



Math Objectives

- Students will be able to identify situations in which stratification might reduce variation.
- Students will recognize that some stratification designs are useful in reducing variation and some are not.
- Students will recognize that random selection of subject(s) must occur within each stratum.
- Students will recognize that the subjects *within* each stratum should be as similar as possible and different from the subjects in the other strata based on a variable that has the potential to affect the outcome of the study.
- Look for and make use of structure (CCSS Mathematical Practices).
- Construct viable arguments & critique the reasoning of others (CCSS Mathematical Practices).

Vocabulary

- random selection
- simple random sample (SRS)
- stratified random sample
- stratum(a)
- survey

About the Lesson

- This lesson involves determining which of three different sampling methods—a simple random selection design and two stratified selection designs—would be most beneficial in selecting a survey sample within a given context.
- As a result, students will:
 - Use a simple random sample (SRS) to select survey participants.
 - Use two different stratified random sampling designs to select survey participants.
 - Compare means and variability for the three sampling designs to determine the most effective design for this context.



TI-Nspire™ Technology Skills:

- Download a TI-Nspire document
- Open a document
- Move between pages
- Grab and drag a point

Tech Tips:

- Make sure the font size on your TI-Nspire handhelds is set to Medium.
- You can hide the function entry line by pressing **ctrl** **G**.

Lesson Files:

Student Activity
Stratified_Sampling_Student.pdf
Stratified_Sampling_Student.doc

TI-Nspire document
Stratified_Sampling.tns

Visit www.mathnspired.com for lesson updates and tech tip videos.



TI-Nspire™ Navigator™ System

- Transfer a File.
- Use Screen Capture to examine patterns that emerge.
- Use Teacher Edition computer software to review student documents.
- Use Quick Poll to assess students' understanding.

Related Activities

- TI Statistics Nspired Blocking Introduction
- TI Statistics Nspired Effective Blocking

Discussion Points and Possible Answers

Tech Tip: Pages 1.2 and 1.3 contain directions for students to “seed” their handhelds for the activity.

Teacher Tip: This activity is most effective using software or handhelds that show colors.

Teacher Tip: Once they have seeded their handhelds, it is not necessary to do so again unless you have a reason for wanting them all on the same page with the same results. Because students are generating a random selection of subjects to survey, results will vary. Occasionally a student will get different overall conclusions than what is described here due to this randomness.

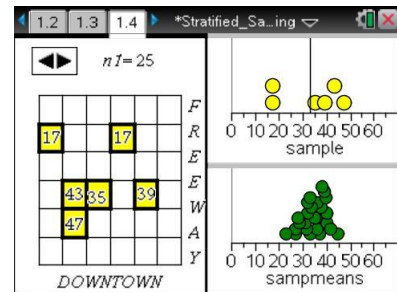
The Happy Sundae Ice Cream Company is looking to expand its ice cream truck business into new neighborhoods. From previous business experience, the company knows how many children need to live in a neighborhood for an ice cream truck to be profitable.

The neighborhood being considered is bounded by the downtown on the south and a freeway on the east and consists of 36 approximately square city blocks. Sean, Happy Sundae's Director of Marketing, wants to survey six of the city blocks to get an estimate of the total number of children living in the neighborhood.



Move to page 1.4.

- The first sampling method Sean decides to use is a simple random sample (SRS). The left work area displays the 36 city blocks in which Sean is interested. Use the right arrow once to randomly select six city blocks to survey (you are actually *simulating* what would happen if these surveys were really conducted).



Note: The left arrow will erase all the samples and allow you to begin again.

- A number is displayed in each "block", representing the simulated number of children in the city block surveyed.
- The top dotplot in the right work area of the screen displays the number of children found in each of the city blocks surveyed.
- The lower dotplot displays the mean of the sample.

Teacher Tip: Make sure students click the arrow and wait a few seconds for the sample to be selected.

TI-Nspire Navigator Opportunity: Screen Capture

See Note 1 at the end of this lesson.

- The vertical line on the top dotplot represents the mean number of children in the sample of six city blocks surveyed. This value for the mean also appears as a dot in the lower half of the screen. What would a mean of 40 represent in the context of the entire neighborhood and the ice cream truck business?

Sample Answers: A mean of 40 for the six city blocks in one sample would suggest an estimate of 1440 children in the entire neighborhood. Six of the 36 city blocks were sampled and an average of 40 was found for those six. Applying that average of 40 to all 36 city blocks would give 1440 for the neighborhood. The company could compare 1440 to the number of children required for a successful ice cream truck route and make their decision.

Teacher Tip: Students will generate different random samples based on the seeding of their calculators. Answers in this Teacher Document reflect only one sample answer, and student results will vary. All students will likely generate the same overall conclusions in the activity regardless of



the random samples they generated.

- b. Click on the right arrow to create another simulation of a simple random sampling for 6 city blocks. Write down any observations you have regarding the number of children in the neighborhood for this sample.

Sample Answers: The numbers of children in each of the six city blocks sampled was different. The mean this time was around 32.

- c. Click the arrow to create 23 more simulated samples (of 6 city blocks each) for a total of 25 simulations, each time observing the number of children in each city block sampled. Fill in the table below with the results from each simulated sample. The number of children in each city block is fixed, so you only fill in a cell once. What pattern(s) do you see in the number of children? Explain your reasoning.

Sample Answers: Students will not get all cells of the table filled in with their samples, but the patterns should be visible (table below shows all 36 blocks). There might be two patterns in the neighborhood regarding location of children. First, more children appear to be closer to downtown and fewer farther from downtown. Second, there might be a difference in children based on proximity to the freeway, with fewer children closer to the freeway.

8	5	15	14	7	12
17	16	23	17	21	24
35	30	26	29	32	28
43	43	35	39	39	37
47	47	53	49	48	52
65	63	56	59	61	61

- d. What is the mean number of children per city block? Describe the number of children possibly living in the entire 36-block neighborhood according to this strategy.

Sample Answers: The plot displays the mean number of children for different samples of city blocks. The mean of the distribution of sample means is 32, which would say the number of children would be about 36 times 32 or 1152 children.



2. In thinking about the neighborhood, Sean suspects that there might be fewer children in the blocks close to the freeway and that a simple random sample might not be the best sampling method to use. Not having a block near the freeway as part of the sample could potentially overestimate the mean number of children in the neighborhood. He decides to redesign his sampling to account for this factor. Stratified random sampling involves dividing the population into similar sub-populations or strata before sampling. The strata need to be as homogeneous (similar) as possible and not overlap.

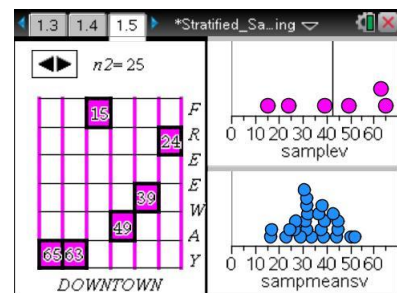
Teacher Tip: With stratification, we take a population and subdivide into two or more sub-groups called strata. Here the strata were based on proximity to a geographical feature, downtown or the freeway. The purpose is to decrease the diversity or noise in the subgroups to get a better representation of the population and reduce the variability in the possible samples.

- a. If Sean suspects that there are fewer children closer to the freeway, describe how he might create strata to reflect this and explain your reasoning.

Sample Answers: Since the freeway runs parallel to the neighborhood on the east side, Sean could divide the neighborhood into six vertical strata, each one containing city blocks approximately the same distance from the freeway.

Move to page 1.5.

- b. This page shows a simulation of stratified random sampling using the freeway as the factor. Six city blocks need to be selected. How would you suggest selecting them using vertical strata?



Sample Answers: One city block should be randomly selected from each of the vertical strata, i.e., one from each column.

Teacher Tip: Some students might think that you select all of the city blocks within one stratum. This would be cluster sampling.



- c. Click on the arrow to see the number of children per block in a stratified random sample of city blocks surveyed (freeway as the stratifying factor). Observe the plot and the dotplot on the right, and write down any observations you have about the number of children.

Sample Answers: The mean number of children per city block was 37.5, close to the number I found earlier.

- d. Click the arrow 24 more times for a total of 25 simulations, each time observing the mean number of children in the plot and the dotplot on the right. The vertical line in the top panel represents the mean of the sample. What is the mean number of children per city block? Describe the possible number of children in the entire neighborhood if using this stratified sampling strategy. Compare this to the answer you found for simple random sampling.

Sample Answers: Applying an average of 34.5 to all 36 city clocks would give a prediction of 1242 children for the neighborhood. For simple random sampling, I found the mean would be around 32, which would give me 90 fewer children. The company could compare this to the number of children required for a successful ice cream truck route and make their decision.

- e. Explain the variation in the plot of means.

Sample Answers: The plot of means represents different combinations of city blocks selected to be sampled. Because these are randomly selected within the strata, the number of children found in the selected city blocks changes, and therefore the mean number of children changes.

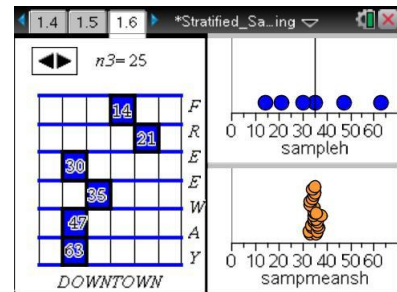
3. Sean remembers an article he read in last Sunday's newspaper that suggested there are fewer children in neighborhoods farther from downtown. He decides to redesign his sampling one more time based on this information.
- a. The strata need to be as homogeneous as possible and not overlap. If Sean suspects that there are more children in neighborhoods closer to downtown and fewer children farther from downtown, describe how to create six strata to reflect this and how to choose a sample from the strata.

Sample Answers: Since downtown runs parallel to the neighborhood on the south side, Sean should divide the neighborhood into six horizontal strata, each one containing city blocks approximately the same distance from the downtown. He should then randomly select one block from each strata.



Move to page 1.6.

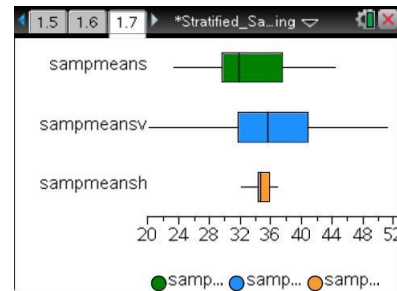
- b. Click on the arrow 25 times to see a simulated sampling distribution of the number of children in a stratified random sample (downtown as the stratifying factor) of city blocks surveyed. Observe the variation in the mean number of children per city block for this stratified sampling strategy. How does the mean number of children compare to the means you found earlier? What would this suggest about the total number of children in the 36-block neighborhood?



Sample Answers: The mean number of children per city block, 34, is similar to the number I found in both earlier samplings. Applying an average of 34 to all 36 city clocks would give 1224 for the neighborhood. The company could compare this estimate to the number of children required for a successful ice cream truck route and make their decision.

Move to page 1.7.

4. This page shows boxplots of the sample means for all three of the sampling designs used—simple random sample, stratified vertically (parallel to the freeway), and stratified horizontally (parallel to downtown).
- a. Compare the centers of the three boxplots. Are they similar, different? How many total children does each represent? Explain your answers.



Sample Answers: From the boxplots, the simple random sample had a median of 36.58, representing about 1317 total children; the samples stratified vertically (parallel to the freeway) had a median of 34.33, representing about 1236 children; and the samples that were stratified horizontally (parallel to downtown) had a median of 35.25, representing about 1269 children. The three median values are relatively similar.

- b. Compare the variation in the three boxplots.

Sample Answers: From the boxplots, the simple random sample had an IQR of approximately 5, stratified vertically (parallel to the freeway) had an IQR of approximately 8.8, and stratified horizontally (parallel to downtown) had an IQR of approximately 2.2. The horizontal stratification produced a much smaller variation.



- c. Which factor, the freeway or downtown, seems to be relevant to the sample selection? Explain.

Sample Answers: Downtown seems to be the relevant factor in sample selection. When samples were stratified parallel with the downtown, there was much less variation in the samples.

Teacher Tip: You might want to skip question #5 and see whether the sense of overall agreement—about 35 children as a mean—comes through the work rather than being told.

5. The actual mean is 34.89 children per block. How many total children does this represent for the entire neighborhood? Do each of the sampling designs reflect this total? Explain.

Sample Answers: The total number of children in the neighborhood would be 1256. All three of the sampling distributions have medians reasonably close to the actual median and would give good values for the entire neighborhood.

6. The goal of the survey was to estimate how many children are in the neighborhood. Sean will only get to select one sample on which to base his decision regarding the new ice cream route. Which method—simple random sample, stratified horizontally, or stratified vertically—should the Director of Marketing use to most effectively eliminate the influence of a factor? Take into consideration the center and the spread of the simulated sampling distributions as you explain your reasoning.

Sample Answers: All three sampling distributions had reasonable estimates of the center (and therefore total number of children). The Director of Marketing will only be given one chance at the real survey, and he wants to minimize the risk of being wrong. The horizontal stratification had the smallest variability, so it is the one that can best eliminate the effects of the downtown, the factor that appeared to have a strong influence on the numbers of children.

TI-Nspire Navigator Opportunity: Quick Poll

See Note 2 at the end of this lesson.

Wrap Up

Upon completion of the lesson, the teacher should ensure that students are able to understand:

- One reason for using a stratified sample design is to eliminate the effects of another outside variable on the study.
- Stratification might reduce variation, and some stratification designs are useful in reducing variation while others are not.
- Random selection of subject(s) must occur within each stratum.



Assessment

1. The first step in a stratified sample design is to create strata neighborhood blocks. If the downtown affects the numbers of children, how would you select the strata so that each contains neighborhood blocks that are as similar as possible? Fill in the chart below to show the strata you would select and give a brief rational for your choices.

Sample Answers:

a	a	a	a	a	a
b	b	b	b	b	b
c	c	c	c	c	c
d	d	d	d	d	d
e	e	e	e	e	e
f	f	f	f	f	f

The cells labeled f are all closest to the downtown area and should form one strata; similarly the other rows labeled with the same letters form the other strata for sampling purposes. Strata should be as similar as possible within and as different as possible without—the distance from downtown appeared to be a useful variable for stratification.

2. One reason for using a stratified sample design is to eliminate the effects of another outside variable on the study. Explain why this is important.

Sample Answers: An outside variable sometimes creates a lot of variability in a population. Using a stratified sample might help reduce that variability and might select a sample that is more representative of the population than a simple random sample might be.

TI-Nspire Navigator

Note 1

Question 1, Name of Feature: Screen Capture

Students will get different data because of the random selection of subjects. Use Screen Capture to facilitate discussion of these differences and the randomness behind them is important.

Note 2

Question 6, Name of Feature: Quick Poll

A Quick Poll can be given at the conclusion of the lesson. You can save the results and show a Class Analysis at the start of the next class to discuss possible misunderstandings students might have.