Lesson 14: Sampling in a fair way

Goals

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

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.

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 have a hidden bias that restricts the sample from being representative of the population. Finally, students practise recognising when 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.

Addressing

- Understand that statistics can be used to gain information about a population by examining a sample of the population; generalisations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
- Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.
- Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy.

Building Towards

- Understand that statistics can be used to gain information about a population by examining a sample of the population; generalisations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

Instructional Routines

- Collect and Display
- Discussion Supports
- Poll the Class
- Think Pair Share

Required Materials

Paper bags
Rulers marked with inches
Straws

Required Preparation

For the That's the First Straw activity, prepare one paper bag containing straws cut to the specified lengths in the table for a demonstration.

$\frac{1}{2}$	1	2	3	4	5
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.

Student Learning Goals

Let's explore ways to get representative samples.

14.1 Ages of Filmgoers

Warm Up: 5 minutes

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.

Instructional Routines

- Think Pair Share

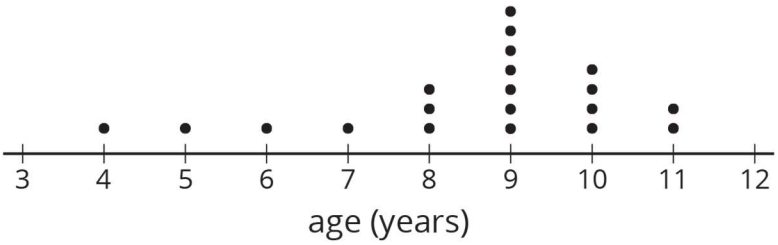
Launch

Arrange students in groups of 2. Give students 1 minute of quiet think time, and then 1 minute to discuss the things they notice with their partner, followed by a whole-class discussion.

Student Task Statement

A survey was taken at a cinema to estimate the average age of filmgoers.

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



1. What questions do you have about the data from the survey?
2. What assumptions would you make based on these results?

Student Response

Answers vary. Sample responses:

- Why is everyone so young?
- Why is there a 4 year old at the cinema?
- Why were parents not surveyed?
- There must be at least one U-rated film playing at the cinema.
- Maybe the survey was taken during a child's birthday party.
- There are a lot of young people at this cinema.

Activity Synthesis

The purpose of the discussion is for students to express their expectations for who would be at the cinema 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.

14.2 Comparing Methods for Selecting Samples

10 minutes

In the previous lesson, students learned that it is very difficult to select representative samples when the population data is unknown. In this lesson, 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 tend to produce samples that are not representative of the population.

Instructional Routines

- Poll the Class

Launch

Arrange students in groups of 2. Give students 5 minutes of partner work time followed by a whole-class discussion.

Action and Expression: Internalise Executive Functions. Chunk this task into more manageable parts. 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: Organisation; Attention

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 year 8. She wants to predict her chances of winning. She has the following ideas for surveying a sample of the students who will be voting:
    a. Ask everyone on her basketball team who they are voting for.
    b. Ask every third girl waiting in the lunch line who they are voting for.
    c. Ask the first 15 students to arrive at school one morning who they are voting for.
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:

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.

b. Every 30 minutes, ask the first adult who comes into a coffee shop about the average amount of caffeine they consume each day.

Student Response

Answers vary. Sample responses:

1. The different methods would probably lead to different conclusions. The basketball team would probably be convenient for Lin to do, but her teammates may be more likely to vote for her since they play together. The girls in the lunch line would also not be too hard to find, but would miss the opinions of any boys or those who bring their lunch from home. The first students to school might be easy to ask if Lin gets to school early, but may miss bus riders or students who get to school later for other reasons. The first students to arrive at school may be the best of these methods since they are least likely to have a direct relationship with Lin, so may represent more of the school. A better way to sample the students might be to ask a student at each different lunch table to get a wide range of students.

2. The different methods would probably lead to different conclusions. The grocery store method would probably be lower than what might be expected and the coffee shop method might be higher. Asking people at the grocery store would be a good way to get a number of responses fairly quickly, but 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. People entering a coffee shop might be more likely to know how much caffeine they have each day, but this method would not talk to 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 sample people 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 the 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 scenario. Record these answers for all to see.

Select several students to explain benefits and drawbacks of each of the sampling methods. After each method has been analysed 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: 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, if you want to send a survey out for people to respond to questions, you may not reach people who do not have email addresses. Due to the (sometimes hidden) biases, the best method for selecting samples is to remove as much of the personal selection as possible. In the rest of this lesson, we will explore methods for generating samples that avoid biases.

14.3 That’s the First Straw

10 minutes

In the previous activity, students saw that some methods for taking samples are more likely to produce samples that are not representative of the population than others. In this activity, students see an example of a hidden bias. Although the method of selecting straws by taking out the first one in the bag touched 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.

Instructional Routines

- Discussion Supports

Launch

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	$\frac{1}{2}$	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 followed by a whole-class discussion.

Representation: Internalise Comprehension. Activate or supply background knowledge of calculating measures of centre. Allow students to use calculators to ensure inclusive participation in the activity.
Supports accessibility for: Memory; Conceptual processing

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.

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

2. Estimate the mean length of all the straws in the bag based on:
   a. the mean of the first sample.
   b. the mean of the second sample.
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.
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?
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.

Student Response

1.

Answers vary. Sample response:

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

2. a. 4 inches since $(5 + 5 + 4 + 4 + 2) \div 5 = 4$
   b. 3.8 inches since $(5 + 5 + 4 + 3 + 2) \div 5 = 3.8$

3. 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.

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

5. 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 would it mean for a process of selecting straws to be 'fair?'" (There should be an equal chance for each item to be selected.)
- "Was this selection process fair?" (No, the shorter pieces probably fell to the bottom of the bag and were less likely to be touched first.)

Reveal the contents of the bag.

- "Were your samples representative of the contents of the bag? Explain your reasoning."
- "Did every straw in the bag have an equal chance of being selected?" (Longer straws were probably touched first so they were probably over represented in our sample.)
- "If we increased 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: 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.

Listening, Speaking, Conversing: Discussion Supports. To help students respond to the questions during the discussion, provide sentence frames such as: "Selecting straws from a bag would be "fair" when . . ." "I believe this selection process was/was not fair because . . ." and "I would change the selection process by . . . ." This will help students discover the hidden bias involved in the activity by communicating about the process.
Design Principle(s): Support sense-making; Cultivate conversation

14.4 That's the Last Straw

10 minutes

In the previous activity, students selected straws using a method that might have seemed fair at first, but did not produce a representative sample since the method was flawed. 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.

Instructional Routines

- Collect and Display

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	0.5
2	0.5
3	0.5
4	0.5
5	0.5
6	0.5
7	1.0
8	1.0
9	1.0
10	1.0
11	1.0
12	1.0
13	2.0
14	2.0
15	2.0
16	2.0
17	2.0
18	2.0
19	2.0

20	2.0
21	3.0
22	3.0
23	3.0
24	3.0
25	3.0
26	3.0
27	4.0
28	4.0
29	4.0
30	4.0
31	4.0
32	5.0
33	5.0
34	5.0
35	5.0

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.)

Following the discussion, allow students quiet work time followed by a whole-class discussion.

Representation: Internalise Comprehension. Use colour and annotations to illustrate student strategies. As students describe their reasoning, use colour and annotations to scribe their thinking on a display. Ask students how they knew the situation would be fair or not, and label each accordingly.
Supports accessibility for: Visual-spatial processing; Conceptual processing Speaking: Collect and Display. Create a visual display with the heading "random sample." As students discuss whether the sampling methods are fair, write down the words and phrases students use to explain why the sample methods are fair or not fair. Listen for and amplify words that help define the term "random sample" such as "every possible sample" and "equal chances." This will help students use appropriate mathematical language when constructing a definition for the new term.
Design Principle(s): Support sense-making; Maximise meta-awareness

Student Task Statement

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

1. Select the straws numbered 1 through 5.
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.
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.)
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.

Student Response

Answers vary. Sample response:

1. 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, it will over represent the shorter straws.
2. This is a fair method. Each straw length has an equal chance of being chosen and should represent the straws fairly well.
3. 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. This is a fair method. Each straw length has an equal chance of being chosen and 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 02 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 03 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 13 54 78 97 30

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 is not between 01 and 35, ignore it, and move on to the next number.

Student Response

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

Activity Synthesis

Define random sample: 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 you wanted a sample? For example, 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 was 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:

- 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. Since 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 you 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.)

14.5 Sampling Spinach

Cool Down: 5 minutes

The cool-down assesses whether students understand how to select a random sample and why it would be useful to use a random sample.

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.
2. Why would a random sample be useful in this situation?

Student Response

Answers vary. Sample responses:

1. She could number the plants from 1 to 5 000 and have a computer select 10 random numbers between 1 and 5 000, then test the plants that correspond to the numbers the computer generated.

2. 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.

Student 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.

Lesson 14 Practice Problems

Problem 1 Statement

The meat department manager at a grocery store is worried some of the packages of ground beef labelled 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.

Solution

Answers vary. Sample response: The manager should pick a method that will result in a random sample. One way is to write the numbers from 1 to 100 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 might also use random digits or another type of random number generator.

Problem 2 Statement

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. A random sample is always the easiest way to select a sample from a population.

c. A random sample is likely to give you a sample that is representative of the population.

d. A random sample is a fair way to select a sample, because each person in the population has an equal chance of being selected.

e. If you use a random sample, the sample mean will always be the same as the population mean.

Solution ["C", "D"]

Problem 3 Statement

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.

Solution

Yes. Explanations vary. Sample explanation: 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 Statement

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.

Solution

Answers vary. Sample response: Since the entire group wins a prize, label each group 1, 2, 3, 4, or 5, and put those 5 pieces of paper in the bag to draw one. The probability of Group 1 winning is $\frac{1}{5}$, which is the same probability as the simulation using 100 slips of paper $\left(\frac{20}{100}\right)$.

Problem 5 Statement

Data collected from a survey of American teenagers aged 13 to 17 was used to estimate that 29% of teens believe in ghosts. This estimate was 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.

Solution C

Problem 6 Statement

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

Solution

Answers vary. 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.