### Sampling in a Fair Way

### Goals

- Describe (orally and in writing) methods to obtain a random sample from a population.
- Explain (orally) that random sampling tends to produce representative samples and support valid inferences.
- Justify (orally) whether a given sampling method is fair.

### **Learning Targets**

- I can describe ways to get a random sample from a population.
- I know that selecting a sample at random is usually a good way to get a representative sample.

# Access for Students with Diverse Abilities

Action and Expression (Activity 1)

### **Access for Multilingual Learners**

 MLR8: Discussion Supports (Activity 2)

### **Instructional Routines**

· Poll the Class

### **Required Materials**

### **Materials to Gather**

- Paper bags: Activity 2
- Rulers marked with inches: Activity 2
- Straws: Activity 2

### **Required Preparation**

### **Activity 2:**

Prepare one paper bag containing straws cut to the specified lengths in the table for a demonstration.

length of straw in inches	1/2	1	2	3	4	5
number of straws	6	6	8	6	5	4

The demonstration will also require a ruler marked with inches to measure the straw pieces chosen in a sample.

### **Lesson Narrative**

In this lesson, students consider different methods of selecting a sample. Students begin by critiquing different sampling methods for their benefits and drawbacks. In particular, students notice that some sampling methods are more biased than others. A follow-up activity shows that some methods may seem to be unbiased at first, but they actually have a hidden bias that restricts the sample from being representative of the population. Finally, students practice constructing arguments for why they believe a sampling method is likely to be biased, and they see that selecting a sample at random is more likely to produce a representative sample.

### **Student Learning Goal**

Let's explore ways to get representative samples.

### **Lesson Timeline**



Warm-up



**Activity 1** 



**Activity 2** 



**Activity 3** 



**Lesson Synthesis** 



5 min

Cool-down

### Warm-up

### **Ages of Moviegoers**



### **Activity Narrative**

The purpose of this Warm-up is for students to begin to see that different samples are more or less representative of the population from which they are drawn. Students are asked to look at a dot plot and reason about the context of the sample by matching it to their expectations about what the population should be.

# Launch

Arrange students in groups of 2.

Give students 1 minute of quiet think time, then 1 minute to discuss the things they notice with their partner.

Follow with a whole-class discussion.

### **Student Task Statement**

A survey was taken at a movie theater to estimate the average age of moviegoers.

Here is a dot plot showing the ages of the first 20 people surveyed.

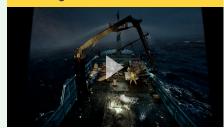


### Sample responses:

- 1. What guestions do you have about the data from survey?
  - Why is everyone so young?
  - · Why is there a 4 year old at the theater?
  - Why were parents not surveyed?
- 2. What assumptions would you make based on these results?
  - There must be at least one G-rated movie playing at the theater.
  - · Maybe the survey was taken during a child's birthday party.
  - There are a lot of young people at this theater.

### Inspire Math

### **Crabbing video**



### Go Online

Before the lesson, show this video to reinforce the real-world connection.

### ilclass.com/l/614220

Please log in to the site before using the QR code or URL.

Student Workbook



# LESSON 14 Sampling in a Fair Way

### **Instructional Routines**

### **Poll the Class**

### ilclass.com/r/10694985

Please log in to the site before using the QR code or URL.



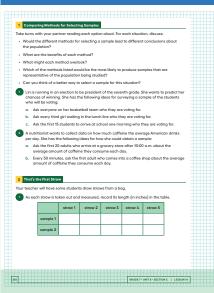
Access for Students with Diverse Abilities (Activity 1, Student Task)

# Action and Expression: Internalize Executive Functions.

To support development of organizational skills in problem-solving, chunk this task into more manageable parts. For example, after students have solved the first 2–3 problems, check in with either select groups of students or the whole class. Invite students to share the strategies they have used so far as well as any questions they have before continuing.

Supports accessibility for: Organization, Attention

### **Student Workbook**



### **Activity Synthesis**

The purpose of the discussion is for students to express their expectations for who would be at the movie theater and whether this group represents that expectation.

Ask several students to report any questions or assumptions they have about the information provided. If possible, display the dot plot so that students can refer to it while giving their answers.

### **Activity 1**

### **Comparing Methods for Selecting Samples**



### **Activity Narrative**

In this activity, students learn that often the best we can do to select a representative sample is to avoid sampling methods that will be inherently biased one way or another. A randomly selected sample is not guaranteed to be representative of the population, but other methods are often biased and thus more likely to produce samples that are not representative of the population.

# Launch 🙎

Arrange students in groups of 2.

Give students 5 minutes of partner work time, and then follow with a whole-class discussion.

### **Student Task Statement**

Take turns with your partner reading each option aloud. For each situation, discuss:

- Would the different methods for selecting a sample lead to different conclusions about the population?
- · What are the benefits of each method?
- · What might each method overlook?
- Which of the methods listed would be the most likely to produce samples that are representative of the population being studied?
- Can you think of a better way to select a sample for this situation?

**1.** Lin is running in an election to be president of the seventh grade. She wants to predict her chances of winning. She has the following ideas for surveying a sample of the students who will be voting:

Sample response: The different methods would probably lead to different conclusions.

- a. Ask everyone on her basketball team who they are voting for.
  - Sample response: The basketball team would probably be convenient for Lin to ask, but her teammates may be more likely to vote for her since they play together.
- b. Ask every third girl waiting in the lunch line who they are voting for.
  Sample response: The girls in the lunch line would also not be too hard to find, but this method would miss the opinions of any boys or those who
- bring their lunch from home.c. Ask the first 15 students to arrive at school one morning who they are

voting for.

- Sample response: The first students to school might be easy to ask if Lin gets to school early, but this method may miss bus riders or students who get to school later for other reasons.
- Asking the first students who arrive at school may be the best of these methods since they are least likely to have a direct relationship with Lin, so they may represent more of the school. To get a wide range of students, a better way to get a sample might be to ask one student from each lunch table.
- **2.** A nutritionist wants to collect data on how much caffeine the average American drinks per day. She has the following ideas for how she could obtain a sample:
  - Sample response: The different methods would probably lead to different conclusions.
  - **a.** Ask the first 20 adults who arrive at a grocery store after 10:00 a.m. about the average amount of caffeine they consume each day.
    - Sample response: The grocery store method would probably give lower results than what might be expected. Asking people at the grocery store would be a good way to get a number of responses fairly quickly, but it would miss out on talking to people who have to get up early and go to work, and they might be more likely to have more caffeine in the morning.
  - **b.** Every 30 minutes, ask the first adult who comes into a coffee shop about the average amount of caffeine they consume each day.
    - Sample response: The coffee shop method might give higher results than expected. People entering a coffee shop might be more likely to know how much caffeine they have each day, but this method would not include people who don't buy coffee and probably have lower caffeine intakes.
    - The grocery store method is probably the better of these two since the coffee shop method would probably produce numbers greater than expected for most people. A better way to get a sample might be to ask people at the mall in the early evening since this includes a wide range of people, and being at the mall at this time is probably not connected to caffeine consumption.

### **Activity Synthesis**

The purpose of the discussion is for students to understand that some methods of sampling are better than others. Although there may be no way to guarantee that a sample is representative of the population, we can certainly avoid methods that will definitely result in some groups being over- or under-represented.

Poll the class on which of the given methods is best for each situation. Record the answers for all to see.

Select several students to explain benefits and drawbacks of each of the sampling methods. After each method has been analyzed for a situation, ask if students have ideas for better ways to get a representative sample for the situation.

Ask students.

"What are some important things to consider when getting a sample?"

Is there a group that this method will show preference for? Is there a group that will automatically be left out of my sample based on the method? If there are groups I didn't even think about, does my method have a way of reaching them?

Explain that people often have biases that may lead them to over- or under-represent some groups in their samples, whether the biases are obvious or not. For example, to find which candidate might be leading in a political race, we might try calling people to survey their preference. This could make our sample biased towards people who are willing to answer their phones when an unknown number calls. Due to (sometimes hidden) biases, the best method for selecting samples is to remove as much of the personal selection as possible.

### **Activity 2**

### That's the First Straw

10 min

### **Activity Narrative**

In this activity, students see an example of a hidden bias. Although the method of selecting straws by taking out the first one touched in the bag appears fair and random, it produces samples that are not representative of the population. In the next activity, students explore ways to resolve the problem by finding other methods of selecting a sample that would be fair for this same context.



Arrange students in groups of 2.

In an opaque bag, include straws cut into 35 pieces according to the table.

length of straw in inches	1/2	1	2	3	4	5
number of straws	6	6	8	6	5	4

Select 5 students to help with a demonstration. One at a time, each student will reach into the bag and remove the first straw piece they touch. They should measure the straw piece to the nearest half inch and announce the value to the class for them to record. Return the straw to the bag and shake the bag. Give the bag to the next student to repeat these steps.

After the class has recorded the 5 lengths, repeat the demonstration and add the second set of 5 straw lengths to the second row of the table in this activity.

Note: Taking out the first one the student touches rather than reaching around in the bag is important for this task.

Following the demonstration, give students partner work time and follow with a whole-class discussion.

### **Student Task Statement**

Your teacher will have some students draw straws from a bag.

**1.** As each straw is taken out and measured, record its length (in inches) in the table.

### Sample responses:

	straw 1	straw 2	straw 3	straw 4	straw 5
sample 1	5	5	4	4	2
sample 2	5	5	4	3	2

- 2. Estimate the mean length of all the straws in the bag based on:
  - **a.** the mean of the first sample.

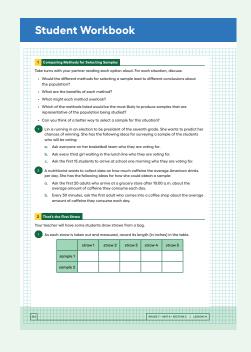
4 inches, since  $(5 + 5 + 4 + 4 + 2) \div 5 = 4$ 

b. the mean of the second sample.

3.8 inches, since  $(5 + 5 + 4 + 3 + 2) \div 5 = 3.8$ 

**3.** Were your two estimates the same? Did the mean length of all the straws in the bag change in between selecting the two samples? Explain your reasoning.

No. Since we drew out different straws the second time and used a different sample, the means were different. The same straws were in the bag each time, though, so the mean of all the straws did not change.





# Access for Multilingual Learners (Activity 2, Synthesis)

### MLR8: Discussion Supports.

Display sentence frames to support whole-class discussion: "Selecting straws from a bag would be "fair" when ..." "I believe this selection process was not fair because ..." and "I would change the selection process by ..."

Advances: Speaking, Conversing

**4.** The actual mean length of all of the straws in the bag is about 2.37 inches. How do your estimates compare to this mean length?

The means from the samples were much longer than the actual mean.

**5.** If you repeated the same process again but you selected a larger sample (such as 10 or 20 straws, instead of just 5), would your estimate be more accurate? Explain your reasoning.

No. Since the task asks to take out the first straw touched, it is more likely to get a long straw than a little one at the bottom of the bag. Taking out more straws is not likely to change that tendency.

### **Activity Synthesis**

### Ask students:

- "What does it mean for a process of selecting straws to be 'fair'?"
  There should be an equal chance for each item to be selected.
- "Is this selection process fair?"

No, the shorter pieces probably fell to the bottom of the bag and are less likely to be touched first.

Reveal the contents of the bag.

- (a) "Are your samples representative of the contents of the bag? Explain your reasoning."
  - No, there are a lot more longer straws in our samples than in the bag.
- O "Does every straw in the bag have an equal chance of being selected?"
  - Longer straws are probably touched first, so they are probably overrepresented in our sample.
- "If we increase the sample size to 10, would that make the sample more representative?"

No, in fact, it may increase the over-representation of the longer straws and be even more misleading.

Tell students that a larger sample does not help the estimate if the selection process is flawed. For example, if someone uses the heights of 40 basketball players instead of only 20 basketball players to determine average height of everyone in the United States, the larger sample probably does not represent the population any better.

Explain that, although the process may seem random since we took out as much of the human element of the choosing process as possible, the longer straws were over-represented in our samples. It is important to try to anticipate all the different ways that the selection process might be biased to avoid it as much as possible.

### That's the Last Straw

10 min

### **Activity Narrative**

In this activity, students determine whether alternate methods of selecting items for a sample from the same population are fair. For the methods that work, the physical objects are linked with numerical values to remove even more of the bias toward selecting certain objects more often than others.

Warm-up

### Launch

Tell students that the straws from the previous task are ordered and numbered with 1 representing the shortest straw and 35 representing the longest. Display the table for all to see.

straw number	length (inches)
1–6	0.5
7–12	1
13–20	2
21–26	3
27–31	4
32–35	5

Before beginning the task, ask students:

- "What would it mean for the sampling method to be fair?"
  Each item has an equal chance of being selected.
- Can you think of a way of sampling the straws that would be fair?"

It is OK for students to struggle with answering this question at this stage. Listen for students who suggest methods using randomness to select straw numbers.

Following the discussion, allow students quiet work time, and then follow with a whole-class discussion.



### **Student Task Statement**

There are a total of 35 straws in the bag. Suppose we put the straws in order from shortest to longest and assign each straw a number from 1 to 35. For each of these methods, decide whether it would be a fair way to select a sample of 5 straws. Explain your reasoning.

- 1. Select the straws numbered 1 through 5.
  - Sample response: This is not a fair method since there is no chance that straws numbered 6 through 35 will get selected. Also, because the straws are numbered in order of length, the sample will over-represent the shorter straws.
- 2. Write the numbers 1 through 35 on pieces of paper that are all the same size. Put the papers into a bag. Without looking, select five papers from the bag. Use the straws with those numbers for your sample.
  - Sample response: This is a fair method. Each straw length has an equal chance of being chosen, and the sample should represent the straws fairly well.
- **3.** Using the same bag as the previous question, select one paper from the bag. Use the number on that paper to select the first straw for your sample. Then use the next 4 numbers in order to complete your sample. For example, if you select number 17, then you also use straws 18, 19, 20, and 21 for your sample.
  - Sample response: This is not a fair method since the straw lengths will tend to be grouped together rather than represent the spread of the real straw lengths.
- **4.** Create a spinner with 35 sections that are all the same size, and number them 1 through 35. Spin the spinner 5 times, and use the straws with those numbers for your sample.
  - Sample response: This is a fair method. Each straw length has an equal chance of being chosen, and the sample should represent the straws fairly well.

### **Are You Ready for More?**

Computers accept inputs, follow instructions, and produce outputs, so they cannot produce truly random numbers. If you knew the input, you could predict the output by following the same instructions the computer is following. When truly random numbers are needed, scientists measure natural phenomena such as radioactive decay or temperature variations. Before such measurements were possible, statisticians used random number tables, like this:

 85
 67
 95
 62
 42
 61
 21
 35
 15
 34
 41

 85
 94
 61
 72
 53
 24
 15
 67
 85
 94
 12

 67
 88
 15
 32
 42
 65
 75
 98
 46
 25
 13

 07
 53
 60
 75
 82
 34
 67
 44
 20
 42
 33

 99
 37
 40
 33
 40
 88
 90
 50
 75
 22
 90

 00
 33
 84
 57
 91
 15
 70
 08
 90
 03
 02

 78
 07
 16
 51
 13
 89
 67
 64
 54
 05
 26

 62
 06
 61
 43
 02
 60
 73
 58
 38
 53
 88

 02
 50
 88
 44
 37
 05

Use this table to select a sample of 5 straws. Pick a starting point at random in the table. If the number is between 01 and 35, include that number straw in your sample. If the number has already been selected, or it is not between 01 and 35, ignore it, and move on to the next number.

Sample response: The sample includes straws 24, 15, 12, 32, and 25.

### **Activity Synthesis**

Tell students that a random sample from a population is a sample that is selected in a way that gives every different possible sample of the same size an equal chance of being the sample selected.

- "Which of the four methods proposed would be a random sample?"
  putting the papers in the bag or using the spinner
- "Would the techniques described here work for other situations in which we wanted a sample? For example, would they work to select 50 people in a large city to represent the views of the city residents?"

Although they would work in theory for large populations, it would be too time-consuming to write over a million numbers [or names] on pieces of paper and put them in a bag. Similarly, a spinner that is divided into a million sections would be difficult to manage. Computers can be used to generate random numbers for larger populations.

- "The most common straw in the bag is the 2-inch straw. When selecting one of the straw numbers (not lengths) at random, what is the probability of selecting a 2-inch straw?"
  - $\frac{8}{35}$ , since there are 8 straws that are 2 inches long and 35 total straws





### Explain that:

- A representative sample would have more of the more common lengths, and there is also a higher probability of selecting these lengths, so a random selection should be a good way to select a representative sample.
- A random sample does not guarantee a representative sample, but
  it avoids methods that might over- or under-represent items of the
  population. Because we do not know the data for the population,
  a random sample usually provides the best opportunity to get a
  representative sample.
- While it is the most ideal method, it is not always possible to generate a
  random sample. For example, if we wanted to know the average size of
  salmon in the wild, it is impossible to know how many there are, much
  less identify them individually, select a few at random from the list,
  then capture and measure those exact individuals. In these cases, it is
  important to try to intentionally reduce bias as much as possible when
  selecting the sample.

### **Lesson Synthesis**

Consider asking these discussion questions:

"What makes a sample selected at random the best way to select individuals for a sample?"

It avoids biases that might be introduced using other methods.

"As part of an English project, you want to look at the length of lines in Shakespeare's plays. What are some methods of selecting a random sample of lines from these plays?"

Assign each line in the plays a number, and use a computer to select several random numbers that correspond to the lines.

### **Lesson Summary**

A sample is selected at random from a population if it has an equal chance of being selected as every other sample of the same size. For example, if there are 25 students in a class, then we can write each of the students' names on a slip of paper and select 5 papers from a bag to get a sample of 5 students selected at random from the class.

Other methods of selecting a sample from a population are likely to be biased. This means that it is less likely that the sample will be representative of the population as a whole. For example, if we select the first 5 students who walk in the door, that will not give us a random sample because students who typically come late are not likely to be selected. A sample that is selected at random may not always be a representative sample, but it is more likely to be representative than using other methods.

It is not always possible to select a sample at random. For example, if we want to know the average length of wild salmon, it is not possible to identify each one individually, select a few at random from the list, and then capture and measure those exact fish. When a sample cannot be selected at random, it is important to try to reduce bias as much as possible when selecting the sample.

### Cool-down

### **Sampling Spinach**



### **Student Task Statement**

A public health expert is worried that a recent outbreak of a disease may be related to a batch of spinach from a certain farm. She wants to test the plants at the farm, but it will ruin the crop if she tests all of them.

- **1.** If the farm has 5,000 spinach plants, describe a method that would produce a random sample of 10 plants.
  - Sample response: She could number the plants from I to 5,000 and have a computer select IO random numbers between I and 5,000, then test the plants that correspond to the numbers the computer generated.
- 2. Why would a random sample be useful in this situation?
  - Sample response: Since it is not known where the disease may have originated, a random sample would hopefully produce a wide selection of plants that would be representative of the entire crop.

### **Responding To Student Thinking**

### **Press Pause**

By this point in the unit, there should be some student mastery of identifying representative samples and random sampling. If students struggle, make time to revisit the work of the activity referred to here. See the Course Guide for ideas to help students re-engage with earlier work.

Unit 8, Lesson 13, Activity 2 Screen Time

# Student Workbook TAGE PRACTICE PROBLEMS The most deportment monoger at a grocery store is worried some of the pockages of ground best between 50 environ ground most produced field. He decides to take a sample of 5 pockages from a shipment containing 100 pockages of ground best. The pockages were more and the pockages were more of the work produced the bas of them the bas of them the bas of them the pockages were produced to they were put in the bas. Is each one has different number of between the same field of the pockages of ground best. The pockages were produced to the work produced to the pockages of ground best. The pockages were produced to the pockages of ground best. The pockages of ground best in the pockages of ground best in the pockages of ground best. The pockages of ground best in the pockages of ground best

### **Practice Problems**

6 Problems

### **Problem 1**

The meat department manager at a grocery store is worried some of the packages of ground beef labeled as having one pound of meat may be under-filled. He decides to take a sample of 5 packages from a shipment containing 100 packages of ground beef. The packages were numbered as they were put in the box, so each one has a different number between 1 and 100.

Describe how the manager can select a fair sample of 5 packages.

Sample response: The manager should pick a method that will result in a random sample. One way is to write the numbers from I to IOO on slips of paper and put them in a bag. Mix them well, then select 5 slips of paper from the bag. Use the numbers on these slips to identify which packages in the shipment will be in the sample. The manager could also use random digits or another type of random number generator.

### Problem 2

Select **all** the reasons why random samples are preferred over other methods of getting a sample.

- **A.** If you select a random sample, you can determine how many people you want in the sample.
- **B.** If a random sample is selected, the desired number of people in the sample can be determined.
- **C.** A random sample is likely to give a sample that is representative of the population.
- **D.** A random sample is a fair way to select a sample because each member of the population has an equal chance of being selected.
- **E.** If a random sample is used, the sample mean will always be the same as the population mean.

### Problem 3

Jada is using a computer's random number generator to produce 6 random whole numbers between 1 and 100 so she can use a random sample. The computer produces the numbers 1, 2, 3, 4, 5, and 6. Should she use these numbers or have the computer generate a new set of random numbers? Explain your reasoning.

### Yes

Sample reasoning: Unless she has reason to believe the computer is messed up, she should use these numbers. To be random, every possible set of numbers should have a chance to be selected, so this set of numbers should be considered to be as random as any other set of numbers.

### Problem 4

from Unit 8, Lesson 6

A group of 100 people is divided into 5 groups with 20 people in each. One person's name is chosen, and everyone in their group wins a prize. Noah simulates this situation by writing 100 different names on papers and putting them in a bag, then drawing one out. Kiran suggests there is a way to do it with fewer paper slips. Explain a method that would simulate this situation with fewer than 100 slips of paper.

Sample response: Since the entire group wins a prize, label each group I, 2, 3, 4, or 5, and put those 5 pieces of paper in the bag to draw one. The probability of Group I winning is I5, which is the same probability as choosing one of the 20 people in Group I from IOO slips of paper  $(\frac{20}{100})$  in the simulation.

### **Problem 5**

from Unit 8, Lesson 12

Data collected from a survey of American teenagers aged 13 to 17 is used to estimate that 29% of teens believe in ghosts. This estimate is based on data from 510 American teenagers. What is the population that people carrying out the survey were interested in?

- A. all people in the United States
- B. the 510 teens that were surveyed
- C. all American teens who are between the ages of 13 and 17
- D. the 29% of the teens surveyed who said they believe in ghosts

### Problem 6

from Unit 8, Lesson 7

A computer simulates flipping a coin 100 times, then counts the longest string of heads in a row.

Based on these results, estimate the probability that there will be at least 15 heads in a row.

trial	most heads in a row			
1	8			
2	6			
3	5			
4	11			
5	13			

Sample response: Between 0 and  $\frac{1}{5}$ . It did not happen at all in these trials, but it is not impossible, so it is probably somewhere between these two values.

