Exploring Fundamental Definitions with a Study of Fingerprint Types in the US

Katina Gothard, Eastern Florida and Megan Mocko, University of Florida
Published: October 2022

Overview of Lesson

Students will randomly sample beads out of a bin to simulate random samples of people's fingerprints. Students will see that the sample proportion of fingerprints with arches (one of the three common fingerprint patterns) varies from sample to sample. The purpose is to guide students in the development of, and identify any misconceptions about, the definitions of *population*, *parameter*, *sample*, *statistic* and *random*. In this lesson, the students will draw simple random samples, calculate sample proportions, construct dotplots and identify a population, parameter, sample, and statistic. By the end of the activity, students will be able to define these terms and describe what random means in the context of sampling.

Type of Data

- One categorical variable
- Data collected as a class

Learning Objectives

After completing this lesson, the student will be able to

- define the terms: sample, random, statistic, population, and parameter.
- identify the difference between the distribution of categorical data and the sampling distribution for a sample proportion.
- construct the sampling distribution of a sample proportion.
- describe the shape, location of the center, and symmetry of a sampling distribution.
- compare sampling distributions of sample proportions for different sample sizes.
- CCSS.MATH.CONTENT.HSS.ID.A.1: "Summarize, represent, and interpret data on a single count or measurement variable."
- CCSS.MATH.CONTENT.HSS.IC.A.1: "Understand and evaluate random processes underlying statistical experiments."

Audience

- Students in a first semester of statistics with minimal if any background.
- *Prerequisites:* This activity should be done within the first two weeks of the class; therefore, only basic math skills are required. Students will need to know the difference between categorical and quantitative variables. Although not required, it would be helpful to have heard the terms sample, population, and random sampling.

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/blue, 5% of the beads are black)
  - Note: To ensure color-blind 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 various sizes (n = 20, n = 100)
  - 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.
- Draw an axis on the board or on a poster board.
- Sticky notes, magnets or stickers, 2 per person

  For small classes, access to the web might be needed for the instructor before class. Otherwise, no technology is needed.

Additional Resources

- We briefly introduce the Investigative Cycle, which is the framework used for this activity. For more information about how to introduce the investigative cycle to your class, see the following resources:

  https://dataschools.education/about-data-literacy/ppdac-the-data-problem-solving-cycle/
  https://new.censusatschool.org.nz/wp-content/uploads/2012/11/data-detective1.pdf
  Velleman, Paul. (2008). Math is Music; Statistics is Literature. Amstat News. 375, 54.

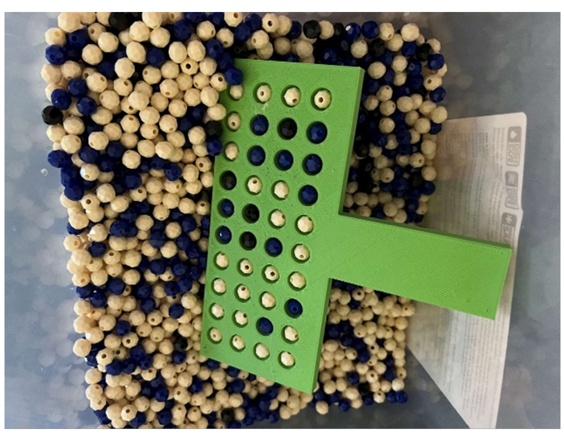
Lesson Plan

The idea of the lesson is to introduce students to the fundamental statistical terms on the first day of class and have the students witness the idea of sampling variability. This example can then be used repeatedly as an example of key terms, such as sample, statistic, population, and parameter throughout the semester. In general, this activity is designed to have the students complete a guided exploration on their own or in a small group. The students should be encouraged to read through the handout and to begin to think about the concepts on their own before being told the answers.

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 Appendix B. 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 portion, the students will then discuss as a class. The activity is marked with stop signs so that the students know to wait for the class discussion. There is also one section in which the instructor gives a short lecture before the students proceed with the material. The lesson is broken down into three stages: Question and Data, Analysis and Conclusions.

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.

Question and Data

1. The instructor has the students form pairs and triads. The instructor encourages the students to introduce themselves to each other.
2. The instructor instructs the students to read the first page of the student handout and discuss the first two questions. The students will be discussing the concept of randomness as well as how to take a sample from a population.
3. After five minutes, the instructor has a volunteer report out their discussion about the first two questions. As the students report out, make sure to clarify that randomness does not mean haphazard and to make sure that the students are distinguishing between a sample and a population correctly.
4. The instructor explains about the three different types of fingerprints and has the students examine example images. The instructor than encourages the students to look at their own fingers and determine which type of fingerprint they may have.
5. The instructor will then introduce the three primary questions of the activity verbally and direct students to where it is written in the handouts.
    - *If you were to take several random samples of American adults, would the sample proportion of individuals with <u>arches</u> change from sample to sample? Why is this?*
    - *What would the most typical value of the sample proportions be? Why does this make sense?*
    - *How do the shape and center of a sampling distribution for a proportion change as the sample size is increased?*
6. The instructor demonstrates that each bead represents someone's fingerprint. The white beads represent loops, the green (or blue) represent whorls and the black beads represent arches. The instructor then dips the paddle deep into the bin and pulls out a sample. The paddle will include beads of different colors in each of the holes. (See Appendix B for pictures.) The instructor then demonstrates counting the number of black beads. The instructor also demonstrates tipping the paddle over so that the beads softly land back in the bin.
7. The instructor should clearly define the bin as the population, and the selected beads with the paddle as a sample.
8. The instructor asks the students how to compute a proportion from this example sample and shows the correct way to compute this on the board or overhead. Explain that the number of individuals with arches would be x and the sample size is n, the number of individuals.

Suppose that the sample had 2 black beads, the sample proportion would be $x/n = 2/20 = 0.10$. The instructor should state that this value is a statistic and explain how this is different from a parameter.

9. The instructor starts the students on the Data section of the activity (questions 1-3). The students discuss why it might be reasonable to take a random sample of beads rather than look at all the beads, then they start the process of randomly sampling from the bins using the paddles. They will take two samples: one of size 20 and one of size 100. It is important for the instructor to help organize the sampling procedure to make the best use of time. It is best to place the bins throughout the room so that students can easily get access to one of the bins, rather than getting congested all at one location. Depending on how many bins you have, you can either allocate bins to a particular group or you can have groups share a bin. Occasionally, a paddle will come up with a bead missing from a hole. Instruct the students to re-sample the beads again. The students should draw the bar chart of their own sample of size 20 and size 100 on the handout and compute the sample proportions. The students will note whether they got the same p-hat value every time and whether this variability would also be observed in a real-world sampling context.
10. After the students have completed questions 1-3, have different volunteers report out what they found when sampling. Make sure to emphasize that each group didn't get that same sample proportion and that these results demonstrate sampling variability. The instructor should assure students that this is expected and not a mistake.

Analysis

1. The instructor leads the process of having the students plot their proportions on the board using the sticky notes, stickers, or magnets. Another method for recording these results is for the instructor to have students raise their hand for certain values. This process is often faster. The instructor can also use a large poster board and keep it around for further discussion later in the term. If the class is small, the instructor might need to help add additional points to see a more complete display of values.

   If you are implementing this activity in a small class ($n < 50$), you will need to supplement the samples. At least 50 samples (between yours and the students) are recommended for each graph. A quick way to create extra samples:
   - Click on http://www.lock5stat.com/StatKey/
   - Next to **Sampling Distributions**, select Proportions
   - Click Edit Proportion and enter 0.6
   - Click Choose samples of size n = and enter the paddle size (20 or 100)
   - Click Generate 10 samples 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. The instructor asks the students to replicate the drawing onto their handout and answer questions 4 – 9. The purpose of these questions is to illustrate the differences between the bar chart and the dotplot. Specifically, the bar chart (Figure 1) shows the results of one sample, but the dotplot (Figure 3) shows the sample proportions from many samples.
3. After 10 minutes, the instructor asks for volunteers to share their results. The instructor clarifies the differences between the bar chart and the dotplot.

4. The students should discuss questions 10 – 11 with their group. These questions ask students what values of $\hat{p}$ they should reasonably expect for difference sample sizes and why the sample proportion of arches is rarely above 0.20. These two questions can be quite challenging. If students get stuck on number 10, the instructor can hint for the students to look at the Figure 3 where n = 20 and Figure 4 where n=100 for ideas. If students get stuck on number 11, the instructor can hint that they examine the initial proportions of those who had arches on their fingerprints and see if this is related to their results in Figures 3 where n = 20 and 4 where n =100.
5. Because students will see skewed and bell-shaped sampling distributions, we took this as an opportunity to introduce four basic shapes: bell-shaped, uniform, skewed-left, and skewed-right, along with some basic properties. Because the students only see sampling distributions, we intentionally did not define <u>sample distribution</u> in this activity. We plan to introduce it at a later point in the class. The instructor lectures on the definition of a distribution and discusses four types of distributions whose images are already presented on the handout (bell shaped, right skewed, left skewed and uniform). One method of doing this would be to have an overhead of each of the graphics presented in this section. The filled-in handout below includes notes for the instructor's lecture.
6. The instructor continues to lecture about the concepts of center, symmetry, tails, and sampling distribution. As you define each of the properties of center, symmetry, tail(s), and sampling distribution of a statistic have the students locate/mark them on each of the above distributions.
7. After finishing the direct instruction, the instructor marks the centers, symmetry, and tails on the graphs of the four types of distributions described before.
8. The instructor tells the groups to answer questions 12 and 13 for about 5 minutes. These questions ask students to describe the shape, location of center, and symmetry for each of the sampling distributions (n = 20 and n=100). After the five minutes, the instructor can ask a group to report that hasn't yet about the answers to these two questions. It is important that each of the three characteristics are described for both figures 3 and 4.

Conclusion

1. As a pair or triad, the students should answer the questions 14 - 17 in the conclusions up to the STOP sign in about 5 minutes. These questions ask students to revisit the questions posed at the start of the activity and consider whether the center and shape of the sampling distribution changed based on the sample size.
2. The instructor can then call on any groups that haven't volunteered yet or ask for volunteers again if everyone has provided an answer today. For these questions, the instructor should emphasize that these graphs represent the results of many samples, not just one. If the students struggle with this, the instructor can revisit the plots made on the board and ask what each magnet, sticker, dot, or sticky note represents.
3. Additionally, emphasize that for these types of graphs, the centers remained the same regardless of the sample size; however, shape and symmetry did change.
4. We decided not cover variability in this activity because we felt that it was already complex enough. However, you may choose to include a discussion about variability at this point. Alternatively, you could postpone the discussion of variability but refer back to this example later in the course.

5. For the final five minutes, the instructor tells the students to define the terms *random sampling, sample, statistic, population,* and *parameter* using what they did in the activity. Students often struggle with the term parameter. Be prepared to use the bins to help illustrate the difference between sample, statistic, population, and parameter. It is also a good time to remind students that the parameter didn't change throughout the entire activity. If you would like to construct the definitions that follow as a class, consider the following:

    You teacher will provide you with a stickynote. On the sticky note provide a definition for each of these terms. As a class, you will then combine these ideas to write the definitions.

    OR

    Time permitting: If your students have access to the internet, make a copy of this Jamboard. https://jamboard.google.com/d/1_u5hPEYxhB2o-NjwbxJYTvHvyn9ow2jSsBGjakRFUUA/edit?usp=sharing
    The students can then select the sticky note icon (write under the arrow button) and then place a virtual sticky note on word's page. The instructor can then show everyone's sticky notes to the class and the class can crafter the definition.

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 Fundamental Definitions with a Study of Fingerprint Types in the US

Investigative Cycle: When conducting any research, you will need to follow an investigative cycle. In general, an investigative cycle includes *asking a question*, *collecting data* to answer that question, *analyzing the data* and then *making conclusions*. This activity will follow this format.

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.

a. What does "randomness" mean?

b. How would you take a sample from a population?

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
		

Here are some actual prints. Can you correctly identify the shape of each?

___whorl_________ ___arches________ ____loop___________

In this activity, we will address the following questions:

- *If you were to take several random samples of American adults, would the sample proportion of individuals with arches change from sample to sample? Why is this?*
- *What would be the most typical value of the sample proportions? Why does this make sense?*
- *How do the shape and center of a sampling distribution for a proportion change as the sample size is increased?*

We will use a bin filled with beads to represent the American population and explore these questions.

Data

Collecting a sample of beads:

1. Why might it be reasonable to take a random sample of beads rather than look at all the beads?

   Answers will vary – too many beads to count, not enough time, etc.

In the bin of beads, each bead represents an individual with a certain type of pattern on their thumb: White for loops, Green for whorls, and Black for arches. We are going to collect a sample of 20 beads and 100 beads using the given paddles.

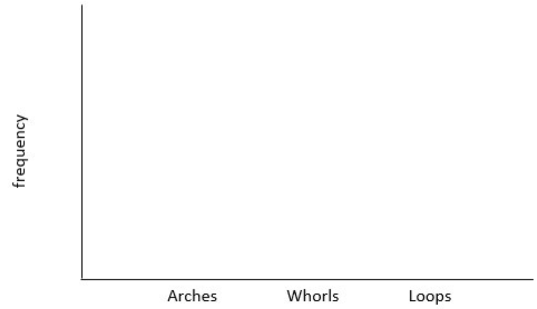
Draw a sample of size n=20 and a sample of size n=100. Complete the table provided below.

	n = 20	**n = 100**
x = Number of "Arches"		
n = Sample Size	**20**	**100**
Sample Proportion of "Arches" from your sample ($\hat{p} = x/n$)		

Now construct a bar graph to display the results of each sample.

Figure 1. *Results for n=20*

Figure 2. *Results for n=100*



Let's continue to summarize our findings with some descriptive statistics. Complete the third row of the above table.

2. Draw three more samples for n=20 and calculate their associated p-hats. Do you get the same value every time? Why is this?

   No, because of the randomness of the selection process

3. If you were to take several random samples of American adults, would the sample proportion of individuals with arches change from sample to sample? Why is this?

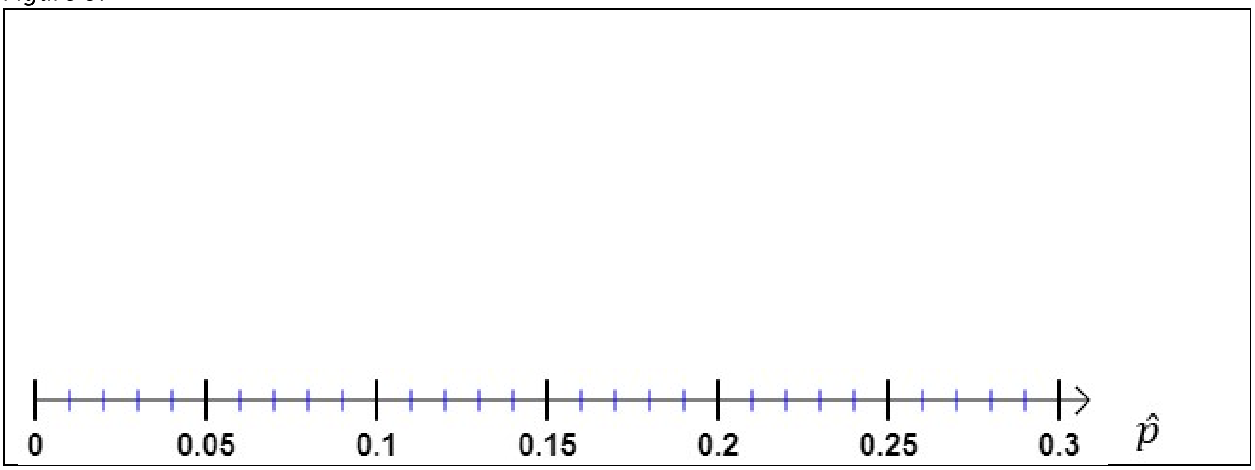
   Yes, because we have a new group of people. Randomness of the selection process makes it possible to have a slightly different proportion each time.

Discuss questions 1-3 as a class.

Analysis
Let's look at the results of everyone in the class. Your instructor will put two axes on the board. Make a dot for each proportion of arches from your samples of size n = 20 and n = 100. Recreate the plots below.

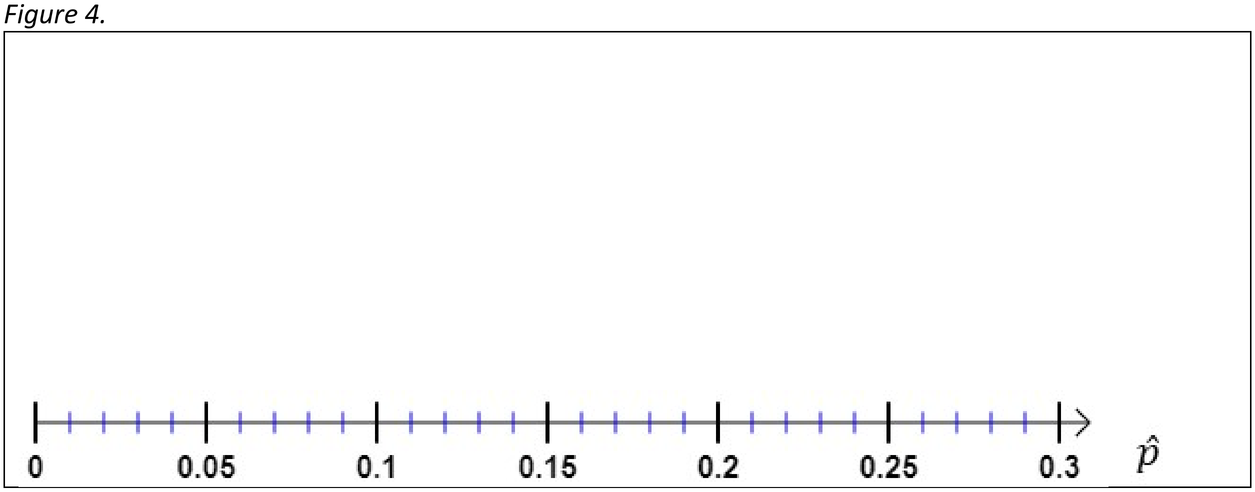
n = 20

Figure 3.



n = 100

Figure 4.



Let's now compare the features of the graphs in Figures 1 & 3.

For Figure 1:
4. What does the x-axis represent?

The variable X = fingerprint type

5. What does the y-axis represent?

The number of people with each type of fingerprint (frequency)

6. Does this graph show the result of one sample or results from many samples?

One sample

For Figure 3:

7. What does the x-axis represent?

The values of the sample proportions (p-hat's) for the samples

8. Does this graph show the result of one sample or results from many samples?

Many samples

9. What does each dot on the graph represent?

Each dot represents the sample proportion for <u>one</u> sample.

<u>Discuss questions 4-9 as a class.</u>

10. What values of $\hat{p}$ should you reasonably expect when sampling? [Consider Figures 3 & 4.]

For n=20 at a time: [0, 0.10]

For n=100 at a time: [0.03, 0.07] – closer to the actual value of the parameter

11. Why do you think the sample proportions we calculated for each iteration (draw) were rarely (if ever) above 0.20?

The sample proportions were rarely above 0.20 because the value of the population parameter is 0.05.

Discuss questions 10-11 as a class.

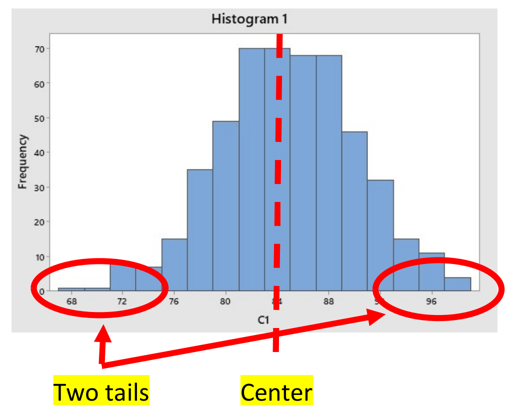
Let's establish some definitions:

Distribution –

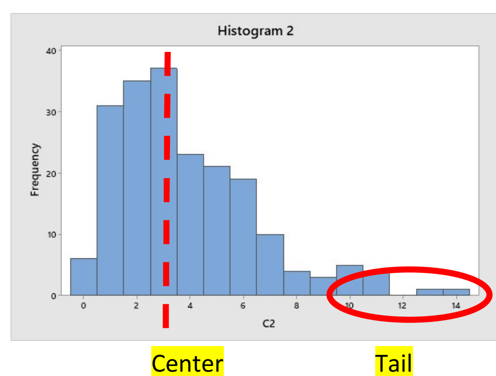
A way to organize the observed values in a sample or samples, along with the frequency of each value

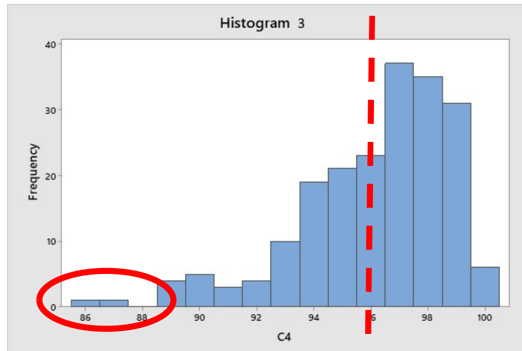
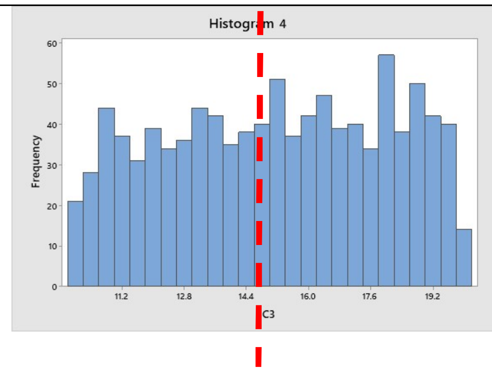
Certain distributions tend to show up often. Here are a few:

Bell-shaped
Symmetric



Skewed-right (positively skewed)
Not symmetric
One tail



Skewed-left (negatively skewed) Not symmetric One tail	 Tail Center
Uniform Symmetric No tails	 Center

Let's define the following properties and identify them on each of the above graphs:

Center (graphically) –

Value on the x axis at which half of the data is to the left and the other half is to the right, graphical estimation of the median

Symmetry –

A graph in which there is a point on the x axis at which the left- and right sides are mirror images of each other)

Tail(s) –

Where a severe drop-off happens in the frequency of the data, will be on the far left and/or far right side of the graph

Sampling distribution of a statistic –

Distribution of a statistic from repeated sampling – the two graphs we created in figures 3 & 4 display sampling distributions for p-hat

Important Note: The above properties are not used to describe graphs that display distributions of categorical or qualitative data (x-axis shows categories), as shown in the bar graphs we created in Figures 1 & 2.

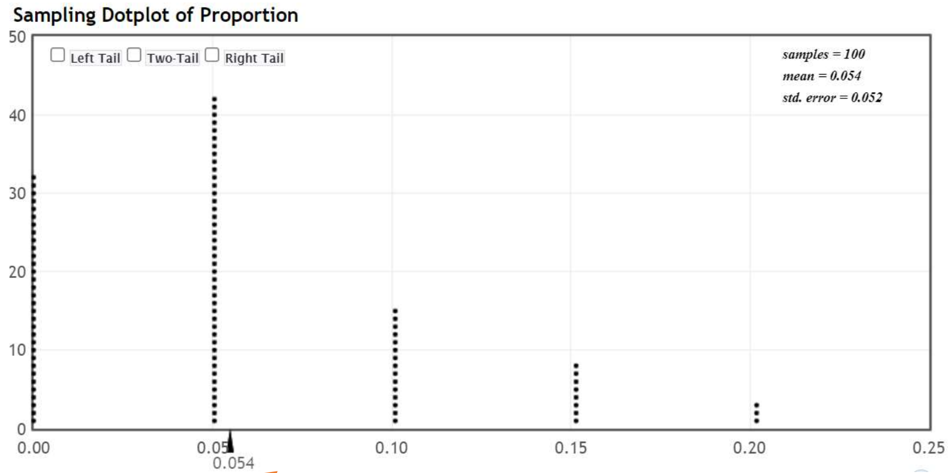
These properties are only used to describe graphs that display distributions of quantitative data (x-axis is a number line). Examples of acceptable graphs include the four histograms provided above and the dotplots we created in Figures 3 through 6 to display our sampling distributions.

Below are the results from using *StatKey* to run 100 samples for the two sample sizes, n=20 and n=100.

Comparing these graphs to the ones we constructed for our earlier samples (Figures 3 & 4), we see that as the number of samples is increased, the shape becomes more clear in each sampling distribution.

Figure 5.

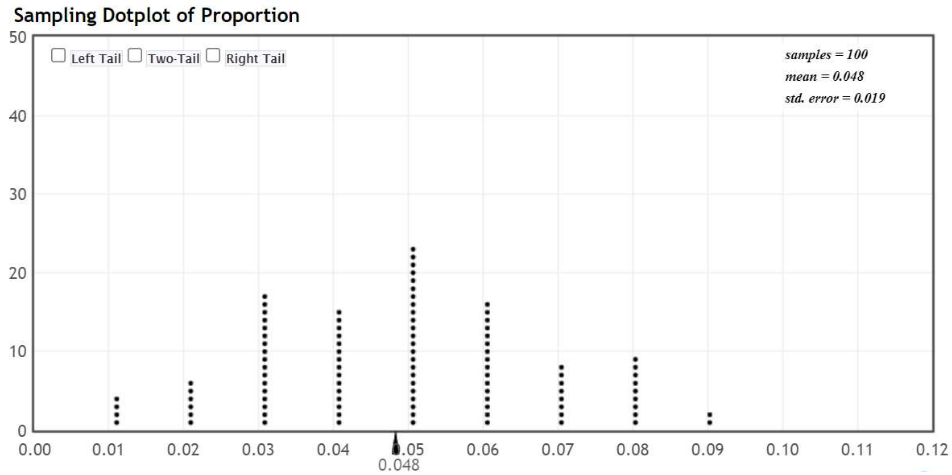
n = 20:



NOTE: The 0.054 shown below the x-axis is the mean of the 100 p-hats in this sampling distribution.

Figure 6.

n = 100:



NOTE: When comparing Figures 5 & 6 to our sampling distributions in Figures 3 & 4, we see that as the number of runs is increased, the shape becomes more pronounced in each sampling distribution.

12. Describe the shape, location of center, and symmetry (if any) of the sampling distribution for n=20 (Figure 5).

    The sampling distribution for n=20 is skewed right with a center at 0.05 and is not symmetric.

13. Describe the shape, location of center, and symmetry (if any) of the sampling distribution for n=100 (Figure 6).

    The sampling distribution for n=100 is bell-shaped and symmetric, with a center at 0.05.

Discuss questions 12-13 as a class.

Conclusion

Let's return to the questions posed at the start of this activity.

14. If you were to take several random samples of American adults, would the sample proportion of individuals with arches change from sample to sample? Why is this?

    Yes, the same people will not be selected from sample to sample due to the random nature of the sampling method; therefore, the sample proportion will likely be different each time.

15. What would be the most typical value of the sample proportions? Why does this make sense?

    If the sample is representative of the population, the 0.05 should be most typical, because 0.05 is the value of the population parameter.

16. What property was the same in the n = 20 and n = 100 sampling distributions?

    The center was roughly 0.05 for both sampling distributions. Since this is a simulation and we don't have every possible sample, the center for the simulated sampling distribution may be slightly off from the theoretical mean of 0.05.

17. What properties changed in the n = 20 and n = 100 sampling distributions?

    The shape and associated symmetry changed.

Discuss questions 14-17 as a class.

18. Describe each of these terms **in the context of this investigation**.

    a. *random* sampling

       the use of a chance process to ensure each person had an equal chance of being selected

    b. sample

       n=20 (or n=100) people in the United States

    c. *statistic* (give a written description and the associated symbol)

       proportion of people from the sample whose fingerprint shapes are arches: $\hat{p}$ "p-hat"

    d. population

       All people in the United States

    e. *parameter* (give a written description, the associated symbol, and the value)

       proportion of people from the population whose fingerprint shapes are arches
       p = 0.05

Exploring Fundamental Definitions with a Study of Fingerprint Types in the US

Investigative Cycle: When conducting any research, you will need to follow an investigative cycle. In general, an investigative cycle includes *asking a question*, *collecting data* to answer that question, *analyzing the data* and then *making conclusions*. This activity will follow this format.

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.

a. What does “randomness” mean?

b. How would you take a sample from a population?

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
		

Here are some actual prints. Can you correctly identify the shape of each?

________________ ________________ ________________

In this activity, we will address the following questions:

- *If you were to take several random samples of American adults, would the sample proportion of individuals with arches change from sample to sample? Why is this?*
- *What would be the most typical value of the sample proportions? Why does this make sense?*
- *How do the shape and center of a sampling distribution for a proportion change as the sample size is increased?*

We will use a bin filled with beads to represent the American population and explore these questions.

Data

Collecting a sample of beads:

1. Why might it be reasonable to take a random sample of beads rather than look at all the beads?

In the bin of beads, each bead represents an individual with a certain type of pattern on their thumb: White for loops, Green for whorls, and Black for arches. We are going to collect a sample of 20 beads and 100 beads using the given paddles.

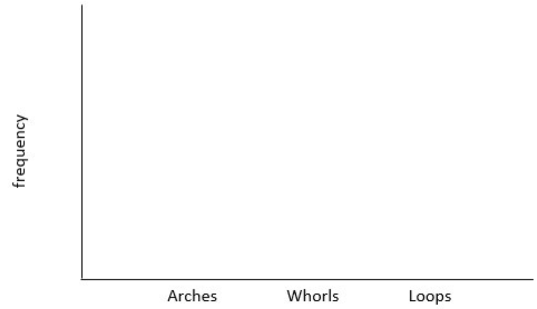
Draw a sample of size n=20 and n=100. Complete the table provided below.

	n = 20	**n = 100**
x = Number of "Arches"		
n = Sample Size	**20**	**100**
Sample Proportion of "Arches" from your sample ($\hat{p} = x/n$)		

Now construct a bar graph to display the results of each sample.

Figure 1. *Results for n=20*

Figure 2. *Results for n=100*



Let’s continue to summarize our findings with some descriptive statistics.

2. Draw three more samples for n=20 and calculate their associated p-hats. Do you get the same value every time? Why is this?

3. If you were to take several random samples of American adults, would the sample proportion of individuals with <u>arches</u> change from sample to sample? Why is this?

<u>Discuss questions 1-3 as a class.</u>

Analysis
Let's look at the results of everyone in the class. Your instructor will put two axes on the board. Make a dot for each proportion of arches from your samples of size n = 20 and n = 100. Recreate the plots below.

Figure 3. **n = 20**

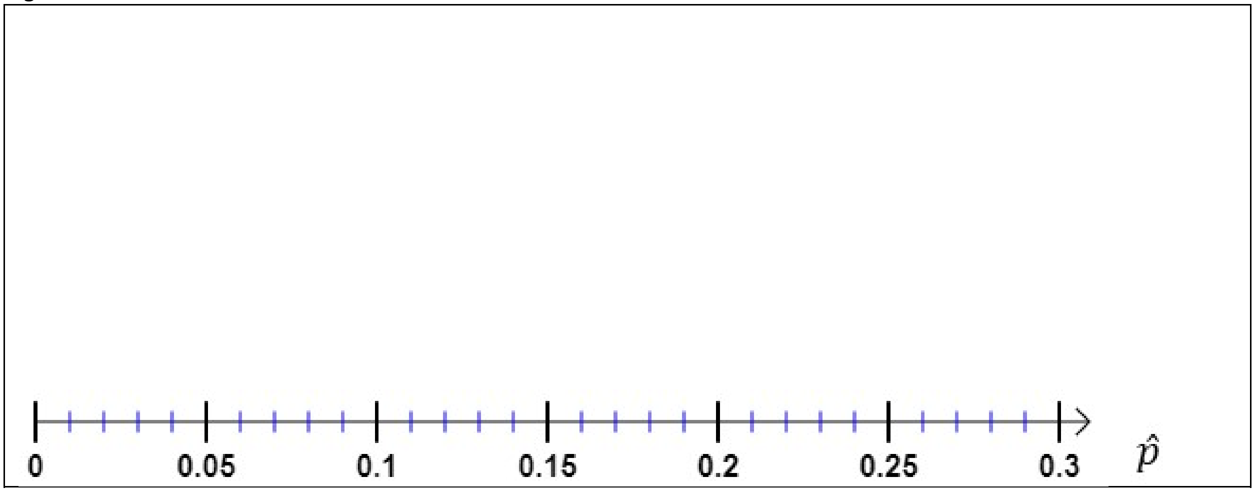
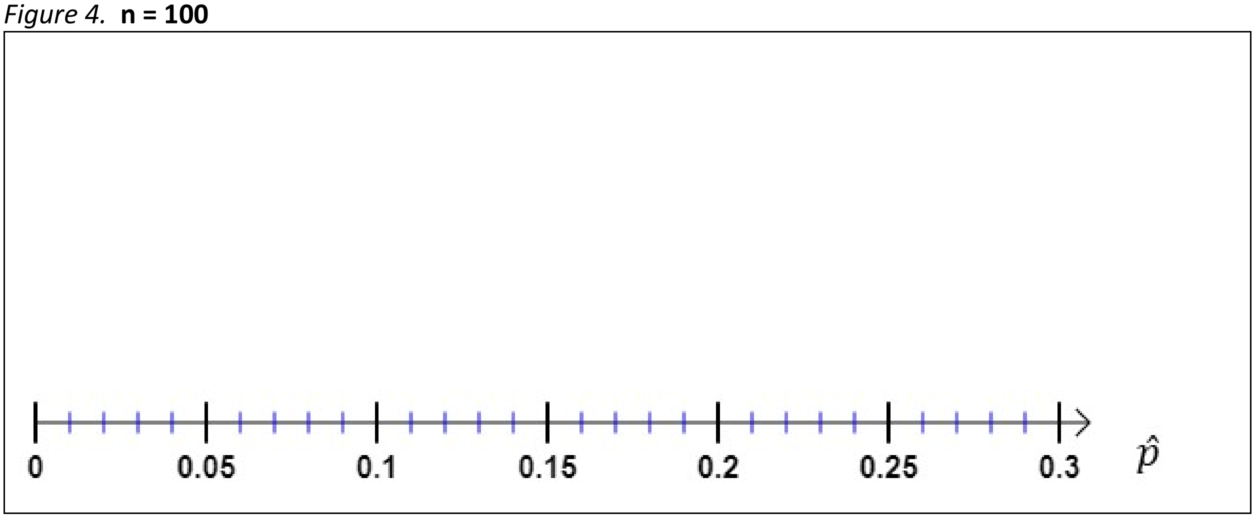


Figure 4. **n = 100**



Let's now compare the features of the graphs in Figures 1 & 3.

For Figure 1:

4. What does the x-axis represent?

5. What does the y-axis represent?

6. Does this graph show the result of one sample or results from many samples?

For Figure 3:

7. What does the x-axis represent?

8. Does this graph show the result of one sample or results from many samples?

9. What does each dot on the graph represent?

<u>Discuss questions 4-9 as a class.</u>

10. What values of $\hat{p}$ should you reasonably expect when sampling? [Consider Figures 3 & 4.]

    For n=20 at a time:

    For n=100 at a time:

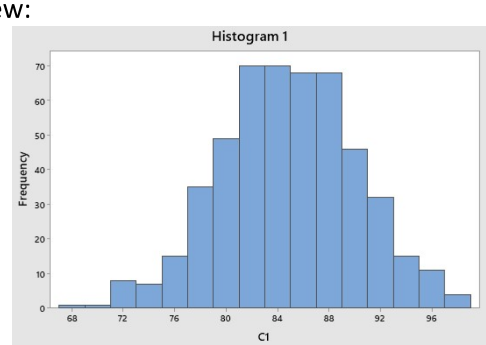
11. Why do you think the sample proportions we calculated for each iteration (draw) were rarely (if ever) above 0.20?

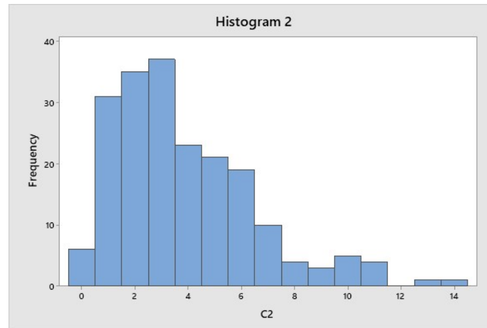
<u>Discuss questions 10 - 11 as a class.</u>

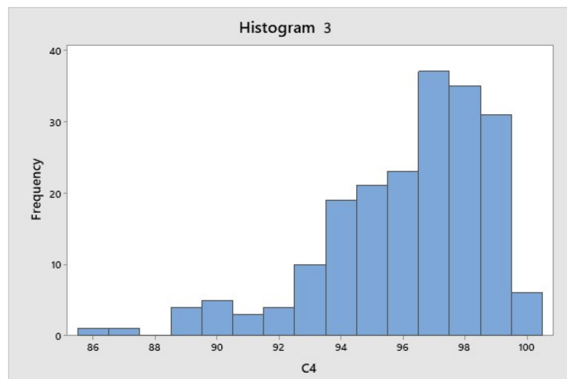
Let's establish some definitions:

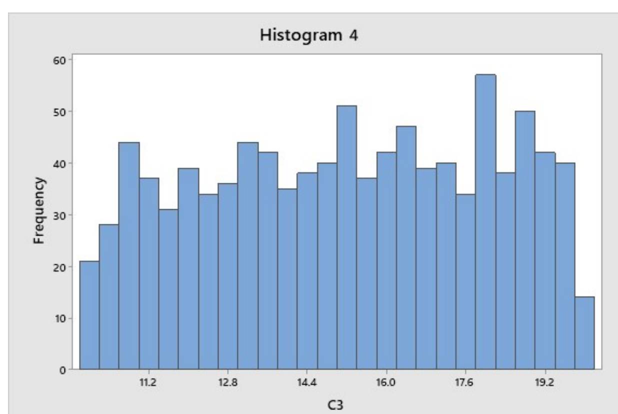
Distribution –

Certain distributions tend to show up often. Here are a few:




Histogram 2
Frequency
C2


Histogram 3
Frequency
C4


Histogram 4
Frequency
C3

Let's define the following properties and identify them on each of the above graphs:
Center (graphically) –

Symmetry –

Tail(s) –

Sampling distribution of a statistic –

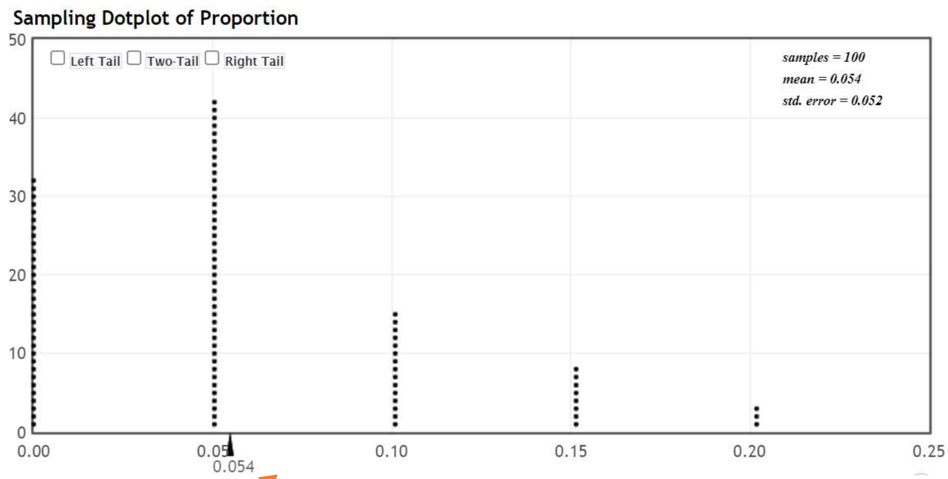
Important Note: The above properties are <u>not</u> used to describe graphs that display distributions of categorical or qualitative data (x-axis shows categories), as shown in the bar graphs we created in Figures 1 & 2.

These properties are only used to describe graphs that display distributions of quantitative data (x-axis is a number line). Examples of acceptable graphs include the four histograms provided above and the dotplots we created in Figures 3 through 4 to display our sampling distributions.

Below are the results from using *StatKey* to run 100 samples for the two sample sizes, n=20 and n=100. Comparing these graphs to the ones we constructed for our earlier samples (Figures 3 & 4), we see that as the number of samples is increased, the shape becomes more pronounced in each sampling distribution.

Figure 5.

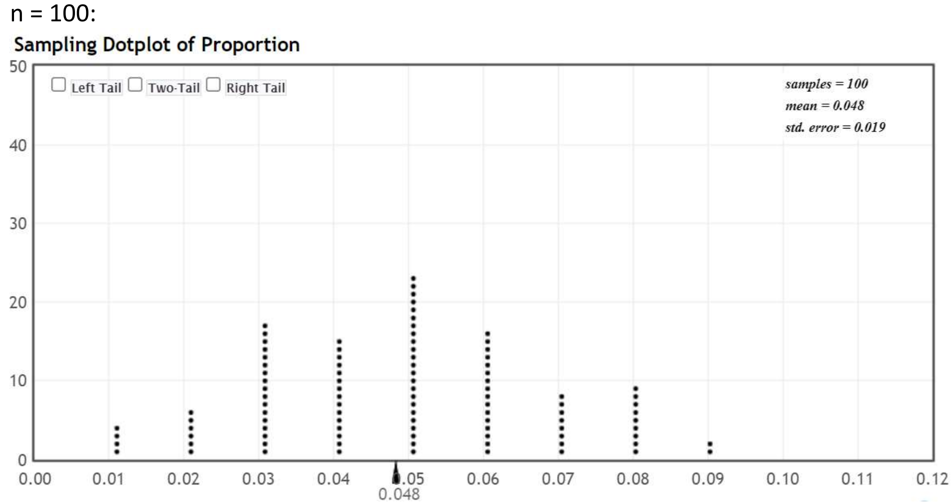
n = 20:



NOTE: The 0.054 shown below the x-axis is the mean of the 100 p-hats in this sampling distribution.

Figure 6.

n = 100:



NOTE: When comparing Figures 5 & 6 to our sampling distributions in Figures 3 & 4, we see that as the number of samples is increased, the shape becomes clearer in each sampling distribution.

12. Describe the shape, location of center, and symmetry (if any) of the sampling distribution for n=20 (Figure 5).

13. Describe the shape, location of center, and symmetry (if any) of the sampling distribution for n=100 (Figure 6).

Discuss questions 12-13 as a class.

Conclusion

Let's return to the questions posed at the start of this activity.

14. If you were to take several random samples of American adults, would the sample proportion of individuals with arches change from sample to sample? Why is this?

15. What would be the most typical value of the sample proportions? Why does this make sense?

16. What property was the same in the n = 20 and n = 100 sampling distributions?

17. What properties changed in the n = 20 and n = 100 sampling distributions?

Discuss questions 14-17 as a class.

18. Describe each of these terms **in the context of this investigation**.

f. *random* sampling

g. sample

h. *statistic* (give a written description and the associated symbol)

i. population

j. *parameter* (give a written description, the associated symbol, and the value)

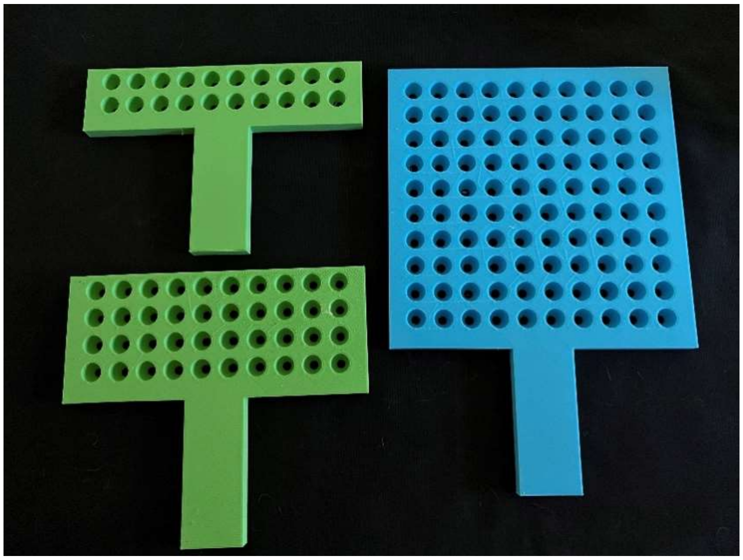
Appendix A

References

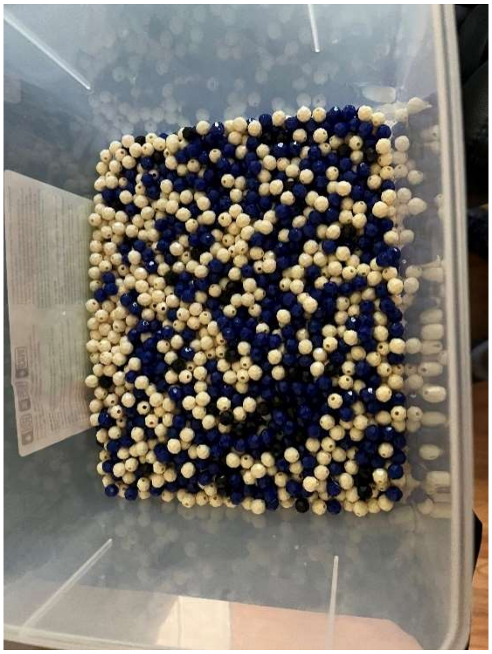
- Cohen, Cathy J., Matthew D. Luttig, and Jon C. Rogowski. 2017. "The Economic Lives of Millennials: GenForward June 2017 Report"
- Lock5 Homepage: http://www.lock5stat.com/
- Malloy, E. J., Bennett, R., Girard, J. E. (2017, May, 20). *Student Designed Data-Oriented Class Projects for a Forensic Science Course* [Poster session]. USCOTS 2017, State College, PA.
- National Forensic Science technology Center, *Principles of Fingerprint Analysis*, Florida International University, retrieved on 6/3/2020 at http://www.forensicsciencesimplified.org/prints/principles.html
- Malloy, E. J., Bennett, R., Girard, J. E. (2017, May, 20). *Student Designed Data-Oriented Class Projects for a Forensic Science Course* [Poster session]. USCOTS 2017, State College, PA.

Appendix B

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.



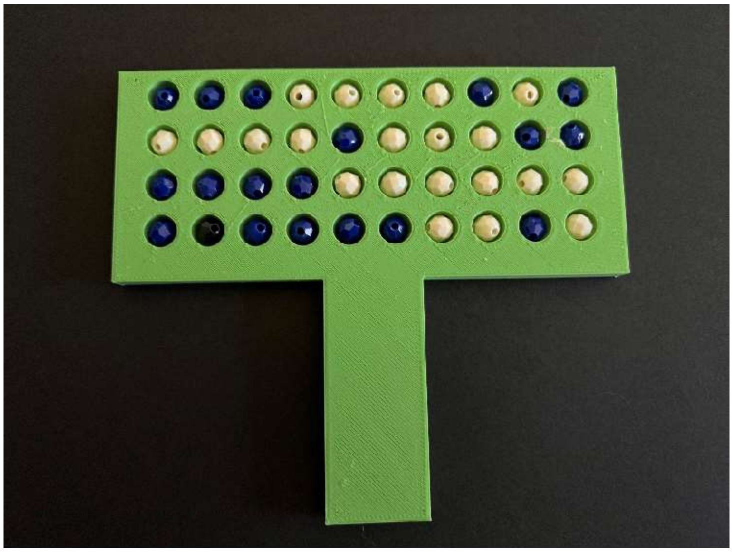
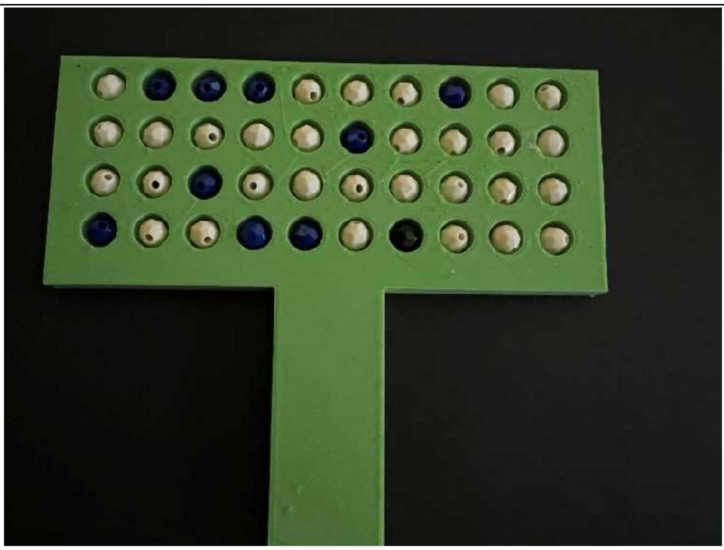
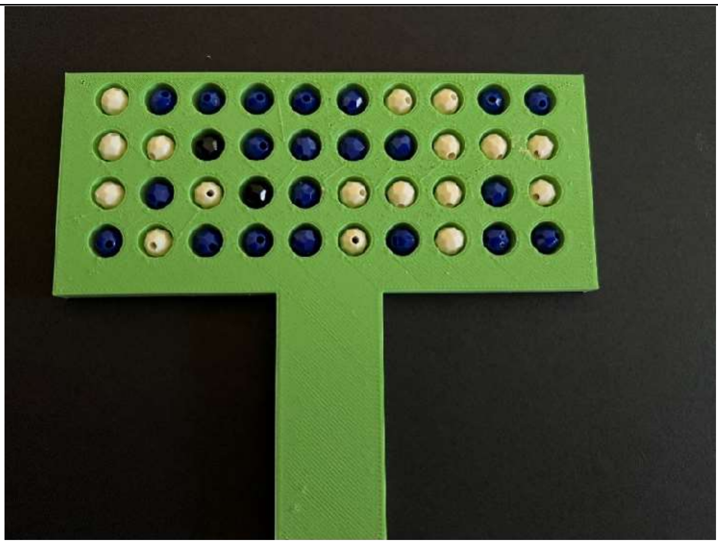
Beads
8mm beads fit into the paddles well. I purchased them through an online retailer so that I can get mass quantities for a more reasonable price. I then put the beads into a Tupperware container with a handle. In this case, I have 60% white, 35% blue and 5% black.



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 were n = 40.

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