Exploring the Sampling Distribution of the Sample Proportion with a Study of Fingerprint Types in the US

Megan Mocko, University of Florida, and Katina Gothard, Eastern Florida State College
Published: March 2023

Overview of Lesson

In this activity, we will explore the population distribution of fingerprints with loops and the sample distribution of fingerprints with loops. Students will also explore how the sample proportion of fingerprints with loops varies from sample to sample. After collecting many sample proportions, they will then explore the distribution of many sample proportions.

This activity is meant to be implemented as an introduction to the definitions of and contrasts the differences between *population distributions*, *sample distributions*, *sampling distributions,* and *standard error*. In this lesson, the students will draw simple random samples, calculate sample proportions, construct histograms, and compare sampling distributions' shape, mean, and variability. By the end of the activity, students will be able to distinguish between these terms and estimate the standard error for a sampling distribution.

Type of Data

- One categorical variable
- Data collected as a class

Learning Objectives

After completing this activity, students will be able to:

- Identify the sampling distribution of the sample proportion
- Describe what happens to the mean, standard deviation, and shape for sampling distributions as the sample size is increased
- Compare the differences between the population distribution, sample distribution, and sampling distribution.
- CCSS.MATH.CONTENT.HSS.ID.A.2: Use statistics appropriate to the shape of the data distribution to compare center and spread of two or more different datasets.
- CCSS.MATH.CONTENT.HSS.ID.A.3: Interpret differences in shape, center, and spread in the context of the datasets.

Audience

- Students in a first semester of statistics.
- *Prerequisites:* This activity is to be completed after studying the Normal distribution but before the Central Limit Theorem. Students will need to know how to identify a variable of interest and compute a proportion.

Time Required

- 75 minutes

Materials and Preparation Required

- Clear bucket with at least 10,000 beads (60% of the beads are white, 35% of the beads are green, and 5% of the beads are black)
  - Note: To ensure colorblind students can distinguish between the colors, use white, black, and one other color only. Do not use two primary colors because they will be indistinguishable to students who are colorblind.
- Paddles of size n = 40. The paddles will be used to randomly select beads from the clear bucket.
  - Instructions on buckets, paddles, and beads are in Appendix B. Instructions for 3D printing the paddles are also attached.
- Masking tape to cover 8 rows on half of the paddles if you only have paddles with n = 100.
- Computer with access to the internet to access the free *StatKey app*: http://www.lock5stat.com/StatKey/ The instructor may show the applet at the front of the room or students can use the applet on their own devices. Although the buckets of beads provide a great tactile demonstration, the lesson can be completed with *StatKey* alone.

Lesson Plan

In this activity, the students will explore the population distribution of fingerprints with loops and the sample distribution of fingerprints with loops. Students will also explore how the sample proportion of fingerprints with loops varies from sample to sample. After collecting many sample proportions, they will then explore the distribution of many sample proportions.

The purpose of this activity is for the students to understand the sampling distribution and its properties. The students will complete a guided study of the mean and standard deviation of the sampling distribution and its shape.

This activity is meant to be implemented before discussing sampling distributions but after discussing the Normal distribution.

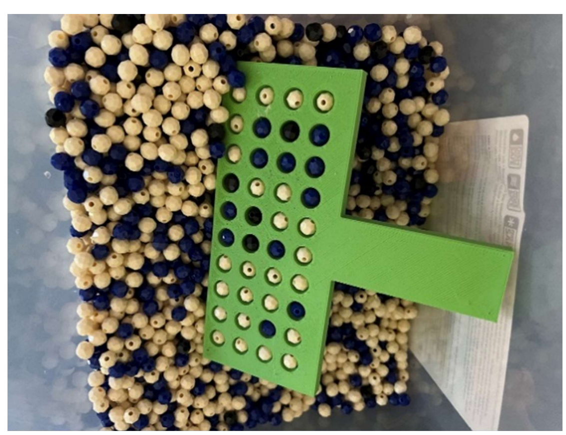
Context

Each person has a unique fingerprint. If we look at the fingerprint of the thumb, there are three common characteristics: loops (lines that bend back on themselves), whorls (such as circles or ellipses), and arches (wave-like patterns). In the United States, 60% of the population has loops, 35% whorls, and 5% arches.

Loop	Whorl	Arch
		

For this activity, you will have a container of beads that represents the population. Each bead represents someone's fingerprint. The white beads represent loops, the green (or blue) represent whorls, and the black beads represent arches. A paddle can then be used to take a random selection. With each new scoop of the paddle, a new random selection will appear on the paddle. The student can then count the number of beads for a certain color. They can then discover and witness sampling variability.

Instructions on buckets, paddles, and beads are in the appendix. Instructions for 3D printing the paddles are also attached.



This lesson consists of a handout in which students complete portions in pairs and triads. After completing a part, the students will discuss it as a class. The activity is marked with stop signs, so the students know to wait for the class discussion.

This lesson plan file begins with a detailed description of what the teacher and student should be doing at each stage of the lesson, followed by the teacher's version of the handout (with solutions). The student's version of the handout is included at the end.

Introducing the Question

1. The instructor should have the students form pairs or triads. The instructor encourages the students to introduce themselves to each other.

2. Below are some optional pre-activity questions, time permitting, for students to answer in groups.

   Directions: With your partner(s), discuss and record answers to the questions that follow. Assign someone the responsibility of sharing your answers with the class. Do not reference any notes or technology. You will have 5 minutes to discuss.
   - If everyone in your group took a random sample of students at the same school, would they get the same sample?
   - Suppose that they collected information about the proportion who had brown eyes. Would each sample have the same proportion of people with brown eyes?

   After five minutes, the instructor has a volunteer report out their discussion about the first two questions. As the students report, make sure to clarify that each sample obtained would be slightly different, and each proportion from the samples would be somewhat different.

3. The instructor explains the three different types of fingerprints and has the students examine example images. The instructor then encourages the students to look at their own fingers and determine which type of fingerprint they may have.

4. The instructor will then verbally introduce the three primary questions of the activity and direct students to where it is written in the handouts.
   - *What is the difference between population distribution, sample distribution, and sampling distribution?*
   - *How do the mean, standard deviation, and shape for sampling distributions change as the sample size increases?*
   - *Why is repeated sampling needed to understand the sampling distribution but not required for inference?*

Part I: Explore the Population Distribution

1. Have the students read through part one of the handout and answer the questions. This part of the handout describes the how the container of beads represents the population of fingerprints and provides definitions for a few key terms: *population, population*

distribution, success, and population proportion. It asks students to identify the population proportion of loops and make a plot of the population distribution.

2. After about five minutes, have the students share their answers. Emphasize that the bucket of beads is the entire population and have them consider how long it might take to find the proportion of loops (white beads) for the whole bucket.

Part II: Explore the Sample (Data) Distribution

1. The instructor should pick two student representatives to randomly sample from the bins.

   As an alternative to the beads and paddles, consider using a free app such as *StatKey*. Directions for *StatKey*:

   - Click on http://www.lock5stat.com/StatKey/
   - Next to **Sampling Distributions**, select Proportions
   - For Loops:
     - Click Edit Proportion and enter 0.60.
     - Click Choose samples of size n = and enter the paddle size (40)
     - You can generate 1, 10, or 100 samples at a time until you have enough sample proportions to see the general shape of the sampling distribution.
     - If you hover your pointer over each dot, you will see the proportion to the far right of **Count**, **Sample Size**, **Proportion**

2. After the results of the two samples are given, ensure that the students fill in this information on the worksheet. There is space to record the number of loops and the sample proportion and create a plot of the sample distribution for each student. This section also provides definitions for a few more terms: *sample, sample proportion, and sample (data) distribution*. The students should then have time to answer questions 1 – 6. Questions 1-2 ask students to identify the random variable and classify it as categorical. Questions 4-6 draw attention to the fact that the population distribution and parameter are unchanging, but the sample distribution and statistic will change from sample to sample.

3. When most students have finished parts II, the instructor should lead the class through a discussion of the answers to questions 1 – 6. For questions 1 and 2, ensure that the students concentrate on the fingerprint type as the variable. For question 3, encourage the students to look at the graph of the population distribution on each other's worksheets. The instructor can ask the students if everyone has the same graph drawn here and ask them to explain why that would be the case. It might be helpful to remind the students that the graph for this problem represents the population – the entire bucket of beads. For question 4, ensure students are now focusing on the samples – the results of each paddle draw. Continue the discussion about the answers for questions 5 and 6, make sure that students carefully distinguish between the sample and the population.

Part III: Explore the Sampling Distribution of the Sample Proportion

1. This section of the handout asks students to consider what would happen if they drew many samples of size 10, 40, and 300 for parameters $p = 0.6$ and $p = 0.05$. If time allows and you can access technology in the classroom, you could have students build the sampling distributions for the sample proportion with the different values of p and n. If not, graphs of the sampling distribution for each combination of sample size and parameter are provided on the handout.

   Directions for *StatKey*:

   - Click on http://www.lock5stat.com/StatKey/
   - Next to **Sampling Distributions**, select Proportions
   - For Arches:
     - Click Edit Proportion and enter 0.05.
     - Click Choose samples of size n = and enter the paddle size (10, 40 or 300)
     - You can generate 1, 10, or 100 samples at a time until you have enough sample proportions
     - If you hover your pointer over each dot, you will see the proportion to the far right of **Count**, **Sample Size**, **Proportion**
   - For Loops:
     - Click Edit Proportion and enter 0.6.
     - Click Choose samples of size n = and enter the paddle size (10, 40 or 300)
     - You can generate 1, 10, or 100 samples at a time until you have enough sample proportions
     - If you hover your pointer over each dot, you will see the proportion to the far right of **Count**, **Sample Size**, **Proportion**

2. For Part III, the students should work on each part up until the stop sign (questions 7 - 11). These questions ask students to consider how the shape of the sampling distribution changes based on the sample size and the parameter. The handout provides definitions for a few key terms: *sampling distribution of a statistic* and *sampling distribution of the sample proportion*. It also provides a rule of thumb for when the sampling distribution of the sample proportion can be considered approximately Normal.

   Important Note: The Rule of Thumb for when the sampling distribution of the sample proportion can be assumed approximately Normal is kept blank for the students to complete. This is because some textbooks use the requirement that the number of expected successes and failures is 15, whereas others require 10. The instructor should either fill in this portion of the worksheet before copying or write the relevant requirements based on their textbook on the board.

3. When most students get to the stop sign, the instructor should lead a class discussion about the answers. For questions 7 – 9, the instructor should emphasize that regardless of

the initial shape of the distribution, the sampling distribution's shape for large n is approximately normal (bell–shaped). For questions 10 – 11, ensure that the students check for np and n(1-p). Some students want to only check np. If needed, use StatKey to give an example where np is large enough, but n(1-p) is not large enough for the sampling distribution to be approximately bell-shaped.

4. Now, the instructor should have the students work on each part until the next stop sign (questions 12-16). These questions draw attention to the fact that the mean of the sampling distribution doesn't change as the sample size changes. Keep an eye on the groups. Some students will try to find a trend rather than stating that it stays the same. When the students have finished, go over the questions and emphasis that the mean stayed the same regardless of the sample size.

5. The students should now move on to questions 17 – 19. These questions draw attention to the fact that the variability of the sampling distribution decreases as the sample size increases. They are also given a formula to calculate the standard deviation of the sampling distribution. Note: Some of the students tend to get stuck on the part that asks which characteristics of the graph influenced their decision. It is crucial to watch the groups and ensure that the students proceed on this section. They may need a hint or two to get them started

   Hints for questions 17 - 18 How can you describe to someone else that the data is less variable? Would the range of the values help you discuss this?

6. After most of the students have finished, discuss these answers. Ensure that the students give enough detail in questions 17 and 18 about why they stated that variability had decreased. Also, for question 19, be prepared to provide a calculator demonstration if students don't have the correct standard deviation for the sampling distribution of the sample proportions.

Part IV: Next Steps

1. If time allows, have students discuss if they think they always need to take repeated samples or if one random sample would be sufficient. Have a few students report from the groups. The idea that repeated samples are now required for research is a common misconception, and time is well spent to keep this misconception from taking hold.

Part V: Conclusion

1. This last portion of the lesson asks students to synthesize the information that they learned. There are 14 questions that summarize the main ideas of the lesson. Let the students have time to grapple with the terminology and have the correct answers before they leave for the day.

2. When reviewing the answers to this portion, it is a good idea to bring students back to the three main questions of the activity.
   - *What is the difference between population distribution, sample distribution, and sampling distribution?*

- *How do the mean, standard deviation, and shape for sampling distributions change as the sample size increases?*
- *Why is repeated sampling needed to understand the sampling distribution but not required for inference?*

Attached Materials

- Teacher version of the handout (with solutions)
- Student version of the handout (blank)
- Instructions about paddles and beads
- Instructions for 3D printing paddles (attached as a separate file)

Exploring the Sampling Distribution of the Sample Proportion with a Study of Fingerprint Types in the US

Question

Fingerprints are often used by law enforcement to fight crime and apprehend wrong doers. Each person has a unique fingerprint. If we look at the fingerprint of the thumb, there are three common characteristics: loops (lines that bend back on themselves), whorls (such as circles or ellipses), and arches (wave like patterns). In the United States, 60% of the population has loops, 35% whorls, and 5% arches.

Loop	Whorl	Arch
		

In this activity, we will answer the following questions:

- *What is the difference between population distribution, sample distribution, and sampling distribution?*
- *How do the mean, standard deviation, and shape for sampling distributions change as the sample size increases?*
- *Why is repeated sampling needed to understand the sampling distribution but not required for inference?*

Data

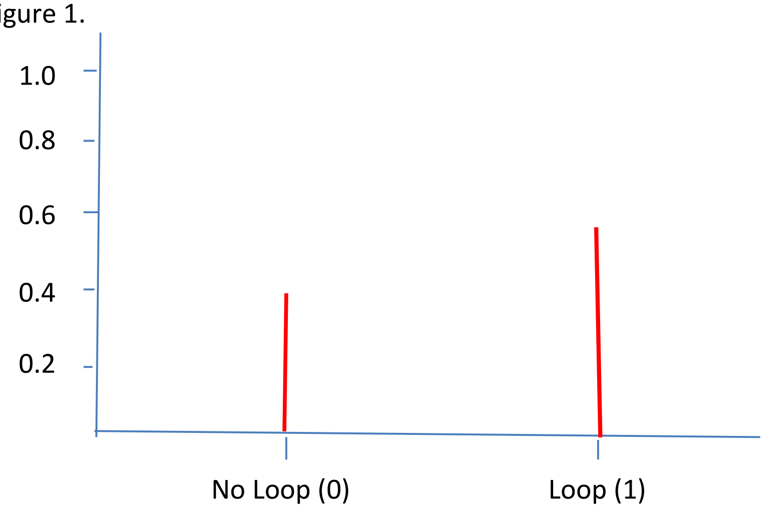
In this activity, we are going to explore three different types of distributions: *population distribution*, *sample (data) distribution,* and *sampling distribution* of the sample proportion.

Part I: Let's Explore the Population Distribution.

We will use a bin filled with beads to represent the American population. In the bin, you will find 60% of the beads are white, 35% of the beads are green, and 5% of the beads are black. This corresponds to loops, whorls, and arches, respectively. Imagine that you were to take out all the beads from the container and record if each bead represented a person with a loop or without a loop on their fingerprint.

- What is the population proportion of loops? _p = 0.60___
- Make a plot of the **population distribution** where 1 is "loops" and 0 is "not loops".

Figure 1.



Population - group of people or objects in which we are interested

Population Distribution - For the population, it shows all possible values of the variable and how often these values occur

Success – person or object displaying the characteristic of interest

Population Proportion (p) – the number of successes divided by the total number in the group in which you are interested (p = X/N)

Part II: Let's Explore the Sample (Data) Distribution.

Now we are going to explore the sample distribution by taking a sample of size 40 from the bin. We will have two students from the class do this. Calculate the sample proportion for each draw.

Sample - subset of the population for whom we have (or plan to have) data, often randomly selected

Sample Proportion – the number of successes divided by the number of outcomes in the sample ($\hat{p}$ = x/n)

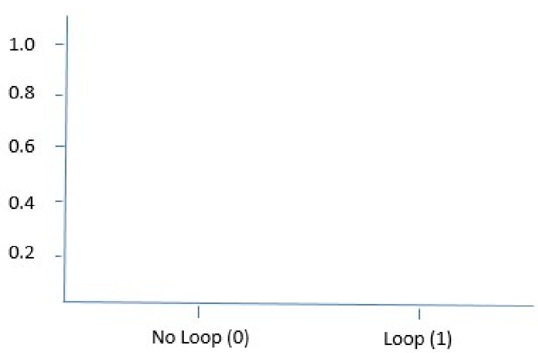
Sample (data) Distribution - For the one sample collected, it shows all possible values of the variable that occurred in the sample and how often these values occurred

Student 1

	n = 40
Number of loops (x)	Answers will vary, close to 24
Sample proportion ($\hat{p}$)	Answers will vary, close to 0.6

Make a plot of the **sample distribution** where 1 is "loops" and 0 is "not loops".

Figure 2.

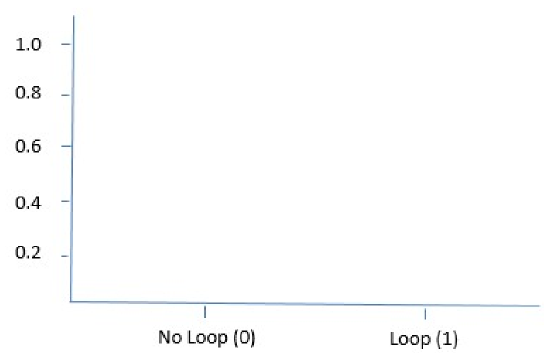


Student 2

	n = 40
Number of loops (x)	Answers will vary, close to 24
Sample proportion ($\hat{p}$)	Answers will vary, close to 0.6

Make a plot of the **sample distribution** where 1 is "loops" and 0 is "not loops".

Figure 3.



Before we go on to the next type of distribution, answer the following questions.

1. Identify the random variable and its possible values. X = Fingerprint type = {loop, not a loop}

2. Is the variable quantitative or categorical? Categorical

3. Should the graph of the population distribution (Figure 1) be the same or different for these two students? Explain. They should all be the same. The population distribution would not vary.

4. Should the graphs (Figure 2 and 3) of the sample distributions be the same or different for these two students? What if all students in the class had taken a sample, would the graphs have been the same or different? Explain.

   They will likely be different. Samples will vary from one to another.

5. Should the graphs in Part I (Figure 1) and Part II (Figures 2 & 3) be the same? Explain. No, Part I shows the population distribution, and Part II's graphs show sample distributions.

6. If every student in the class had taken a separate sample and computed the corresponding proportion for that sample, would these values be parameters or statistics?
   statistics

Discuss questions 1 –6 as a class.

Part III: Let's Explore the Sampling Distribution of the Sample Proportion by looking at the shape, mean, and variability of the sampling distributions.

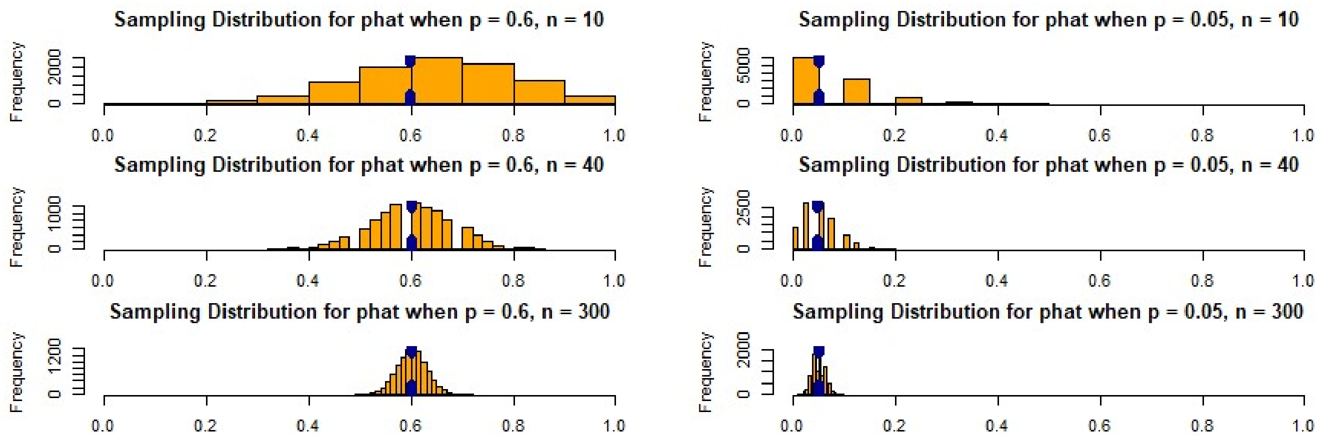
<u>Shape</u>

Suppose that we had 200 college students draw samples with paddles that had 10 beads (representing the fingerprints of 10 individuals), 40 beads, and 300 beads. Their sample proportions were then plotted below. Compare the shapes of the sampling distributions of the sample proportions for n = 10, n = 40, and n = 300 for both p = 0.05 and p = 0.6.

Sampling Distribution of a statistic- shows all possible values of the sample statistic and how often these values are expected to occur upon repeated sampling

Sampling Distribution of the sample proportion – probability distribution for the possible values of the sample proportion

Figure 4.



7. Using Figure 4, compare the shapes of the sampling distributions of the sample proportions when p = 0.6 for the sample sizes of 10, 40, and 300.
   They are all roughly bell shaped and symmetric.

8. Using Figure 4, compare the shapes of the sampling distributions of the sample proportions when p = 0.05 for the sample sizes of 10, 40, and 300. As the sample size increased, the distribution become more bell shaped and symmetric.

9. What does this suggest about the shape of sampling distributions of ANY sample proportion? What shape might they be?
   It looks like that they might be Normally distributed if the sample size is big enough.

Statisticians often use a Rule of Thumb to determine if the sampling distribution is approximately Normally distributed.

Rule of Thumb
The sampling distribution of the sample proportion is approximately Normal
if (answers will vary by textbook) np and n(1-p) greater than or equal to 15.

10. We saw that the sampling distribution of the sample proportion of arches (p=0.05) for n = 10 was right skewed. Verify that this case does not satisfy the above Rule of Thumb.
    10*0.05 = 0.5 and 10*0.95 = 9.5 The Rule of Thumb is not met. We should not assume that the sampling distribution of the sample proportion is Normally distributed.

11. We saw that the sampling distribution of the sample proportion of arches (p = 0.05) for n = 300 was an approximately Normal distribution. Verify that this case satisfies the above Rule of Thumb.
    300*0.05 = 15 and 300(1-0.05) = 285 Both of these are at least 15, so we can assume that the sampling distribution of the sample proportion is approximately Normally distributed.

Discuss questions 7 – 11 as a class.

<u>Mean</u>

12. A dotted line is shown at the mean for each of the graphs. Estimate the value of the mean and record in the table below.

	Mean		Mean
n =10, p = 0.60	Close to 0.60	n =10, p = 0.05	Close to 0.05
n =40, p = 0.60	Close to 0.60	n =40, p = 0.05	Close to 0.05
n =100, p = 0.60	Close to 0.60	n =100, p = 0.05	Close to 0.05

13. For p = 0.60: As the sample size increases, the value of the mean
    ☐Increases
    ☐Decreases
    ☐Stays approximately the same

14. For p = 0.05: As the sample size increases, the value of the mean
    ☐Increases
    ☐Decreases
    ☐Stays approximately the same

15. What does this suggest about the effect of sample size on the mean of sampling distributions of the sample proportion?

    The mean stays the same regardless of an increase or decrease in sample size.

16. Do we have a symbol for the mean of the sampling distribution of the sample proportion? If so, what is it? ______p______

Variability

Now let us talk about the variability in the sampling distributions of the sample proportion. Visually inspect the sampling distributions in Figure 4.

17. When p = 0.05, what happens to the variability as n increases from 10 to 300?
    - ☐Increases
    - ☐Decreases
    - ☐Stays approximately the same

    What characteristics of the graph influenced your decision?
    The sampling variability is less for the larger sample size (n = 300)

    The sample proportion values are more condensed. For p = .05 and n = 10, the values of the sample proportions range from about 0 to a little over 0.4. Where n =300, the values range from 0 to less than 0.1.

18. When p = 0.60, what happens to the variability as n increases from 10 to 300?
    - ☐Increases
    - ☐Decreases
    - ☐Stays approximately the same

    What characteristics of the graph influenced your decision?

    This also happened with p = .6. For p = .6 and n = 10, the values ranged from 0.2 to 1, but for n = 300, the values were between 0.5 and 0.7.

We saw in Figure 4 that as the sample size increased (beads per paddle), the variability in the distribution decreased. Mathematically, it can be proven that the standard deviation for the sampling distribution for a sample proportion is $\sqrt{\frac{p(1-p)}{n}}$. Let's see what this looks like for our sampling distributions.

19. Calculate the standard deviation for each of the three sampling distributions for p = 0.05.

- n=10: Standard Deviation = __0.0689________
- n=40: Standard Deviation = __0.0345________
- n=300: Standard Deviation = __0.0126________

As you can see, sample size can greatly reduce the standard deviation in a sampling distribution.

Discuss answers to questions 17 - 19.

Part IV. Next Steps

Throughout today's activity, we have used n to represent the sample size. This value represents how many fingerprints were examined or how many beads were on the paddle, not the number of times that the paddle was put into the bin to draw out the beads.

Is repeated sampling (drawing more than one set of beads from the bin) necessary for the properties we have been studying to hold true? The short answer is no, which is great because repeated sampling would most likely take too much time and money to complete.

Part V. Conclusion

Match the following terms with their definitions.

A. Sampling Distribution of the Sample Proportion
B. Population Distribution
C. Sample Distribution
D. Sample Proportion
E. Population Proportion

__B__ 1. Shows all possible values of a random variable and how often these values occur

__D__ 2. The statistic that represents the proportion of the group, randomly selected from the whole, with a particular characteristic of interest

__C__ 3. For the one sample collected, it shows all possible values of variable that occurred and how often these values occurred

__E__ 4. The parameter that represents the proportion of the whole group with a particular characteristic of interest

__A__ 5. The distribution of sample proportions shows all possible values of the sample proportion and how often these values occur with repeated sampling

Fill in the blanks.

6. The sampling distribution of the sample proportion has a mean of __p_____ and standard deviation of $\sqrt{\frac{p(1-p)}{n}}$

7. This distribution will have an approximately Normal distribution when the Rule of Thumb (np and n(1-p) greater than or equal to 15) is met.

8. As n increases, what happens to the mean of the sampling distribution of the sample proportion?
   ☐Increases
   ☐Decreases
   ☐Stays approximately the same

9. As n increases, what happens to the standard deviation of the sampling distribution of the sample proportion?
   ☐Increases
   ☐Decreases
   ☐Stays approximately the same

10. Suppose that Jose' wanted to take a sample of his classmates at his school and find the proportion of students who had fingerprints with loops. Would he need to take one random sample of 100 students or 1000 samples of 100 randomly selected students? Why?

    Jose' would only take one sample of 100 students. He can then compute the sample proportion from that sample. It would be impractical to take 1000 samples of 100 students. That would be 10,000 students!!

According to the WorldAtlas.com, 86% of all the people who live in Ireland and Scotland have green eyes. Sean, who did not know this value, took a random sample of 100 people living in Ireland and Scotland. For the 100 people that he selected, 85 had green eyes.

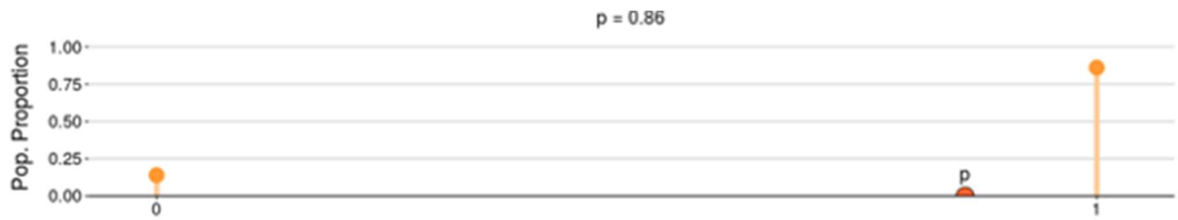
Match the images with the terms.

__A_ 11. Population distribution

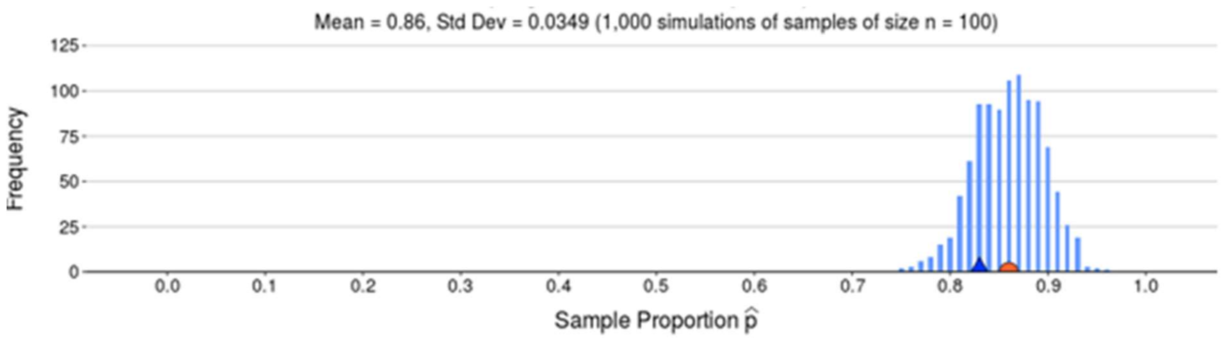
__C_ 12. Sample distribution

__B_ 13. Sampling distribution of the sample proportion

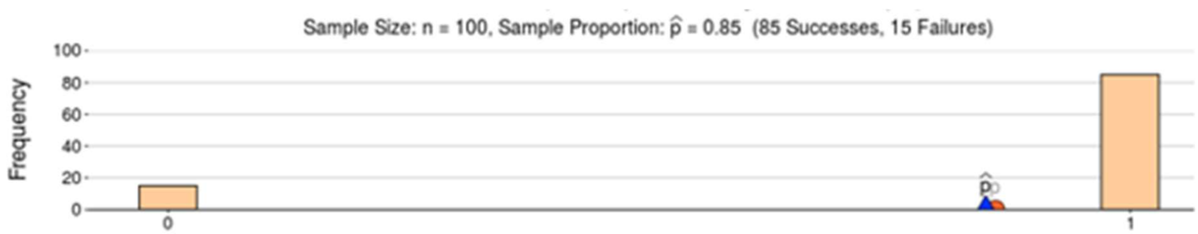
Graphic A



Graphic B



Graphic C



14. Calculate the mean and standard deviation for the sampling distribution of the sample proportion of green eyes given in the previous exercise.
The sampling distribution would be approximately Normal with a mean of 0.86 and a standard error equal to sqrt(0.86*(1-.86)/100) = 0.035.

Name: ____

Student Handout: Exploring the Sampling Distribution of the Sample Proportion with a Study of Fingerprint Types in the US

Question

Fingerprints are often used by law enforcement to fight crime and apprehend wrong doers. Each person has a unique fingerprint. If we look at the fingerprint of the thumb, there are three common characteristics: loops (lines that bend back on themselves), whorls (such as circles or ellipses), and arches (wave like patterns). In the United States, 60% of the population has loops, 35% whorls, and 5% arches.

Loop	Whorl	Arch
		

In this activity, we will answer the following questions:

- *What is the difference between population distribution, sample distribution, and sampling distribution?*
- *How do the mean, standard deviation, and shape for sampling distributions change as the sample size increases?*
- *Why is repeated sampling needed to understand the sampling distribution but not required for inference?*

Data

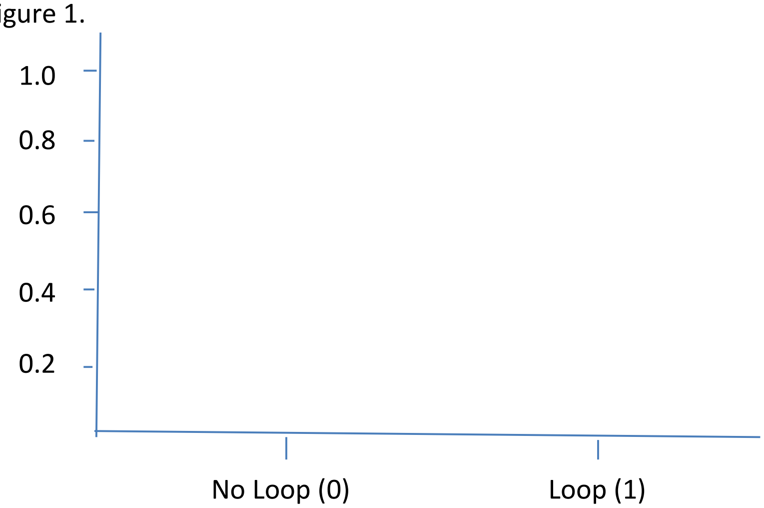
In this activity, we are going to explore three different types of distributions: *population distribution*, *sample (data) distribution,* and *sampling distribution* of the sample proportion.

Part I: Let's Explore the Population Distribution.

We will use a bin filled with beads to represent the American population. In the bin, you will find 60% of the beads are white, 35% of the beads are green, and 5% of the beads are black. This corresponds to loops, whorls, and arches, respectively. Imagine that you were to take out all the beads from the container and record if each bead represented a person with a loop or without a loop on their fingerprint.

- What is the population proportion of loops? ______
- Make a plot of the **population distribution** where 1 is "loops" and 0 is "not loops".

Figure 1.



Population - group of people or objects in which we are interested

Population Distribution - For the population, it shows all possible values of the variable and how often these values occur

Success – person or object displaying the characteristic of interest

Population Proportion (p) – the number of successes divided by the total number in the group in which you are interested (p = X/N)

Part II: Let's Explore the Sample (Data) Distribution.

Now we are going to explore the sample distribution by taking a sample of size 40 from the bin. We will have two students from the class do this. Calculate the sample proportion for each draw.

Sample - subset of the population for whom we have (or plan to have) data, often randomly selected

Sample Proportion – the number of successes divided by the number of outcomes in the sample ($\hat{p}$ = x/n)

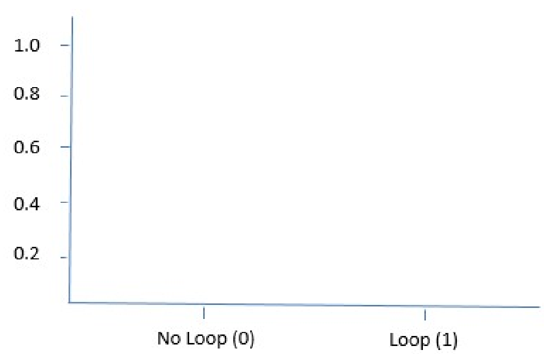
Sample (data) Distribution - For the one sample collected, it shows all possible values of the variable that occurred in the sample and how often these values occurred

Student 1

	n = 40
Number of loops (x)	
Sample proportion ($\hat{p}$)	

Make a plot of the **sample distribution** where 1 is "loops" and 0 is "not loops".

Figure 2.

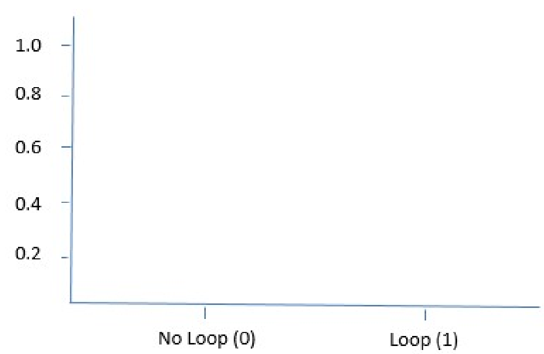


Student 2

	n = 40
Number of loops (x)	
Sample proportion ($\hat{p}$)	

Make a plot of the **sample distribution** where 1 is "loops" and 0 is "not loops".

Figure 3.



Before we go on to the next type of distribution, answer the following questions.

1. Identify the random variable and its possible values.

2. Is the variable quantitative or categorical?

3. Should the graph of the population distribution (Figure 1) be the same or different for these two students? Explain.

4. Should the graphs (Figure 2 and 3) of the sample distributions be the same or different for these two students? What if all students in the class had taken a sample, would the graphs have been the same or different? Explain.

5. Should the graphs in Part I (Figure 1) and Part II (Figures 2 & 3) be the same? Explain.

6. If every student in the class had taken a separate sample and computed the corresponding proportion for that sample, would these values be parameters or statistics?

Discuss questions 1 – 6 as a class.

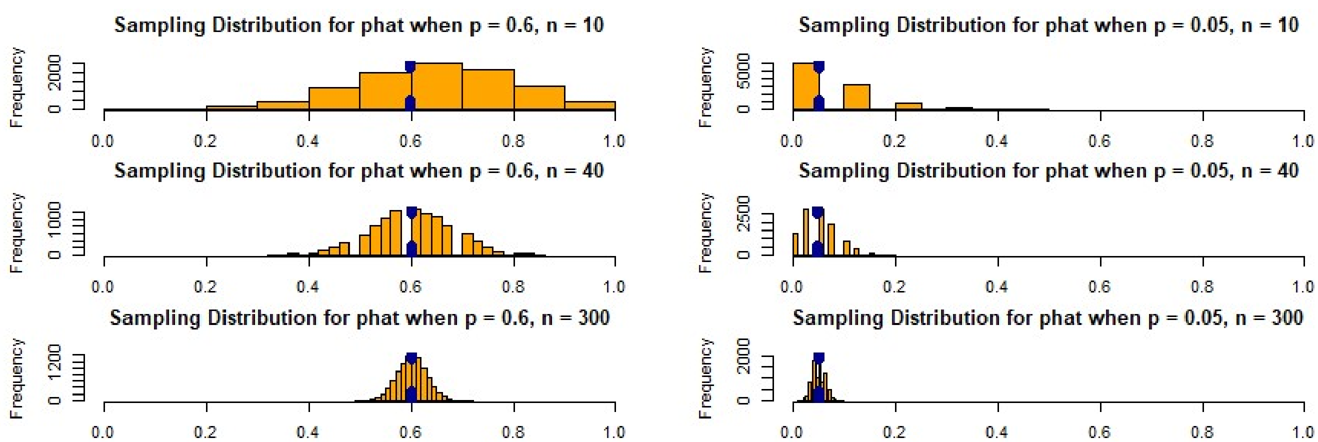
Part III: Let's Explore the Sampling Distribution of the Sample Proportion by looking at the shape, mean, and variability of the sampling distributions.

<u>Shape</u>

Suppose that we had 200 college students draw samples with paddles that had 10 beads (representing the fingerprints of 10 individuals), 40 beads, and 300 beads. Their sample proportions were then plotted below. Compare the shapes of the sampling distributions of the sample proportions for n = 10, n = 40, and n = 300 for both p = 0.05 and p = 0.6.

> *Sampling Distribution of a statistic*- shows all possible values of the sample statistic and how often these values are expected to occur upon repeated sampling
>
> *Sampling Distribution of the sample proportion* – probability distribution for the possible values of the sample proportion

Figure 4.



7. Using Figure 4, compare the shapes of the sampling distributions of the sample proportions when p = 0.6 for the sample sizes of 10, 40, and 300.

8. Using Figure 4, compare the shapes of the sampling distributions of the sample proportions when p = 0.05 for the sample sizes of 10, 40, and 300.

9. What does this suggest about the shape of sampling distributions of ANY sample proportion? What shape might they be?

Statisticians often use a Rule of Thumb to determine if the sampling distribution is approximately Normally distributed.

Rule of Thumb
The sampling distribution of the sample proportion is approximately Normal if:

10. We saw that the sampling distribution of the sample proportion of arches (p=0.05) for n = 10 was right skewed. Verify that this case does not satisfy the above Rule of Thumb.

11. We saw that the sampling distribution of the sample proportion of arches (p = 0.05) for n = 300 was an approximately Normal distribution. Verify that this case satisfies the above Rule of Thumb.

Discuss questions 7 – 11 as a class.

Mean

12. A dotted line is shown at the mean for each of the graphs. Estimate the value of the mean and record in the table below.

	Mean		Mean
n =10, p = 0.60		n =10, p = 0.05	
n =40, p = 0.60		n =40, p = 0.05	
n =100, p = 0.60		n =100, p = 0.05	

13. For p = 0.60: As the sample size increases, the value of the mean
    ☐Increases
    ☐Decreases
    ☐Stays approximately the same

14. For p = 0.05: As the sample size increases, the value of the mean
    ☐Increases
    ☐Decreases
    ☐Stays approximately the same

15. What does this suggest about the effect of sample size on the mean of sampling distributions of the sample proportion?

16. Do we have a symbol for the mean of the sampling distribution of the sample proportion? If so, what is it? ________

Variability

Now let us talk about the variability in the sampling distributions of the sample proportion. Visually inspect the sampling distributions in Figure 4.

17. When p = 0.05, what happens to the variability as n increases from 10 to 300?
    ☐Increases
    ☐Decreases
    ☐Stays approximately the same

    What characteristics of the graph influenced your decision?

18. When p = 0.60, what happens to the variability as n increases from 10 to 300?
    - ☐Increases
    - ☐Decreases
    - ☐Stays approximately the same

    What characteristics of the graph influenced your decision?

We saw in Figure 4 that as the sample size increased (beads per paddle), the variability in the distribution decreased. Mathematically, it can be proven that the standard deviation for the sampling distribution for a sample proportion is $\sqrt{\frac{p(1-p)}{n}}$. Let's see what this looks like for our sampling distributions.

19. Calculate the standard deviation for each of the three sampling distributions for p = 0.05.

- n=10: Standard Deviation = ___________
- n=40: Standard Deviation = ___________
- n=300: Standard Deviation = __________

As you can see, sample size can greatly reduce the standard deviation in a sampling distribution.

Discuss answers to questions 17 - 19.

Part IV. Next Steps

Throughout today's activity, we have used n to represent the sample size. This value represents how many fingerprints were examined or how many beads were on the paddle, not the number of times that the paddle was put into the bin to draw out the beads.

Is repeated sampling (drawing more than one set of beads from the bin) necessary for the properties we have been studying to hold true? The short answer is no, which is great because repeated sampling would most likely take too much time and money to complete.

Part V. Conclusion

Match the following terms with their definitions.

A. Sampling Distribution of the Sample Proportion
B. Population Distribution
C. Sample Distribution
D. Sample Proportion
E. Population Proportion

____ 1. Shows all possible values of a random variable and how often these values occur

____ 2. The statistic that represents the proportion of the group, randomly selected from the whole, with a particular characteristic of interest

____ 3. For the one sample collected, it shows all possible values of variable that occurred and how often these values occurred

____ 4. The parameter that represents the proportion of the whole group with a particular characteristic of interest

____ 5. The distribution of sample proportions shows all possible values of the sample proportion and how often these values occur with repeated sampling

Fill in the blanks.

4. The sampling distribution of the sample proportion has a mean of _______ and standard deviation of:

5. This distribution will have an approximately Normal distribution when the Rule of Thumb

   ____ is met.

6. As n increases, what happens to the mean of the sampling distribution of the sample proportion?

   ☐Increases

   ☐Decreases

   ☐Stays approximately the same

7. As n increases, what happens to the standard deviation of the sampling distribution of the sample proportion?

   ☐Increases

   ☐Decreases

   ☐Stays approximately the same

8. Suppose that Jose' wanted to take a sample of his classmates at his school and find the proportion of students who had fingerprints with loops. Would he need to take one random sample of 100 students or 1000 samples of 100 randomly selected students? Why?

According to the WorldAtlas.com, 86% of all the people who live in Ireland and Scotland have green eyes. Sean, who did not know this value, took a random sample of 100 people living in Ireland and Scotland. For the 100 people that he selected, 85 had green eyes.

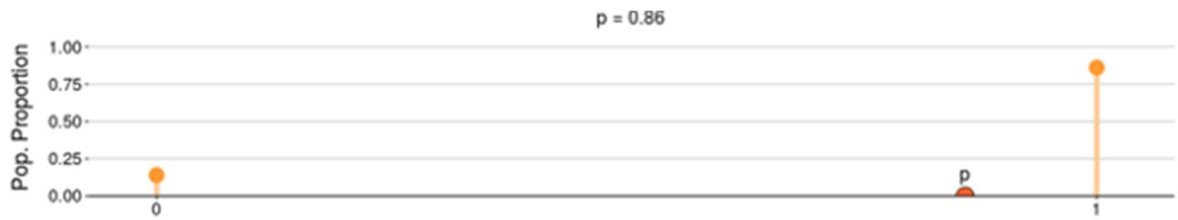
Match the images with the terms.

_____ 11. Population distribution

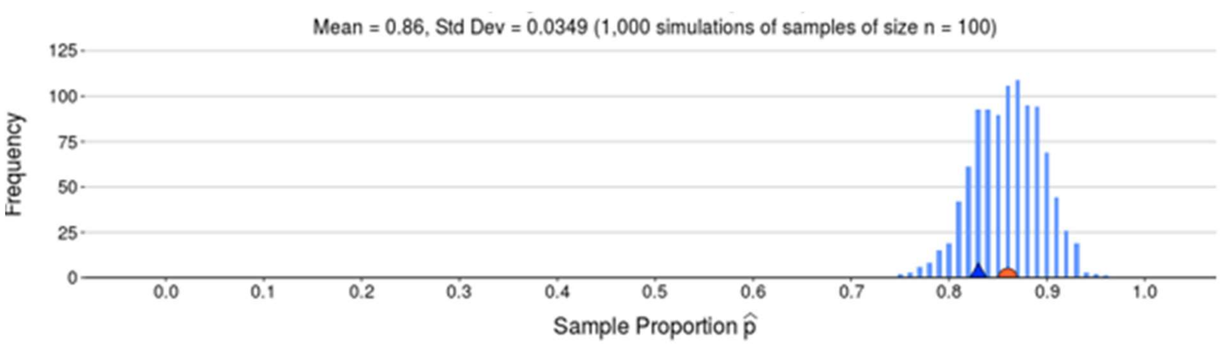
_____ 12. Sample distribution

_____ 13. Sampling distribution of the sample proportion

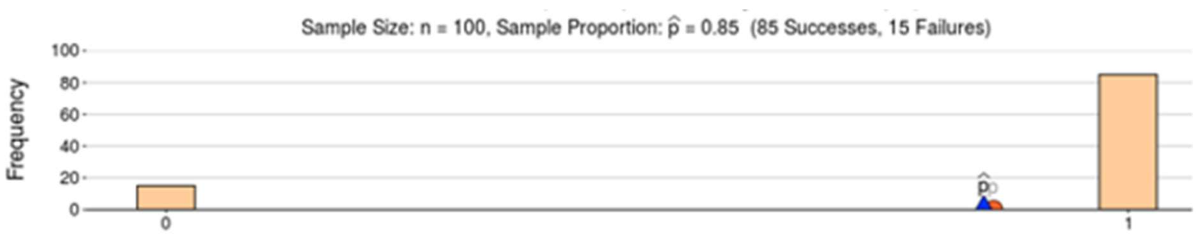
Graphic A



Graphic B



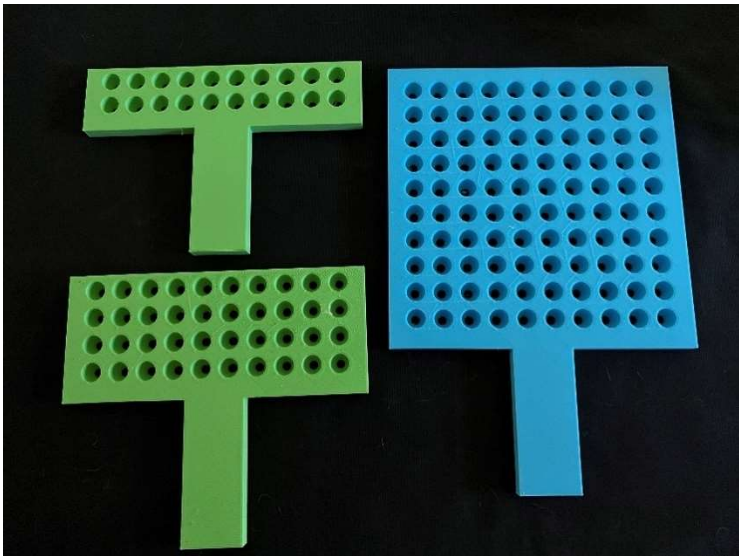
Graphic C



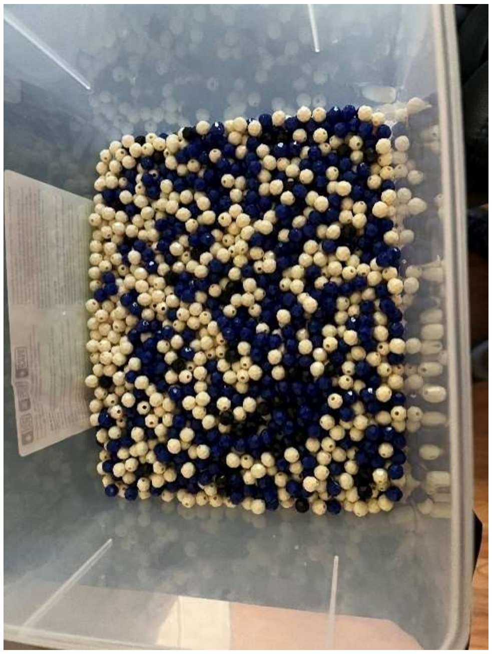
14. Calculate the mean and standard deviation for the sampling distribution of the sample proportion of green eyes given in the previous exercise.

Appendix

Paddles
The paddles can be creating using a 3D printer. If your school doesn't have one, check with a local library or neighboring college/university. The instructions for the 3D printing are also posted with this activity. You can either create a 3D printer with 100 holes or two sets of paddles (20 and 100 holes). If you use the 100-hole paddles, you can then use masking tape to cover up any unneeded holes.



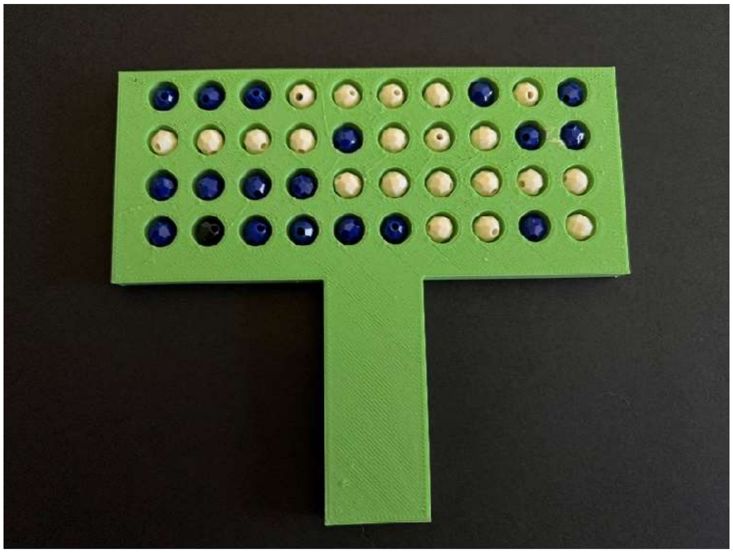
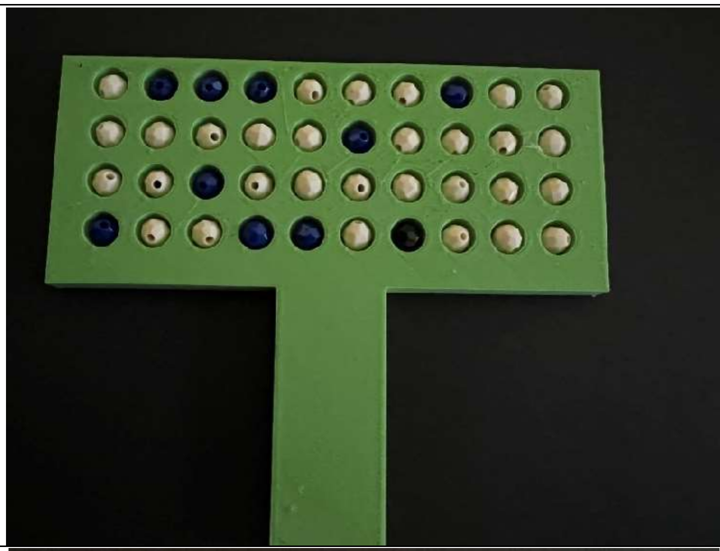
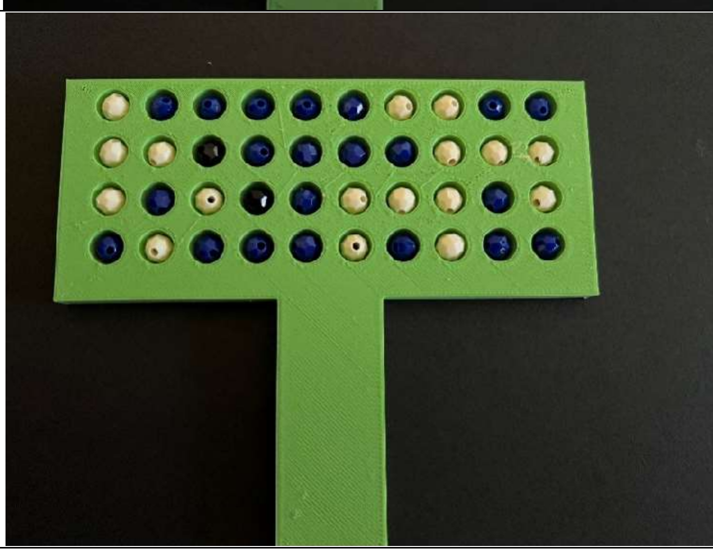
Beads
8mm beads fit into the paddles well. I purchased them through an online retailer so that I can get mass quantities for a more reasonable price. I then put the beads into a Tupperware container with a handle. In this case, I have 60% white, 35% blue and 5% black.



Sampling Examples

Here is one example sample from the bin with a sample of 40 beads. In this sample, there were 4 black beads.

Other examples of samples where n = 40.

	1 black bead out of the 40
	1 black bead out of the 40
	2 black beads out of 40