of Statistics Education*. 16(2).  
Online: <http://www.amstat.org/publications/jse/>

# The Gumball Machine Activity Sheet

Names: \_\_\_\_\_  
\_\_\_\_\_

1. Imagine that a gumball machine contains 1 red, 2 green, 3 yellow and 4 blue gumballs that were thoroughly mixed before they were put into the machine. If you got one gumball from the machine, which color do you think would come out? \_\_\_\_\_

Explain your answer.

2. Suppose that after each gumball comes out, the gumball wizard magically puts another one of the same color into the machine so that the gumball machine always has the same number of each color of gumball. If you took 10 gumballs out of this magic machine, with the gumball wizard replacing your gumball each time, how many times do you think each color would come out? Blue \_\_\_\_ Red \_\_\_\_ Green \_\_\_\_ Yellow \_\_\_\_

Explain your answer.

3. Imagine that 5 students select 10 gumballs each from the magic gumball machine with 4 blue, 3 yellow, 2 green and 1 red gumballs. Predict the numbers of blue gumballs the students get.

| Student A   | Student B   | Student C   | Student D   | Student E   |
|---|---|---|---|---|
|  |  |  |  |  |
| Select 10   | Select 10   | Select 10   | Select 10   | Select 10   |
| ↓   | ↓   | ↓   | ↓   | ↓   |
| Number of Blues _____   | _____   | _____   | _____   | _____   |

4. Do the following experiment to simulate the gumball machine:
- Put 4 blue chips, 3 yellow chips, 2 green chips, and 1 red chip in a paper cup.
  - Mix up the chips thoroughly.
  - Draw one chip out (without looking), tally its color in the middle row of the table below, replace the chip in the cup, and mix the chips again.
  - Repeat these steps until you have tallied 10 trials.

| Color of Chip      | Blue | Yellow | Green | Red |
|--------------------|------|--------|-------|-----|
| Tally              |      |        |       |     |
| Number (out of 10) |      |        |       |     |

5. How close were your results to the predictions you made in Task 2? \_\_\_\_\_ Explain.

6. Repeat the experiment one more time and record your data below.

| Color of Chip      | Blue | Yellow | Green | Red |
|--------------------|------|--------|-------|-----|
| Tally              |      |        |       |     |
| Number (out of 10) |      |        |       |     |

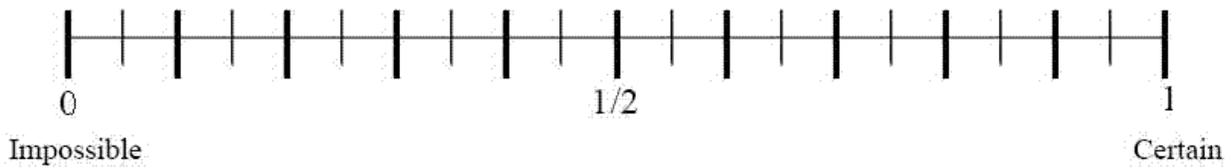
7. Combine the results for all 20 trials below.

| Color of Chip      | Blue | Yellow | Green | Red |
|--------------------|------|--------|-------|-----|
| Number (out of 20) |      |        |       |     |

8. Using the table in Task 4, compute the experimental probability of drawing each color. In other words, compute the fraction of blue, yellow, green and red you got:

| Color of Chip            | Blue | Yellow | Green | Red |
|--------------------------|------|--------|-------|-----|
| Experimental Probability |      |        |       |     |

9. Use B, Y, G and R to represent the four colors. On the scale below, show the experimental probability of drawing each color you got in Task 8.



10. Now use the table in Task 7 to compute the experimental probability of drawing each color. In other words, compute the fraction of blue, yellow, green and red you got:

| Color of Chip            | Blue | Yellow | Green | Red |
|--------------------------|------|--------|-------|-----|
| Experimental Probability |      |        |       |     |

11. Use B, Y, G and R to represent the four colors. On the scale below, show the experimental probability of drawing each color you got in Task 10.