SAMPLE MEANS – EXPLORING SAMPLING VARIABILITY

TEACHER VERSION

Subject Level:
Middle School Math

Grade Level:
7

Approx. Time Required:
45-60 minutes

Learning Objectives:

- Students will be able to help create dot plots to represent a collection of sample means.
- Students will be able to select a random sample from a population.
- Students will be able to understand that sample means for different random samples of the same population will differ due to random selection.
- Students will be able to understand that there is less sampling variability in the sample mean when the sample size is large than when the sample size is small.

Activity Description

Students will explore sampling variability in the sample means of different random samples of a population, using data on Internet access from the U.S. Census Bureau's American Community Survey.

Suggested Grade Level:
7

Approximate Time Required:
45-60 minutes

Learning Objectives:

- Students will be able to help create dot plots to represent a collection of sample means.
- Students will be able to select a random sample from a population.
- Students will be able to understand that sample means for different random samples of the same population will differ due to random selection.
- Students will be able to understand that there is less sampling variability in the sample mean when the sample size is large than when the sample size is small.

Topics:

- Random sampling
- Sample mean
- Sampling variability

Skills Taught:

- Calculating sample means
- Constructing and comparing dot plots

Materials Required

- The student version of this activity, 8 pages
- A printed version of the data cards in the teacher version of the activity
- Calculators
- Sticky notes

Activity Items

There are no separate items for this activity.

For more information to help you introduce your students to the U.S. Census Bureau, read "*Census Bureau 101 for Students*." This information sheet can be printed and passed out to your students as well.

Standards Addressed

See charts below. For more information, read "*Overview of Education Standards and Guidelines Addressed in Statistics in Schools Activities*."

Common Core State Standards for Mathematics

Standard	Domain	Cluster
CCSS.MATH.CONTENT.7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations 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.	7 SP – Statistics & Probability	Use random sampling to draw inferences about a population.
CCSS.MATH.CONTENT.7.SP.A.2 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.	7 SP – Statistics & Probability	Use random sampling to draw inferences about a population.

Common Core State Standards for Mathematical Practice

Standard
CCSS.MATH.PRACTICE.MP4. Model with mathematics. Students will help create dot plots to show the variability of Internet access among states and the District of Columbia.
CCSS.MATH.PRACTICE.MP6. Attend to precision. Students will distinguish between two distributions of sample means for the same population to understand sampling variability.
CCSS.MATH.PRACTICE.MP7. Look for and make use of structure. Students will recognize patterns in and draw conclusions about the data.

National Council of Teachers of Mathematics' Principles and Standards for School Mathematics

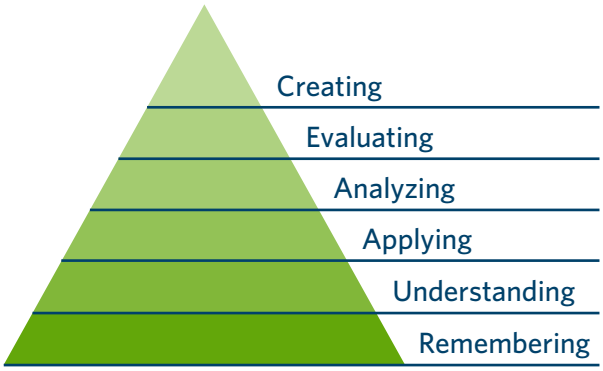
Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Develop and evaluate inferences and predictions that are based on data.	Use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken.

Guidelines for Assessment and Instruction in Statistics Education

GAISE	Level A	Level B	Level C
Formulate Questions	X		
Collect Data		X	
Analyze Data		X	
Interpret Results			X

Bloom's Taxonomy

Students will ***apply*** their knowledge of means and sampling to help create dot plots of sample means. Students will also ***analyze*** the dot plots to determine the center and spread of the data.



Teacher Notes

Before the Activity

Students must understand the following key terms:

- **Population** – the number of people living in an area, representing the entire set of data from which samples can be taken
- **Sample** – a selection of observations from a population
- **Outlier** – a data point that is well outside of the expected range of values or does not follow the overall pattern of the other data points
- **Random sampling** – the process of randomly selecting a representative subset of a population
- **Sample mean** – the numerical average of a sample of data
- **Sampling variability** – the concept that a data set's values will vary in repeated random sampling

Students should have the following skills:

- Ability to calculate means
- Ability to read and interpret dot plots

Teachers should draw two number lines side by side on the board, with values going from 60 to 90 and counting by twos. Teachers should title the first one "Percentage of Homes With Internet Access in Each State and the District of Columbia (n = 3)" and the second one "Percentage of Homes With Internet Access in Each State and the District of Columbia (n = 10)."

Teachers should also print and cut out the data cards on the following pages, which are for parts 2 and 3 of the activity, placing each group of cards in a separate container from which students can draw.

Teachers should ask students, "If I wanted to find the average percentage of homes with Internet access in all 50 states and the District of Columbia, what problems would I encounter in averaging all the values in such a large data set?" Teachers should expect answers like: "It would take a really long time," "It would be really tedious," "It could be easy to make mistakes working with so much data," and "You could lose your place as you go through the data."

Teachers should explain that statisticians often conduct random sampling because it can be difficult to gather all data from a large population. As long as the sample is representative of the population, its data can tell us a lot. Teachers should also tell students that this activity uses data from the American Community Survey (ACS), conducted monthly by the Census Bureau, which is designed to show how communities are changing. Through asking questions of a sample of the population, it produces national data on more than 35 categories of information, such as education, income, housing, and employment. One question the ACS asks is "Do you have a computer at home with Internet access?"

Part 2 Data Cards: Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014

Alabama	71%
Alaska	86%
Arizona	79%
Arkansas	70%
California	83%
Colorado	85%
Connecticut	86%
Delaware	79%
District of Columbia	75%
Florida	80%
Georgia	78%
Hawaii	86%
Idaho	80%
Illinois	81%
Indiana	77%
Iowa	81%
Kansas	80%
Kentucky	75%
Louisiana	72%
Maine	82%
Maryland	84%
Massachusetts	86%
Michigan	79%
Minnesota	85%
Mississippi	64%
Missouri	77%
Montana	79%
Nebraska	81%
Nevada	80%
New Hampshire	87%
New Jersey	86%
New Mexico	71%
New York	81%
North Carolina	77%
North Dakota	81%
Ohio	80%
Oklahoma	74%
Oregon	84%
Pennsylvania	80%
Rhode Island	82%
South Carolina	73%
South Dakota	78%
Tennessee	74%
Texas	76%
Utah	87%
Vermont	82%
Virginia	82%
Washington	86%
West Virginia	74%
Wisconsin	82%
Wyoming	82%

Part 3 Data Cards: Random Sampling of the Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014

1	
Oklahoma	74%
Delaware	79%
Montana	79%
Connecticut	86%
Minnesota	85%
Connecticut	86%
Washington	86%
Washington	86%
Texas	76%
Mississippi	64%

2	
Washington	86%
Hawaii	86%
New Jersey	86%
California	83%
Arkansas	70%
South Dakota	78%
Utah	87%
North Dakota	81%
Oregon	84%
South Dakota	78%

3	
Alaska	86%
Indiana	77%
Maine	82%
Maine	82%
Delaware	79%
Wyoming	82%
Pennsylvania	80%
Massachusetts	86%
Iowa	81%
New Mexico	71%

4	
Kentucky	75%
Minnesota	85%
Rhode Island	82%
South Dakota	78%
North Carolina	77%
District of Columbia	75%
Louisiana	72%
Oregon	84%
New Hampshire	87%
New York	81%

5	
Hawaii	86%
Utah	87%
Kansas	80%
Alaska	86%
Wyoming	82%
Georgia	78%
Idaho	80%
Maryland	84%
Washington	86%
Washington	86%

6	
Georgia	78%
Washington	86%
Hawaii	86%
Nevada	80%
Indiana	77%
New Hampshire	87%
New Mexico	71%
Louisiana	72%
Rhode Island	82%
Massachusetts	86%

Part 3 Data Cards: Random Sampling of the Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014 (Continued)

7	
Kentucky	75%
Georgia	78%
Washington	86%
Illinois	81%
Connecticut	86%
District of Columbia	75%
Rhode Island	82%
Alaska	86%
North Carolina	77%
Mississippi	64%

8	
Montana	79%
South Dakota	78%
Alabama	71%
Hawaii	86%
Virginia	82%
Tennessee	74%
New Jersey	86%
Pennsylvania	80%
Mississippi	64%
Georgia	78%

9	
New Mexico	71%
South Carolina	73%
Wyoming	82%
Maine	82%
North Carolina	77%
Utah	87%
New York	81%
Arkansas	70%
Arizona	79%
New Jersey	86%

10	
Maryland	84%
Kansas	80%
South Carolina	73%
North Carolina	77%
Louisiana	72%
Georgia	78%
Arizona	79%
Iowa	81%
Alaska	87%
New Jersey	86%

11	
Vermont	82%
Louisiana	72%
Alabama	71%
Georgia	78%
New Mexico	71%
Florida	80%
Maine	82%
New Hampshire	87%
Alabama	71%
California	86%

12	
Mississippi	64%
Nebraska	81%
Connecticut	86%
Missouri	77%
Vermont	82%
New Hampshire	87%
Iowa	81%
Utah	87%
Florida	80%
Arizona	79%

Part 3 Data Cards: Random Sampling of the Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014 (Continued)

13	
Connecticut	86%
Georgia	78%
Massachusetts	86%
Connecticut	86%
Wisconsin	82%
Mississippi	64%
New York	81%
South Dakota	78%
Michigan	79%
South Carolina	73%

14	
New Jersey	86%
Nebraska	81%
Kansas	80%
Montana	79%
Tennessee	74%
Louisiana	72%
New Hampshire	87%
Alaska	87%
West Virginia	74%
Colorado	85%

15	
Nevada	80%
New Jersey	86%
New Jersey	86%
District of Columbia	75%
Arizona	79%
Michigan	79%
Montana	79%
Virginia	82%
Connecticut	86%
New Mexico	71%

16	
New Jersey	86%
North Dakota	81%
Oregon	84%
Tennessee	74%
District of Columbia	75%
Alaska	86%
Oklahoma	74%
Hawaii	86%
Hawaii	86%
Nevada	80%

17	
Pennsylvania	80%
North Carolina	77%
Connecticut	86%
Georgia	78%
Montana	79%
New York	81%
Hawaii	86%
Delaware	79%
Maine	82%
Arkansas	70%

18	
Alabama	71%
Iowa	81%
Louisiana	72%
Texas	76%
Montana	79%
Alaska	86%
Texas	76%
New Hampshire	87%
Indiana	77%
Washington	86%

Part 3 Data Cards: Random Sampling of the Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014 (Continued)

19	
Wisconsin	82%
Utah	87%
Minnesota	85%
Washington	86%
California	83%
Indiana	77%
New Mexico	71%
Louisiana	72%
Maryland	84%
Ohio	80%

20	
Virginia	82%
Massachusetts	86%
District of Columbia	75%
Utah	87%
New Mexico	71%
New York	81%
Florida	80%
New Mexico	71%
Virginia	82%
Oklahoma	74%

21	
Kentucky	75%
Minnesota	85%
Virginia	82%
Louisiana	72%
Maine	82%
Pennsylvania	80%
Ohio	80%
Maine	82%
Montana	79%
Ohio	80%

22	
New Jersey	86%
Kansas	80%
Utah	87%
Massachusetts	86%
Missouri	77%
Virginia	82%
Maine	82%
Wisconsin	82%
New York	81%
Rhode Island	82%

23	
Oklahoma	74%
North Carolina	77%
Michigan	79%
Montana	79%
North Dakota	81%
Washington	86%
Pennsylvania	80%
New Hampshire	87%
Indiana	77%
Tennessee	74%

24	
Wyoming	82%
Virginia	82%
California	83%
Delaware	79%
Nevada	80%
Alaska	86%
Rhode Island	82%
Illinois	81%
Missouri	77%
Virginia	82%

Part 3 Data Cards: Random Sampling of the Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014 (Continued)

25	
Indiana	77%
Indiana	77%
Oregon	84%
New Mexico	71%
Oregon	84%
Arizona	79%
New York	81%
Pennsylvania	80%
Vermont	82%
Arizona	79%

26	
Nebraska	81%
Maryland	84%
Louisiana	72%
Massachusetts	86%
Florida	80%
North Dakota	81%
Wisconsin	82%
New Jersey	86%
Kentucky	75%
Louisiana	72%

27	
Colorado	85%
Kentucky	75%
Alabama	71%
North Carolina	77%
Massachusetts	86%
Missouri	77%
North Dakota	81%
New York	81%
New Hampshire	87%
Kansas	80%

28	
South Carolina	73%
Wisconsin	82%
Alaska	86%
Montana	79%
Illinois	81%
Oklahoma	74%
Oregon	84%
Washington	86%
Michigan	79%
Pennsylvania	80%

During the Activity

Teachers should review the example of sampling variability in part 1 with students and then have them complete the remainder of this part in groups of two to four. Then teachers should discuss the answers as a class.

In part 2, teachers should work with students to create the first dot plot on the board. To plot each value, teachers should ask one student to draw three part 2 data cards, a second student to find the average of their percentages (using a calculator, if needed), and a third student to plot this average on the dot plot. Teachers should repeat this process for 10 sample means, checking students' answers on their own calculators if desired to ensure accuracy.

Teachers should discuss with students what they notice about the dot plot and then direct them to complete part 2 in their groups.

In part 3, teachers should work with students to create the second dot plot on the board. Teachers should distribute one part 3 data card to each student (or pair of students, depending on class size) and instruct them to find the average of their data with a calculator, rounded to one decimal place. Teachers should distribute sticky notes for students to write their sample means on and to place on the board. Teachers can check students' answers on their own calculators if desired to ensure accuracy.

Teachers should discuss with students what they notice about the dot plot and then direct them to complete part 3 in their groups.

After the Activity

Teachers should have students complete the reflection questions in their groups.

Extension Idea

- Teachers could use the part 3 data cards to further investigate the influence of sample size. To make smaller samples, teachers could have each student (or pair of students) choose five percentages from their data card — rather than using all 10 — to find the sample mean (n = 5). To make larger samples, teachers could have students use 20 percentages (two data cards for each student or pair) to find the sample mean (n = 20).

Student Activity

Click ***here*** to download a printable version for students.

Activity Items

There are no separate items for this activity.

Student Learning Objectives

- I will be able to help create dot plots to represent a collection of sample means.
- I will be able to select a random sample from a population.
- I will be able to understand that sample means for different random samples of the same population will differ due to random selection.
- I will be able to understand that there is less sampling variability in the sample mean when the sample size is large than when the sample size is small.

Part 1 – Analyze a Dot Plot

Can a sample of 10 help you learn about a population of 51? In this activity, you will investigate sampling variability in the sample mean.

Example of Sampling Variability

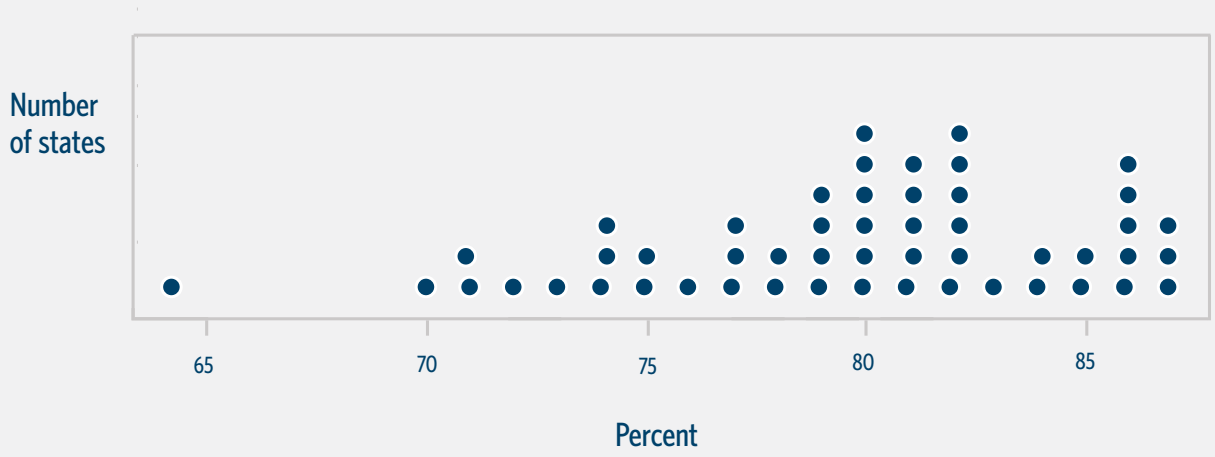
If you average all the numbers from 1 to 99, you get 50. In the table below are eight samples, each with 10 randomly chosen numbers from 1 to 99.

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
59	84	68	88	68	97	9	52
62	77	26	17	12	21	84	19
24	61	25	57	84	88	20	12
36	74	41	62	7	76	52	93
91	26	89	70	26	48	7	64
9	89	49	2	64	17	32	18
14	1	79	70	8	80	70	17
63	35	1	26	42	20	27	9
90	5	87	40	82	50	86	88
36	52	76	22	18	9	16	41
48.4	50.4	54.1	45.4	41.1	50.6	40.3	41.3
Sample means							

Although the values in each sample differ, their means are all clustered close to the actual mean of 50 — though none is exactly 50. This is because of sampling variability.

In 2014, the average percentage of homes with Internet access in all 50 states and the District of Columbia was 79.6 percent. With this information in mind, review the dot plot below and work with your group to answer the following questions:

Percentage of Homes With Internet Access in Each State and the District of Columbia: 2014



1. If the mean is a measure of center, why isn't it at the center of this dot plot?

   The mean is off-center because most of the data values are on the right side of the number line, between 75 and 85, and the most common values have the biggest impact on the mean.

2. What is the range of the data? Show your work.

   The range is the maximum minus the minimum, so: 87 percent - 64 percent = 23 percentage points.

3. How are the data distributed?

   Most of the data values are between 75 percent and 85 percent, but there are a few data points above and below those percentages. Since the value 64 percent is far removed from the others, it could be considered an outlier.

4. Is the data distribution symmetrical: If you folded the dot plot in half from side to side, would the data points on each side match up approximately?

   No, the left and right halves of the dot plot are different.

Part 2 – Create Your First Dot Plot

Working as a class, follow your teacher's guidance to create a dot plot showing the percentage of homes with Internet access in all 50 states and the District of Columbia (where the sample size is $n = 3$, and the number of samples is 10). Use this dot plot to answer the following questions with your group:

No sample dot plot is provided because the sample means for each set will vary.

1. What value represents the center of this dot plot?

   Student answers will vary; some students may find the mean or median while others may just look at the dot plot to see which value appears to be at the center.

2. How does this sample center compare with the actual center of 79.6 percent?

   Student answers will vary.

3. What is the range of these 10 sample means, and how does it compare with the range you found in part 1?

   Student answers will vary, but students should calculate the range and compare it with 23. The range will likely be smaller because the sample size is smaller and because the average of three data points can never be the highest (87) or lowest (64) value in the data set.

4. How are the sample means distributed, and how does this distribution compare with that of the dot plot from part 1?

   Student answers will vary, but students should comment on the data plot's shape, skewness, and/or outliers, making comparisons with the dot plot from part 1.

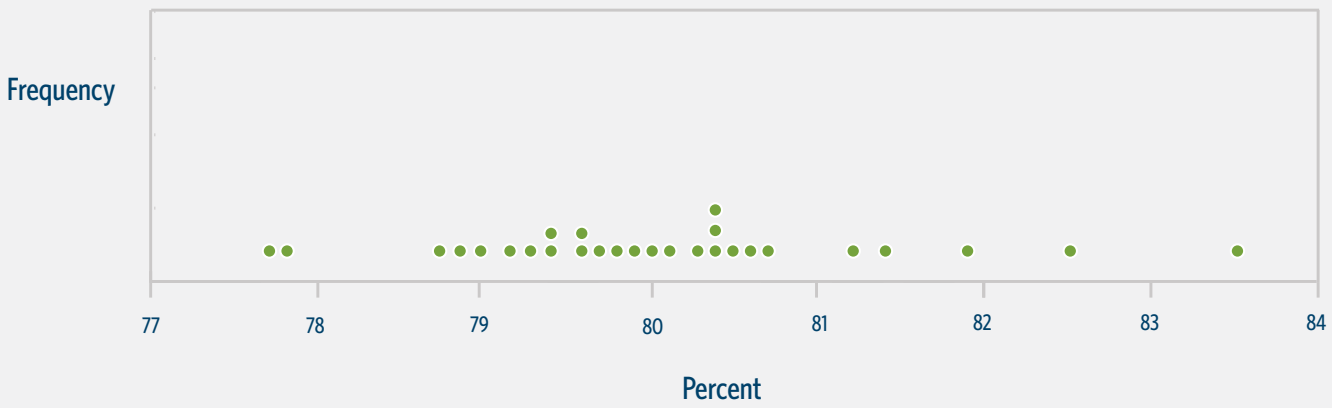
5. Is the distribution more symmetrical or less symmetrical than that from part 1?

   Student answers will vary, but the distribution is unlikely to be symmetrical; however, the distribution of this dot plot will likely be more symmetrical than that from part 1.

Part 3 – Create Your Second Dot Plot

Again working as a class, follow your teacher's guidance to create another dot plot showing the percentage of homes with Internet access in all 50 states and the District of Columbia (where the sample size $n = 10$). Use this dot plot to answer the following questions with your group:

If teachers use all the data cards, the dot plot will look similar to the following:



1. What value represents the center of this dot plot?

   Student answers will vary; some students may find the mean or median while others may just look at the dot plot to see which value appears to be at the center. If teachers use all the data cards, the mean will be 80.1 and the median will be 80.0.

2. How does this sample center compare with the actual center of 79.6 percent?

   Student answers may vary.

3. What is the range of the values, and how does it compare with those found in parts 1 and 2?

   Student answers will vary, but students should calculate the range and compare it with 23 and with the range they found in part 2. If teachers use all the data cards, the range of the means will be 5.8 (83.5 - 77.7), which is much smaller than the range from part 1 and likely much smaller than the range from part 2.

4. How are the sample means distributed, and how does this distribution compare with those of the dot plots from parts 1 and 2?

   Student answers will vary, but students should comment on the data plot's shape, skewness, and/or outliers, making comparisons with the dot plots from parts 1 and 2.

5. Is the distribution symmetrical?

   Student answers will vary. If teachers use all the data cards, the distribution will not be symmetrical; the data will be more spread out on the right.

6. If we repeated the exercise from part 3 with new randomly selected samples of 10 percentages, would you expect the same results?

   Yes. Students should notice that each time a data value was added to the dot plot, it fit in the overall shape.

Reflection Questions:

1. Consider all three dot plots (original, $n = 3$, and $n = 10$):
   a. Which dot plot's distribution is the most spread out?

      Original dot plot

   b. Which dot plot's distribution is the most symmetrical?

      Student answers will vary. It could be the $n = 3$ dot plot, but most likely will be the $n = 10$ dot plot.

2. Of the $n = 3$ and $n = 10$ dot plots, which has a center that is closest to the actual center of 79.6 percent?

   $n = 10$ dot plot

3. Which does a better job of estimating the actual mean of 79.6 percent: one randomly selected value or the average of 10 randomly selected values? Why?

   The average of 10 randomly selected values, because a bigger sample size results in a sample mean that is closer to the actual mean.