Lesson 19: Understanding Variability when Estimating a Population Proportion

Student Outcomes

- Students understand the term *sampling variability* in the context of estimating a population proportion.
- Students know that increasing the sample size decreases sampling variability.

Lesson Notes

In a previous lesson, students investigated sampling variability in the sample mean for numerical data. In this lesson they will investigate sampling variability in the sample proportion for categorical data. Distinguishing between the sample mean for numerical data and the sample proportion for the categorical data is important. This lesson begins with students investigating these differences. It is important that students see the similarity between working with a proportion and what they did in the previous lessons when working with the mean; that is, as the sample size increases, the sampling variability of sample proportions decreases. The dot plots provide students a visual representation of sampling variability. Dot plots are also used to show the connection between sampling variability and sample size.

Before students start the exercises, prepare bags containing   cubes (or beads, or slips of paper, etc.) for each group of four students. Forty cubes should be red, and the remaining   cubes can be any other color (e.g., orange, blue, green, yellow). Also, prepare a number line (illustrated in Example 1) that is visible on the board or on poster paper.

Classwork

As a warm-up activity, discuss how the data used to calculate a proportion (such as the proportion of gym members that are female) is different than the data used to determine a mean (such as the mean time people spend in the gym or the mean number of words in a children's literature book). Other examples of a proportion are: the proportion of people at a particular school who run a marathon, or the proportion of Great Lake perch that are   cm or less. Understanding what a proportion is and how the type of data used to calculate a proportion is different from the type of data they used to calculate a mean is important for the next couple of lessons. This difference is not always easy for students to see, so be deliberate in pointing this out with each new data scenario that is introduced.

Before students begin working with the example in their lesson, discuss the following statistical questions with students. Ask students what data they would collect to answer the questions, and what they would do with the data. Point out to the students those situations in which they would find the average or mean of the data and those situations in which they would find a proportion or a percent.

Question 1: How many hours of sleep do seventh graders get on a night in which there is school the next day?

Data on the number of hours students in a sample of seventh graders sleep would be collected. This would be numerical data and the mean of the numbers would be used to answer this statistical question.

COMMON CORE

Lesson 19:
Date:

Understanding Variability when Estimating a Population Proportion
4/9/14

engage^ny

215

Question 2: What proportion of the students in our school participate in band or orchestra?

A sample of students could be asked to indicate whether or not they participate in band or orchestra. This data would not be numerical. The proportion of the students participating in band for the sample would be used to answer this statistical question. (Point out that it would not be possible to calculate a mean based on this data.)

Question 3: What is the typical weight of a backpack for students at our school?

Data from a sample of students at the school on the weight of backpacks would be collected. This data would be numerical and the mean of the weights would be used to answer this statistical question.

Question 4: What is the likelihood that voters in a certain city will vote for building a new high school?

A sample of voters could be obtained. Each voter in the sample would be asked whether or not they would vote for building a new high school. This data is not numerical. The proportion of the voters who indicated they would vote "yes" for the sample would be used to answer this question.

After the warm-up, proceed with the example in the student lesson.

> In a previous lesson, you selected several random samples from a population. You recorded values of a numerical variable. You then calculated the mean for each sample, saw that there was variability in the sample means, and created a distribution of sample means to better see the sampling variability. You then considered larger samples and saw that the variability in the distribution decreased when the sample size increases. In this lesson, you will use a similar process to investigate variability in sample proportions.

Example 1 (10 minutes): Sample Proportion

Prior to the start of the lesson prepare the following:

- A bag containing ____ cubes (or beads, or slips of paper, etc.) for each group of four students. Forty should be red, and the remaining ____ cubes can be any other color.
- On the board draw a number line similar to the one shown below.

| 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |

Begin the lesson by saying, "Today we are going to learn something about the population of cubes in this bag. It is full of cubes. Based on your previous work, how could you learn something about the whole bag of cubes?" After this discussion, instruct each student to take a random sample of ____ cubes from the bag. This sample should be taken with replacement. You may want to begin by taking a sample as a demonstration.

> **Example 1: Sample Proportion**
>
> Your teacher will give your group a bag that contains colored cubes, some of which are red. With your classmates, you are going to build a distribution of sample proportions.

Each student should record the color of each draw, and then calculate the proportion of red in their sample of 10.

▪ How would you determine the sample proportion of red cubes?

□ _____

Have students place post-it notes on the number line to build a dot plot. After all of the post-it notes have been placed, explain to students that the graph is the *sampling distribution* of the sample proportions.

Exercises 1–6 (10 minutes)

Allow students to work with their groups on Exercises 1–6. Then discuss as a class.

As students discuss the answers to Exercises 4 and 5, note that the shape of the distribution of sample proportions is often mound shaped. For the variability students need to focus on the overall spread in the sample proportions and the main cluster around the center of the distribution.

An extension is to have the students calculate MAD for the distribution.

Exercises 1–6

1. **Each person in your group should randomly select a sample of ___ cubes from the bag. Record the data for your sample in the table below.**

   Students' tables will vary based their samples.

Cube	Outcome (Color)
	Green
	Orange
	Red
	Red
	Blue
	Orange
	Red
	Green
	Blue
	Blue

2. **What is the proportion of red cubes in your sample of ___?**

   This value is called the sample proportion. The sample proportion is found by dividing the number of "successes" (in this example, the number of red cubes) by the total number of observations in the sample.

   Student results will be around ___. In this example, the sample proportion is ___.

---

3. Write your sample proportion on a post-it note and place it on the number line that your teacher has drawn on the board. Place your note above the value on the number line that corresponds to your sample proportion.

The graph of all the students' sample proportions is called a sampling distribution of the sample proportions.

This is an example of a dot plot of the sampling distribution.

Dot Plot of Sample Proportions for n=10
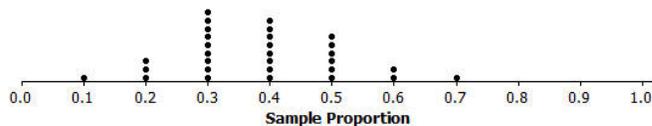
4. Describe the shape of the distribution.

Nearly symmetrical distribution that is clustered around   .

5. Describe the variability in the sample proportions.

The spread of the data is from   to   . Much of the data clusters between   and   .

6. Based on the distribution, answer the following:

a. What do you think is the population proportion?

Based on the dot plot, an estimate of the population proportion is approximately   .

b. How confident are you of your estimate?

Because there is a lot of variability from sample to sample (to   ), there is not a lot of confidence in the estimate.

Example 2 (10 minutes): Sampling Variability

This example parallels Example 1. Students now take a random sample of   cubes from the bag. This sample should be taken with replacement.

Example 2: Sampling Variability

What do you think would happen to the sampling distribution if everyone in class took a random sample of   cubes from the bag? To help answer this question you will repeat the random sampling you did in Exercise 1, except now you will draw a random sample of   cubes instead of   .

Students record their sample results and find the sample proportion of red cubes in the bag. Again they should record the proportion on a post-it note.

Draw another number line on the board using the same scale as in Example 1.

Exercises 7–15 (10 minutes)

Allow students to work with their groups on Exercises 7–10. Then discuss the answers as a class.

Exercises 7–15

What do you think would happen to the sampling distribution if everyone in class took a random sample of   cubes from the bag? To help answer this question you will repeat the random sampling you did in Exercise 1, except now you will draw a random sample of   cubes instead of   .

7.   Take a random sample of   cubes from the bag. Carefully record the outcome of each draw.

   Answers will vary. An example follows:

Blue	Green	Red	Yellow	Blue	Red	Red	Green	Blue	Red

Green	Red	Red	Green	Yellow	Blue	Blue	Yellow	Red	Blue

Red	Green	Red	Red	Blue	Red	Yellow	Green	Green	Yellow

8.   What is the proportion of red cubes in your sample of   ?

   Answers will vary. In this example the sample proportion is ⎯ or   .

9.   Write your sample proportion on a post-it note and place the note on the number line that your teacher has drawn on the board. Place your note above the value on the number line that corresponds to your sample proportion.

   An example of a dot plot:

Dot Plot of Sample Proportions for n=30
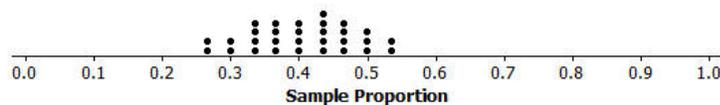
10.   Describe the shape of the distribution.

   Mound shaped, centered around

Students should work independently on Exercises 11–15. Then discuss as a class.

Lesson 19:	Understanding Variability when Estimating a Population Proportion
Date:	4/9/14

11. Describe the variability in the sample proportions.

    Spread of the data is from _____ to _____ . Most of the data clusters between _____ and _____ .

12. Based on the distribution, answer the following:

    a. What do you think is the population proportion?

       Based on the dot plot, an estimate of the population proportion is approximately _____ .

    b. How confident are you of your estimate?

       Because there is a less variability from sample to sample (_____ to _____), there is more confidence in the estimate.

    c. If you were taking a random sample of _____ cubes and determined the proportion that was red, do you think your sample proportion will be within _____ of the population proportion? Explain.

       Answers depend on the dot plots prepared by students. If the dot plot in Exercise 9 is used as an example, note that only about half of the dots are between _____ to _____ . There are several samples that had sample proportions that were farther away from the center than _____ , so the sample proportion might not be within _____ of the population proportion.

13. Compare the sampling distribution based on samples of size _____ to the sampling distribution based on samples of size _____ .

    Both distributions are mound shaped and center around _____ . Variability is less in the sampling distribution of sample sizes of _____ versus sample sizes of _____ .

14. As the sample size increased from _____ to _____ describe what happened to the sampling variability of the sample proportions.

    The sampling variability decreased as the sample size increased.

15. What do you think would happen to the variability of the sample proportions if the sample size for each sample was _____ instead of _____ ? Explain.

    The variability in the sampling distribution for samples of size _____ will be less than the variability of the sampling distribution for samples of size _____ .

Closing (5 minutes)

To highlight the Lesson Summary, ask students the following questions:

- How do you know if the data collected is summarized by a mean or by a proportion?
  - *Anticipate responses that indicate the different types of data collected and how the data is summarized. Examples similar to those discussed in the lesson would indicate an understanding of when a mean is calculated and when a proportion is calculated.*

- What would a dot plot of a sampling distribution of the sample proportion look like if proportions from many different random samples were used to create the plot?
  - *Anticipate responses that describe a dot plot in which each dot represents the sample proportion of one sample. The graph would show variability (or sampling variability) with a cluster around a value close to the value of the proportion for the population. Consider asking students to sketch a dot plot that they might get for one of the examples in which a proportion is used to answer the statistical question.*

- What happens to the sampling distribution as the sample size increases?
  - *Anticipate responses that indicate the sampling variability (spread of the dots in the dot plot) decreases as the sample size increases.*

Lesson Summary

- **The sampling distribution of the sample proportion is a graph of the sample proportions for many different samples.**
- **The mean of the sample proportions will be approximately equal to the value of the population proportion.**
- **As the sample size increases the sampling variability decreases.**
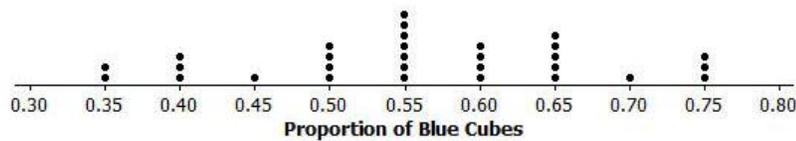

Exit Ticket (5 minutes)

Name _____   Date_____

Lesson 19: Understanding Variability when Estimating a Population Proportion

Exit Ticket

A group of seventh graders took repeated samples of size   from a bag of colored cubes. The dot plot below shows the sampling distribution of the sample proportion of blue cubes in the bag.

Proportion of Blue Cubes

1. Describe the shape of the distribution.

2. Describe the variability of the distribution.

3. Predict how the dot plot would look differently if the sample sizes had been   instead of

Exit Ticket Sample Solutions

A group of seventh graders took repeated samples of size ___ from a bag of colored cubes. The dot plot below shows the sampling distribution of the sample proportion of blue cubes in the bag.

0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80
Proportion of Blue Cubes
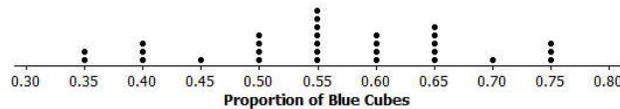
1. Describe the shape of the distribution.

    Mound shaped, centered around   .

2. Describe the variability of the distribution.

    The spread of the data is from   to   with much of the data between   and   .

3. Predict how the dot plot would look differently if the sample sizes had been   instead of   .

    The variability will decrease as the sample size increases. The dot plot will be centered in a similar place but will be less spread out.

Problem Set Sample Solutions

1. A class of seventh graders wanted to find the proportion of M&M's that are red. Each seventh grader took a random sample of   M&M's from a very large container of M&M's. Following is the proportion of red M&M's each student found.

    a. Construct a dot plot of the sample proportions.

Dot Plot of Proportion of Red M&Ms

0.00 0.05 0.10 0.15 0.20 0.25 0.30
Proportion of Red M&Ms
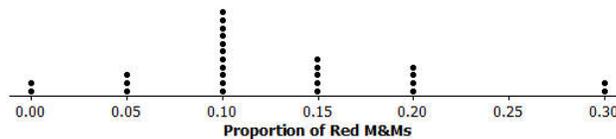
    b. Describe the shape of the distribution.

    Somewhat mounded shaped, slightly skewed to the right.

    c.   **Describe the variability of the distribution.**

        Spread of the data is from   to   . Most of the data clusters between   and   .

    d.   **Suppose the seventh grade students had taken random samples of size   . Describe how the sampling distribution would change from the one you constructed in part (a).**

        The sampling variability would decrease.

2.   **A group of seventh graders wanted to estimate the proportion of middle school students who suffer from allergies. One group of seventh graders each took a random sample of   middle school students, and another group of seventh graders each took a random sample of   middle school students. Below are two sampling distributions of the sample proportions of middle school students who said that they suffer from allergies. Which dot plot is based on random samples of size   ? How can you tell?**
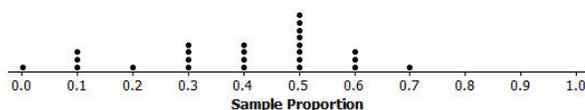
Dot Plot A

Dot Plot of Sample Proportion
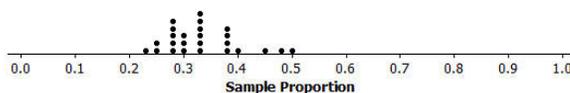
Dot Plot B

Dot Plot of Sample Proportion

    Dot Plot   because the variability of the distribution is less than the variability in Dot Plot   .

3.   **The nurse in your school district would like to study the proportion of middle school students who usually get at least eight hours of sleep on school nights. Suppose each student in your class plans on taking a random sample of   middle school students from your district, and each calculates a sample proportion of students who said that they usually get at least eight hours of sleep on school nights.**

    a.   **Do you expect everyone in your class to get the same value for their sample proportion? Explain.**

        No, we expect sample variability.

    b.   **Suppose each student in class increased the sample size from   to   . Describe how you could reduce the sampling variability.**

        I could reduce the sampling variability by using the larger sample size.