Building Concepts: Sample Means

TEACHER NOTES

Lesson Overview

In this TI-Nspire lesson students use simulation to explore how sample means (and medians) vary from sample to sample when repeated samples are taken from the same population.

The mean of a simulated sampling distribution of sample means for a given sample size is close to the mean of the population from which the samples were drawn.

Learning Goals

1. Distinguish sample variability within a single sample from sampling variability among statistics (i.e., means or medians) that occurs between samples;
2. recognize that sampling distributions of sample means are mound shaped and symmetric with a fairly predictable span for a given sample size;
3. describe how sample size affects the spread of a sampling distribution of sample means.

Prerequisite Knowledge

Sample Means is the twentieth lesson in a series of lessons that explore the concepts of statistics and probability. In this lesson students explore sample variability within a single sample. This lesson builds on the concepts of the previous lessons. Prior to working on this lesson students should have completed *Probability and Simulation*, *Choosing Random Samples* and *Sample Proportions*. Students should understand:

- how to find and interpret measures of central tendency;
- that random sampling is likely to produce a sample that is representative of the population;
- how to use simulation to collect data.

Vocabulary

- **sample:** representative of the population from which it was drawn
- **sample size:** the quantity of data in a sample
- **mean:** the sum of all the data values in a set of data divided by the number of data values
- **sampling distribution:** the collection of sample statistics from all possible samples of a given size from a specific population
- **simulated sampling distributions:** modeling a collection of sample statistics from a specific population

Lesson Pacing

This lesson should take 50–90 minutes to complete with students, though you may choose to extend, as needed.

Lesson Materials

- Compatible TI Technologies:

  TI-Nspire CX Handhelds, TI-Nspire Apps for iPad®, TI-Nspire Software
- Sample Means_Student.pdf
- Sample Means_Student.doc
- Sample Means.tns
- Sample Means_Teacher Notes
- To download the TI-Nspire activity (TNS file) and Student Activity sheet, go to http://education.ti.com/go/buildingconcepts.

Class Instruction Key

The following question types are included throughout the lesson to assist you in guiding students in their exploration of the concept:

Class Discussion: Use these questions to help students communicate their understanding of the lesson. Encourage students to refer to the TNS activity as they explain their reasoning. Have students listen to your instructions. Look for student answers to reflect an understanding of the concept. Listen for opportunities to address understanding or misconceptions in student answers.

Student Activity: Have students break into small groups and work together to find answers to the student activity questions. Observe students as they work and guide them in addressing the learning goals of each lesson. Have students record their answers on their student activity sheet. Once students have finished, have groups discuss and/or present their findings. The student activity sheet can also be completed as a larger group activity, depending on the technology available in the classroom.

Deeper Dive: These questions are provided for additional student practice and to facilitate a deeper understanding and exploration of the content. Encourage students to explain what they are doing and to share their reasoning.

Mathematical Background

In Lesson 19, *Sample Proportions*, students investigated random samples and how sample proportions varied from sample to sample. The primary focus of this lesson is to understand that means (and medians) computed from samples will also vary from sample to sample and making informed decisions based on such sample statistics requires some knowledge of the amount of variation to expect. Just as for proportions, a good way to gain this knowledge is through simulation; beginning with understanding the behavior of samples from a known population. In investigating random samples, students should clearly distinguish between the variation within a sample and the variation from sample to sample when samples of the same size are taken. Without technology, students' thinking may be limited to one observed random sample, which provides no background for thinking in terms of all the possible samples that could have been drawn, the variability across all those samples, and how their sample might fit into that range of possibilities.

Students should realize that one of the reasons for taking a random sample is to learn something about the population from which the sample was selected, but a sample *statistic* is only an estimate of a corresponding population *parameter* and there will be some difference between the two. Understanding variability in the sampling process allows the investigator to gauge the expected size of the difference between the sample statistic and the population parameter it estimates. The variability among samples can be studied by means of simulations.

Three types of distributions are related to sampling: the distribution of a population itself, the distribution of a sample from that population, and the *sampling distribution* of a sample statistic from that population. Each of these distributions can be described in terms of shape, center and spread. The sampling distribution of sample means for a given sample size *n* would include the means of all possible samples of size *n* from the population. Typically, it is not feasible to generate the entire sampling distribution. It is possible to simulate a subset of a sampling distribution of sample means using random numbers to select the samples, which introduces yet a fourth distribution, a *simulated sampling distribution*. A reasonable number of samples must be taken in order for the simulated sampling distribution to closely model the sampling distribution. Students often fail to recognize that a sampling distribution is a distribution of a sample statistic; a typical misunderstanding is to confuse the types of distributions.

Resources:

Source: Natural History Magazine, March 1974, copyright 1974; The American Museum of Natural History; and James G. Doherty, general curator, The Wildlife Conservation Society;
http://www.thetravelalmanac.com/lists/animals-speed.htm;

http://en.wikipedia.org/wiki/Fastest_animals

http://www.demogr.mpg.de/longevityrecords/0203.htm

http://web.stanford.edu/group/stanfordbirds/text/essays/How_Long.html

http://www.vetstreet.com/our-pet-experts/meet-eight-of-the-fastest-dogs-on-the-planet

http://canidaepetfood.blogspot.com/2012/08/which-dog-breeds-are-fastest.html

http://slimdoggy.com/dog-life-spans/

http://users.pullman.com/lostriver/breeddata.htm

http://visual.ly/shark-speed-worlds-fastest-sharks

http://www.elasmo-research.org/education/topics/r_haulin'_bass.htm

http://whaleopedia.org/animalfund/harbor-porpoise/

http://www.enchantedlearning.com/subjects/whales/species/

http://wwf.panda.org/what_we_do/endangered_species/cetaceans/about/fin_whale/

http://www.speedofanimals.com/animals/sea_otter

http://www.thetravelalmanac.com/lists/fish-speed.htm

http://www.dfo-mpo.gc.ca/Library/333800.pdf

http://www.petcarerx.com/article/the-average-lifespan-of-a-cat-breed-by-breed-chart/1698

http://www.infoplease.com/ipa/A0004737.html

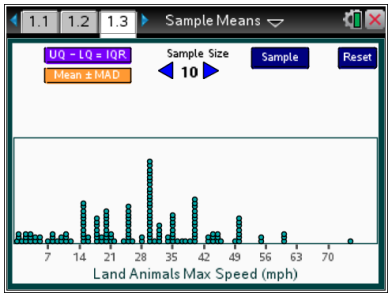
http://a-z-animals.com/animals/donkey/

http://a-z-animals.com/animals/galapagos-giant-tortoise/

http://a-z-animals.com/animals/aldabra-giant-tortoise/

Part 1, Page 1.3

Focus: In general, while random samples vary from sample to sample, a random sample will be fairly representative of the population from which it was drawn.



Page 1.3 shows a dot plot of land animals' maximum speeds.

Population allows a choice of all animals or all land animals maximum recorded speeds or life spans.

Sample displays a random sample from the chosen population.

Measures allows the display of the interval determined by the mean+/-MAD or IQR for a sample drawn from that population.

The arrows on the keyboard or the screen change the sample size.

Reset returns to the original screen.

TI-Nspire Technology Tips

menu accesses page options.

tab moves between Sample and Reset buttons.

enter takes a sample.

ctrl del resets the page.

Class Discussion

These questions explore variation within a sample, how a sample relates to the population from which it was selected, and make the transition from a finding a mean for a single sample to creating a simulated distribution of sample means.

The TNS activity contains the maximum recorded speeds and life spans of types of animals. Many of the animals are the same as those from Lesson 1,* Introduction to Data*, but others, such as reptiles, have been added.

- ***Describe the distribution of the maximum recorded speeds of the types of land animals shown on page 1.3.***

  Answers may vary. The distribution of speeds is spread out and slightly symmetric with most of the speeds between 14 mph and 38 mph. One animal (the cheetah, identified in earlier units) has a speed that is about 15 mph faster than all of the others.

- ***Select Sample. What is the population? What is the attribute of interest? What does the plot on the top represent?***

  Answer: The population is all the land animals in the file and the attribute of interest is the maximum recorded speed. The plot on the top represents a random sample of size 10 of those speeds.

- ***Compare the plot on the top to the population.***

  Answers will vary. In many cases, the distribution of the sample speeds will be fairly representative of the population speeds but just kind of spread out.

Class Discussion (continued)

- ***Generate more samples. Overall, how do the samples compare to each other?***

  Answers will vary. Most of them seem to spread out along the range of the population. One or two of the distributions cluster on the left, and a very few contain the 75 mph measurement.

Student Activity Questions—Activity 1

1. Select *menu> Measures> Mean +/-MAD*.

a. The mean maximum recorded speed of all of the land animals is about 27 mph and the MAD is about 12 mph. How do the mean and mean absolute deviations of the sample and of all the land animals compare?

Answers will vary. For some samples, the mean maximum recorded speed and the MAD of the sample will be fairly close to the mean maximum recorded speed and MAD of all the land animals.

b. Generate another sample and estimate the difference between the mean speed of the animals in the sample and the mean speed of all the land animals. Estimate the difference between the MAD of the sample and the MAD of the population.

Answers will vary. In one sample, the mean was about 24 mph and the MAD about 6 mph so the difference would be 3 mph for the means and 6 mph for the MADs.

c. Generate four more samples and answer question b for each sample. Then compare your answers to the answers someone else found. Overall, how would you describe the difference between the mean speed and the MAD for the sample and for all the land animals?

Answers will vary.

	Estimated sample mean speed	Difference: pop (27)-sample	Estimated sample MAD	Difference: Pop (12)-sample
1	33	-6	$44-33=11$	1
2	30	-3	$44-30=13$	-1
3	28	-1	$44-28=16$	-4
4	27	0	$42-27=15$	-3

Answers will vary. The difference between the mean speed of the sample of land animals and the mean speed of all the land animals in the TNS activity is typically less than 7 or 8 mph. The difference between the MADs is typically small, usually less than 3 or 4.

2. Use the TNS activity to determine which of the following statements seem to be true.

a. Over a lot of samples, the mean maximum speeds of the samples of land animals were close to the mean of the maximum speeds of the land animals.

Answers may vary. This seems to be a true statement, given our work in the previous question, if close is taken to be less than three or four mph difference.

Student Activity Questions—Activity 1 (continued)

b. Many of the sample distributions were relatively similar to the population distribution in terms of center and spread.

Answers may vary. This seems to be true from the samples we did.

c. When the sample contained the cheetah at 75 mph, the mean maximum speed of the sample of land animals was larger than the mean speed of all of the land animals.

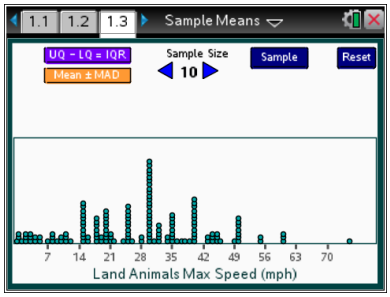
Answers may vary, but this seems to be true because every sample that had the 75 mph in the sample had a higher mean. This might be because there are only 10 speeds in the sample, and 75 is 15 mph larger than any of the other speeds.

d. The mean speed of the sample of land animals was always larger than the mean maximum speed of all of the land animals in the TNS activity.

Answers will vary. This is not supported by the samples generated in the activity.

Part 2, Page 1.3

Focus: Students establish the fact that a sampling distribution of sample means, for a fairly large sample, will be mound shaped and symmetric about the population mean and will not necessarily have the shape of the population.



Class Discussion

The focus of these questions is on establishing the fact that a sampling distribution of sample means, for a fairly large sample, will be mound shaped and symmetric about the population mean and will not necessarily have the shape of the population.

Have students…	Look for/Listen for…
Select ***menu> Population> all animals mph***.	
• ***Describe the distribution of the maximum recorded speeds of all the animals.***	Answers may vary. The distribution of speeds is slightly skewed to the right with at least one outlier, the peregrine falcon at 240 mph (identified in earlier work). Most of the speeds seem to be between 18 mph and 40 mph.

Class Discussion (continued)

- ***Select* Sample*. Compare the plot on the top to the population.***

  Answers may vary. The sample speeds on the top plot were from 0 to about 58 mph and bunched up a bit around 28 or 30. This is less spread out than the population distribution of speeds. The middle is fairly similar since the middle group of speeds for the population seems to be between 28 and 30 as well.

- ***Generate a second sample. How did this sample compare to the first one?***

  Answers will vary. For example, the second sample has a larger range of 70, from about 0 to 70 but the center was similar, 30 to about 32.

Student Activity Questions—Activity 2

1. Select *menu> Measures> IQR = UQ-LQ*.

a. Select *Sample*. How do the IQRs of the sample and the population compare?

Answers will vary. Some students might point out that the medians are about the same, but the IQR for the sample speeds is larger. Others may have a much smaller IQR.

b. Select *Sample* again and record the IQRs. Do this four times and record your results in the table. Make a conjecture about the difference between the IQR of the sample of land animals and the IQR of all the animals.

Answers will vary. Examples are below. A conjecture might be that the IQRs seem to be typically larger for the animals in the samples than for all of the animal types.

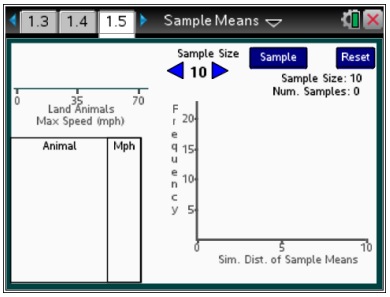
Sample	Sample IQR	Difference: All: IQR about $30-20=10$
1	$40-22$	18
2	$50-28$	22
3	$30-18$	12
4	$46-24$	22

c. Compare your work with several other students and see if your conjectures agree. What might explain your thinking?

Answers may vary. Conjectures might be that the IQR of a sample might have more variability than the distribution of all of the animals because the distribution is skewed and so a small sample is likely to have a lot of variability.

Part 3, Pages 1.5 and 1.6

Focus: Regardless of the shape of the population distribution, a simulated sampling distribution of sample means for a given sample size will be fairly mound shaped and symmetric around the mean of the population from which the samples were drawn.



TI-Nspire Technology Tips

menu accesses page options.

tab cycles through previous samples.

ctrl del resets the page.

On page 1.5, the plot in the upper left corner displays a single sample from the population of interest while the plot in the lower right show the means of the selected samples from the population of interest. The population may be chosen as it is on page 1.3.

Sample allows a single sample, five or ten samples to be drawn then displays a sample chosen from the population on the top left number line and the mean of the sample on the bottom right number line.

The arrows on the keyboard or the screen change the sample size.

Clear clears the distributions but not the population attribute.

Page 1.6 behaves the same as page 1.5 but displays the sample median.

Class Discussion

Have students…	Look for/Listen for…
Select *Sample*.	
• ***Describe the difference between the points on the number line at the top left and the point on the number line at the right.***	Answer: The points on the number line at the top represent a sample of size 10 drawn from the set of land animals' maximum recorded speeds. The number line at the right displays the mean of that sample.
• ***Generate another sample. Use the list to identify the slowest and fastest speeds of the animals in the sample. Which, if any, type of animals' speed surprised you?***	Answers will vary: For one sample, the slowest speed was 2 mph for a sea tortoise and the fastest 50 mph for the hare and blackbuck. I never would have believed that a cow can run at a speed of 25 mph.
• ***Generate samples until you have ten means represented on the number line to the right. Describe the distribution of means in terms of the context including what the points represent.***	Answers may vary. Each point represents the mean maximum speed from a sample of the speeds of different kinds of land animals. For each sample, the mean speed was determined, and the result plotted on the number line. The simulated distribution of the sample mean speeds is stretched out from 19 to 57 mph.

Class Discussion (continued)

• ***Penny claimed the TNS activity was broken because her sample of speeds had the Siberian Tiger at 60 mph, but the number line for the sample means did not have a 60. What would you say to Penny?***	Answer: The number line shows the mean speed of the types of animals in the sample, not all the numbers in each sample.

Return to page 1.3. Reset the page.

• ***Describe the shape of the distribution of maximum speeds of land animals.***	Answer: The distribution of speeds is spread out and slightly symmetric with most of the speeds between 14 mph and 38 mph. One animal (the cheetah, identified in earlier units) has a speed that is about 15 mph faster than all of the others.
• ***Return to page 1.5 and make a conjecture about the shape of the simulated distribution of sample means (on the number line on the right) after taking many, many samples from the population of speeds of land animals: include a description of the general shape of the simulated distribution, the center and the spread, including the smallest and the largest sample means.***	Answers will vary. Some students may think it will be the same shape as the population, mound shaped and symmetric and have a similar center and spread
• ***Generate at least 200 samples. Were you surprised by the simulated distribution of sample means? Compare your simulated distribution to the conjecture you gave for the question above.***	Answers will vary. The simulated distribution of sample means is mound shaped and symmetric around about 26 or 27 mph with a range of 25 mph, from a min of about 15 mph to a max of about 40 mph. Students may not be surprised about the center, but they may be surprised to see the smaller spread.
• ***Reset and generate another simulated distribution of 100 or more sample means. What changed in your description from the question above?***	Answers will vary. The shape, center and spread of the new simulated distribution looks pretty much like the one in d), but the range increased by 3 mph to 35 mph, from 14 to 49 mph.

Student Activity Questions—Activity 3

1. Select *menu> Population> all animals mph*. Then select *Sample*.

a. Describe the difference between the points on the number line at the top left and the points on the number line at the right.

Answer: The number line at the top represents the speeds of types of animals in a sample of size 10 drawn from the set of all animals' maximum recorded speeds. The number line at the right displays the mean of that sample.

Student Activity Questions—Activity 3 (continued)

b. Generate ten samples. Would you expect to get a sample mean of 150 mph in any of your samples? Why or why not?

Answers may vary. It is not possible to get a sample mean of 150 mph in a sample of 10 animal types because even if your sample by chance had all of the fastest speeds, the mean would still not be as large as 150 mph. (about 104 mph).

c. Would you expect to get a sample mean of 15 mph? Why or why not?

Answers may vary. A lot of the speeds are from 0 to 24 mph so it seems reasonable that a sample of 10 speeds might have a mean of 15 mph.

2. Return to page 1.3.

a. Change the population to *all animals mph* and describe the shape of the distribution of maximum speeds of all animals.

Answer: The distribution of speeds for all animals appears skewed to the right with an apparent outlier at 240 mph. A typical speed is around 30 mph and most of the speeds are between 15 and 50 mph.

b. Return to page 1.5. Make a conjecture about the shape of the simulated distribution of sample means (on the number line on the right) after taking many, many samples: include a description of the general shape of the simulated distribution, where you think the central cluster of the sample means will be.

Answers will vary: Some students may think it will be the same shape as the population, skewed to the right with most between 18 and 40 mph.

c. Generate at least 200 samples. Were you surprised by the simulated distribution? Compare your simulated distribution to the conjecture you had for question b.

Answers will vary. Students may be surprised because the distribution should be fairly mound shaped. For example, one simulated distribution was mound shaped and symmetric around 32 mph, with most of the sample means from 25 mph to 50 mph. The smallest sample mean was 15 mph and the largest was 55 mph for a range of 40 mph.

d. Reset and generate another simulated distribution of 100 or more sample means. What changed in your description from question b?

Answer will vary. For one new sample, the shape, center and spread of the new simulated distribution looks pretty much like the one in d), but the range increased by 3 mph to 43 mph, from 15 to 58 mph.

Student Activity Questions—Activity 3 (continued)

3. **Select *menu> Population> all animals mph*.**

   a. **Generate several samples. How do the samples relate to the population? How do the distributions of the speeds in samples of size 10 compare to the distribution of the speeds of all the animals? (Note that to see all of the animal types in a sample, you need to scroll through the list.)**

      Answers will vary. Most samples will be a between 0 and 96, with a few including the outlier and some with a cluster at the left end of the distribution.

   b. **Change the sample size to 20. Generate several samples. How do the distributions of the speeds in samples of size 20 compare to the distribution of the speeds of all the animals?**

      Answers will vary. The samples look a bit more representative of the population. Not all of them, but a lot cluster between 18 and 36 mph with a tail to the right. Some of them include the falcon (about 1 out of every 9 samples).

   c. **Pauleen noted that as the sample size increased, the simulated distribution of sample means seems to get more mound shaped and symmetric. Do you agree?**

      Answers will vary, but students should see that the statement seems to be true.

4. **Kelli was worried about sampling from a population that was skewed, like the maximum speeds or life spans of the animal types. She thought the sampling distributions would be skewed as well. Work with a partner to explain how she might use the files to explore the connection between simulated distributions of sample means and sample size.**

   Answers may vary. Using either the maximum ages or life spans of all animals, she could generate simulated distributions on page 1.5 and observe the change in the shape of the distributions as the sample size increases from 5 to 20. She should see that as the sample size increases, the simulated sampling distributions become more symmetric around the actual mean age or mean life span.

Part 4, Pages 1.3 and 1.5

Focus: The mean and MAD from a single sample of size 20 can give some indication of the population mean. (Note that inferences from smaller samples are less predictable.)

Class Discussion

The following questions introduce the notion that some outcomes are not surprising for a random sample for a given sample size, while others would be far from what might be expected.

Have students…	**Look for/Listen for…**
Go to page 1.3, select all animals maximum speeds, then Mean +/-MAD. Select sample size 20.	

Class Discussion (continued)

- ***Choose Sample and estimate the sample MAD. Repeat this nine more times so you have 10 estimates of sample MADS. The MAD for the speeds of all of the animals is about 17.***

  Answers will vary. One set of ten MADs might be: 14, 20, 20, 10, 22, 19, 18, 18, 9, 22

- ***Sara noted from her samples that, when the sample size was 20, the MAD's were in the interval between 9 and 22 mph. How does her statement compare with your results?***

  Answers will vary. Students may agree that their sample MADs are in that interval.

- ***Generate a lot of samples and note whether the actual population mean was ever outside of the interval determined by the mean+/- MAD.***

  Answer: Students should find that rarely does the population mean fall outside the interval.

Go to page 1.5, generate 200 samples to create a simulated sampling distribution of the mean maximum speeds of all animals for samples of size 20. Use your distribution to decide whether each of the following would be a plausible outcome, given that the mean recorded speed of all types of animals is 31.7 mph.

a. A sample of size 20 with a mean speed of about 25 mph

b. A sample of size 20 with a mean speed of about 40 mph

c. A sample of size 20 with a mean speed of about 15 mph

d. A sample of size 20 with a mean speed of about 33 mph

Answer: Only c would be not be plausible for most of the simulated sampling distributions.

Student Activity Questions—Activity 4

1. **What would you say to each of the following statements? Use examples from the TNS activity to support your thinking.**

   a. **A sample will always have the same shape as the population.**

   b. **A distribution of sample means will have the same shape as the population.**

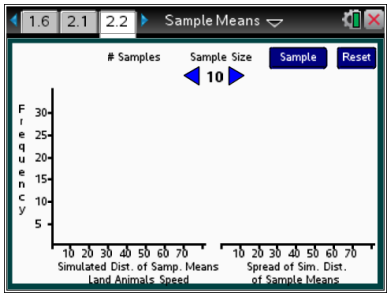
   c. **A distribution of sample means will be fairly mound-shaped and symmetric.**

   d. **A simulated distribution of sample means will be symmetric around the mean of the population.**

   Answers will vary. a) is not true because a sample might not have any of the large numbers or any of the small numbers, like a sample of animal speeds that never had any speeds over 40 mph. b) is not true because we found that the distribution of the speeds of all animals was skewed but the simulated distribution of the sample means was mound shaped and symmetric. c) and d) are both true because that is what happened both with the distribution of the means of samples of the maximum speeds of all animals and the speeds of land animals.

Part 5, Pages 1.3, 1.5, and 2.2

Focus: The larger the sample size, the smaller the variability within the sample.



On page 2.2, use menu options to select a population and the arrow keys a sample size. The spread of the means for each sample is represented by the horizontal bar on the right.

Sample generates 150 random samples of the selected size from the population.

Clear clears the distribution but keeps the population.

Reset returns the original screen. Page 2.3 behaves like 2.2 but with medians.

Class Discussion

Have students…	**Look for/Listen for…**
***Go to page 1.3, select* menu> Population> all animals yrs.**	
• ***Which statement best describes the distribution:*** ***i. skewed right*** ***ii. skewed left*** ***iii. skewed right with two outliers*** ***iv. mound shaped***	Answer: skewed right with two outliers

Class Discussion (continued)

• ***Move to page 1.5, generate at least 200 samples for sample size 10 from the population all animals years. Describe the simulated sampling distribution of sample mean life spans.***	Answers may vary. For example, one simulated distribution of mean life spans was mound shaped but a bit skewed right, with most of the sample mean life spans between 15 years mph and 40 years. The range was 58 years from 10 to 68 years.
• ***Change the sample size to 20. Generate at least 200 samples. How does the new simulated distribution of sample mean life spans compare to the ones from the question above?***	Answers may vary. The simulated distribution of sample mean life spans for samples of size 20 is mound shaped and more symmetric than the simulated sampling distribution of mean life spans for samples of size 10. For the larger sample size most of the mean life spans are from about 22 years to 34 years, with the center around a mean of a 28 year life span. For the smaller sample size most of the sample mean life spans are between 18 and 36 years, a smaller span, with a center around 27 or 28. The range is much smaller, 34 years, for the larger sample size than for the smaller sample size, 58 years.
• ***Make a conjecture about the relationship between sample size and the spread of the distribution.***	Answers will vary. Students may conjecture that just like in Lesson 19, *Sample Proportions*, as the sample size increases, the variability of the distribution of the sample means decreases.

Student Activity Questions—Activity 5

1. Go to page 2.2. Choose *Population> all animals yrs*.

a. Change the sample size to 5. Generate a simulated distribution of the sample mean life spans and describe the spread of the mean life spans represented by the bar on the right.

Answers may vary. One simulated distribution of mean life spans had a spread of 58 years, from 10 to 68 years.

b. Change the sample size to 10 and generate a sample. Repeat for samples of size 15 and samples of size 20. What observations can you make about sample size and the spread of the mean life spans in the simulated sampling distributions?

Answers may vary. One set of simulated distributions of the mean number of years the animals live for different sample sizes had bars representing the spreads for the different sample sizes that decreased from the spread of 58 for sample size 5 to a spread of 22 years, from 15 to 37 years for sample size 20.

Student Activity Questions—Activity 5 (continued)

2. Match the Sentence Starters to the Sentence Enders to make at least four true sentences. Explain your reasoning in each case.

Sentence Starters	Sentence Enders
a. For a simulated distribution of sample means	1. the variability decreases.
b. As the sample size increases,	2. the variability increases.
c. As the sample size decreases	3. the distribution will be skewed.
d. If the population is skewed	4. the variability does not change.
e. If the population is mound shaped and symmetric	5. the distribution of sample means will be mound shaped and symmetric.
	6. the sampling distribution will become more symmetric around the population mean as the sample size increases.

Answers will vary. True statements come from matching a to 5; b to 1; c to 2; d to 6; e to 5.

Deeper Dive — Page 1.3

- ***Find the mean maximum speed for all animals and the mean maximum speed for land animals. What might explain the difference?***

  Answer: The mean maximum speed for all animals is about 32 mph, while for land animals it is about 27 mph. The mean speed is higher for all animals because many of the fastest animals are birds (240 mph for the falcon, 120 mph for the swift, 100 mph for Eurasian Hobby, 105 mph for the Needle tail, 100 mph for pigeon are all birds) and they would not be included in land animals. So the mean speed for all animals will be larger than the mean speed of land animals.

- ***Answer the question above for land animals.***

  Answer: The mean life span for all animals is about 26 years and for land animals about 23 years. The difference is that the animals that live the longest are the sea mammals (sea tortoise 255 years, bowhead whales 200 years, blue whales 110 years, and the orca killer whale 100 years). So the mean life span for all animals is larger because the set has all of the longer living animals.

Deeper Dive — Pages 1.3 and 1.5

Select *all animals mph*.

- ***For a sample of size 20, use page 1.3 for all animals mph to estimate how often the falcon is in the sample- that is, estimate the probability a sample will include 240 mph.***

  Answers will vary. For 60 simulations, about 1 out of every 15 samples

- ***Do you think the answer will be larger, smaller or the same for sample size 10? Explain your reasoning. Then change the sample size to check your thinking.***

  Answers will vary. Some students may note that there are fewer opportunities to include the falcon's 240 mph speed. For 60 simulations, about 1 out of every 30 samples contained the falcon. (Note that this problem is similar to the true false test from Lesson 15: *Probability and Simulation* where the number of questions on the test changed from 10 to 20.)

- ***Will your answers change if you look at the maximum life spans of all the animals and estimate how often the falcon is included?***

  Answer: No. Any particular type animal will occur at the same rate because the number of animals will be the same even though the data are life spans in years and not mph.

- ***Will the probability of getting a speed equivalent to the elephant's maximum speed (25 mph) in the sample be the same as, less than or more than the probability of getting a speed equivalent to the falcon? Explain your thinking.***

  Answer: Using both page 1.3 and 1.5, students should see that the probability of getting a speed of 25 mph will be more than the probability of getting the falcon's speed because lots of animals have a speed of 25 mph (ox, sheet, polar bear) and only one has a speed of 240 mph.

Deeper Dive — Page 1.5

Think about each of the following statements in the context of the simulated sampling distributions of the means of the maximum speeds of all the animals (mean of all animal maximum speeds was 31.7 mph, MAD of 17.5) and the longest recorded life span of all animals (mean of 26.4 years, MAD of about 12). Adapt the TNS activity to support your thinking in each new context.

- ***Sally surveyed a random sample of 20 seventh graders and found the students in the sample spent a mean of $38 per month on clothes with a MAD of $20. She concluded that mean amount all the seventh graders spent on clothes was $38 a month.***

  Answers will vary. Note 38 is a plausible outcome for a sample from a population where the mean is actually 26.4 (from mean speed of land animals) and from a population where the mean was actually 31.7 (from the mean speed of all animals). Given that a result of 38 could have come from either of these populations, Sally might not be making the right claim.

Deeper Dive — Page 1.5 (continued)

- ***Samee took a similar survey of a random sample of 20 students from seventh grade in another school and found that the mean amount of money they spent on clothes per month was $35 with a MAD of $15. He concluded that the typical amount of money students in the seventh grade spent on clothes was likely to be somewhere between $20 and $50 a month.***

Answers may vary. Students should note that a sample result of 35 was a likely outcome for a sample from both a population with a mean of 31.7 and a population with a mean of 26.4 and the interval would be plausible for either population.

Deeper Dive — Pages 1.6 and 2.3

Pages 1.6 and 2.3 allow you to explore simulated sampling distributions of sample medians. Use the pages to answer the following questions.

a. In general, describe simulated distributions for sample medians.

b. What features of simulated sampling distributions for sample means seem to be the same as the simulated sampling distributions for sample medians? Explain your thinking.

c. Describe how sample size might affect simulated sampling distributions of sample medians.

Answers will vary. a. Mound shaped and fairly symmetric around 30 mph for maximum speeds of all animals, but for land animals maximum speeds, the peak is not in the center but at bit to the right at about 31 or 32 mph. The maximum life span for all animals is a skewed right for small samples and always has the most common value right around 20. b. The distributions of sample medians will be mound shaped, but do not seem to be too symmetric. c. An increase in the sample size should cause the spread to decrease.

Sample Assessment Items

After completing the lesson, students should be able to answer the following types of questions. If students understand the concepts involved in the lesson, they should be able to answer the following questions without using the TNS activity.

1. Identify each statement as true or false:

   a. The mean of a distribution of sample means will approach the population mean as more and more samples are taken.

   Answer: True

   b. Distributions of sample means will have the same shape as the population.

   Answer: False

   c. Different samples selected from the same population will have the same sample mean.

   Answer: False

   d. The larger the sample size, the less the variation in the distribution of sample means.

   Answer: True

2. Malika and Adrian prepared containers of potato salad at a deli. Each container was supposed to have a mass of one pound. The manager selected a random sample of containers prepared by each employee to check the mass of each container. The results are shown in the table below.

Malika's Containers (pounds)	Adrian's Containers (pounds)
1.10	1.30
1.08	1.21
1.05	0.79
0.95	0.90
0.98	0.88

   Which inference is best supported by these data?

   a. Malika will produce more containers with a mass of exactly one pound than Adrian will.

   b. Adrian will produce more containers with a mass of exactly one pound than Malika will.

   c. Most of Malika's containers will have a mass closer to one pound than most of Adrian's containers.

   d. Most of Adrian's containers will have a mass closer to one pound than most of Malika's containers.

   Engage New York; New York State Testing Program Grade 7 Common Core Mathematics Test release items 2015

   Answer: c. Most of Malika's containers will have a mass closer to one pound than most of Adrian's containers.

3. Identify the following as a sample, a population, or a distribution of a sample statistic.

   a. all the grade point averages of students in the seventh grade

      Answer: the distribution of the population of grade point averages for seventh graders

   b. the grade point averages of the twenty students in fifth hour seventh grade math

      Answer: the distribution a sample of grade point averages

   c. the mean grade point averages for different samples of 15 seventh grade students

      Answer: the distribution of a sample statistic

4. Identify the statements as true or false.

   a. The larger the sample size the more likely that a sample mean will be farther from the population mean.

      Answer: False

   b. The larger the sample size the more likely a sample mean will approximate the population mean.

      Answer: True

   c. All simulated distributions of sample means, no matter what sample size, will be mound shaped and symmetric.

      Answer: False

Student Activity Solutions

In these activities you will describe and compare the mean and MAD of different samples . After completing the activities, discuss and/or present your findings to the rest of the class.

Activity 1 [Page 1.3]

1. Select **menu> Measures> Mean +/-MAD**.

   a. The mean maximum recorded speed of all of the land animals is about 27 mph and the MAD is about 12 mph. How do the mean and mean absolute deviations of the sample and of all the land animals compare?

   Answers will vary. For some samples, the mean maximum recorded speed and the MAD of the sample will be fairly close to the mean maximum recorded speed and MAD of all the land animals.

   b. Generate another sample and estimate the difference between the mean speed of the animals in the sample and the mean speed of all the land animals. Estimate the difference between the MAD of the sample and the MAD of the population.

   Answers will vary. In one sample, the mean was about 24 mph and the MAD about 6 mph so the difference would be 3 mph for the means and 6 mph for the MADs.

   c. Generate four more samples and answer question b for each sample. Then compare your answers to the answers someone else found. Overall, how would you describe the difference between the mean speed and the MAD for the sample and for all the land animals?

   Answers will vary.

	Estimated sample mean speed	Difference: pop (27)-sample	Estimated sample MAD	Difference: Pop (12)-sample
1	33	-6	$44-33=11$	1
2	30	-3	$44-30=13$	-1
3	28	-1	$44-28=16$	-4
4	27	0	$42-27=15$	-3

Answers will vary. The difference between the mean speed of the sample of land animals and the mean speed of all the land animals in the file is typically less than 7 or 8 mph. The difference between the MADs is typically small, usually less than 3 or 4.

2. Use the TNS activity to determine which of the following statements seem to be true.

   a. Over a lot of samples, the mean maximum speeds of the samples of land animals were close to the mean of the maximum speeds of the land animals.

      Answers may vary. This seems to be a true statement, given our work in the previous question, if close is taken to be less than three or four mph difference.

   b. Many of the sample distributions were relatively similar to the population distribution in terms of center and spread.

      Answers may vary. This seems to be true from the samples we did.

   c. When the sample contained the cheetah at 75 mph, the mean maximum speed of the sample of land animals was larger than the mean speed of all of the land animals.

      Answers may vary, but this seems to be true because every sample that had the 75 mph in the sample had a higher mean. This might be because there are only 10 speeds in the sample, and 75 is 15 mph larger than any of the other speeds.

   d. The mean speed of the sample of land animals was always larger than the mean maximum speed of all of the land animals in the TNS activity.

      Answers will vary. This is not supported by the samples generated in the activity.

Activity 2 [Page 1.3]

1. Select **menu> Measures> IQR = UQ-LQ**.

   a. Select **Sample**. How do the IQRs of the sample and the population compare?

      Answers will vary. Some students might point out that the medians are about the same, but the IQR for the sample speeds is larger. Others may have a much smaller IQR.

   b. Select **Sample** again and record the IQRs. Do this four times and record your results in the table. Make a conjecture about the difference between the IQR of the sample of land animals and the IQR of all the animals.

      Answers will vary. Examples are below. A conjecture might be that the IQRs seem to be typically larger for the animals in the samples than for all of the animal types.

Sample	Sample IQR	Difference: All: IQR about $30-20=10$
1	$40-22$	18
2	$50-28$	22
3	$30-18$	12
4	$46-24$	22

c. Compare your work with several other students and see if your conjectures agree. What might explain your thinking?

Answers may vary. Conjectures might be that the IQR of a sample might have more variability than the distribution of all of the animals because the distribution is skewed and so a small sample is likely to have a lot of variability.

Activity 3 [Pages 1.3 and 1.5]

1. Select **menu> Population> all animals mph**. Then select **Sample**.

a. Describe the difference between the points on the number line at the top left and the points on the number line at the right.

Answer: The number line at the top represents the speeds of types of animals in a sample of size 10 drawn from the set of all animals' maximum recorded speeds. The number line at the right displays the mean of that sample.

b. Generate ten samples. Would you expect to get a sample mean of 150 mph in any of your samples? Why or why not?

Answers may vary. It is not possible to get a sample mean of 150 mph in a sample of 10 animal types because even if your sample by chance had all of the fastest speeds, the mean would still not be as large as 150 mph. (about 104 mph).

c. Would you expect to get a sample mean of 15 mph? Why or why not?

Answers may vary. A lot of the speeds are from 0 to 24 mph so it seems reasonable that a sample of 10 speeds might have a mean of 15 mph.

2. Return to page 1.3.

a. Change the population to **all animals mph** and describe the shape of the distribution of maximum speeds of all animals.

Answer: The distribution of speeds for all animals appears skewed to the right with an apparent outlier at 240 mph. A typical speed is around 30 mph and most of the speeds are between 15 and 50 mph.

b. Return to page 1.5. Make a conjecture about the shape of the simulated distribution of sample means (on the number line on the right) after taking many, many samples: include a description of the general shape of the simulated distribution, where you think the central cluster of the sample means will be.

Answers will vary: Some students may think it will be the same shape as the population, skewed to the right with most between 18 and 40 mph.

c. Generate at least 200 samples. Were you surprised by the simulated distribution? Compare your simulated distribution to the conjecture you had for question b.

Answers will vary. Students may be surprised because the distribution should be fairly mound shaped. For example, one simulated distribution was mound shaped and symmetric around 32 mph, with most of the sample means from 25 mph to 50 mph. The smallest sample mean was 15 mph and the largest was 55 mph for a range of 40 mph.

d. Reset and generate another simulated distribution of 100 or more sample means. What changed in your description from question b?

Answer will vary. For one new sample, the shape, center and spread of the new simulated distribution looks pretty much like the one in d), but the range increased by 3 mph to 43 mph, from 15 to 58 mph.

3. Select **menu> Population> all animals mph**.

a. Generate several samples. How do the samples relate to the population? How do the distributions of the speeds in samples of size 10 compare to the distribution of the speeds of all the animals? (Note that to see all of the animal types in a sample, you need to scroll through the list.)

Answers will vary. Most samples will be a between 0 and 96, with a few including the outlier and some with a cluster at the left end of the distribution.

b. Change the sample size to 20. Generate several samples. How do the distributions of the speeds in samples of size 20 compare to the distribution of the speeds of all the animals?

Answers will vary. The samples look a bit more representative of the population. Not all of them, but a lot cluster between 18 and 36 mph with a tail to the right. Some of them include the falcon (about 1 out of every 9 samples).

c. Pauleen noted that as the sample size increased, the simulated distribution of sample means seems to get more mound shaped and symmetric. Do you agree?

Answers will vary, but students should see that the statement seems to be true.

4. Kelli was worried about sampling from a population that was skewed, like the maximum speeds or life spans of the animal types. She thought the sampling distributions would be skewed as well. Work with a partner to explain how she might use the files to explore the connection between simulated distributions of sample means and sample size.

Answers may vary. Using either the maximum ages or life spans of all animals, she could generate simulated distributions on page 1.5 and observe the change in the shape of the distributions as the sample size increases from 5 to 20. She should see that as the sample size increases, the simulated sampling distributions become more symmetric around the actual mean age or mean life span.

Activity 4 [Page 1.5]

1. What would you say to each of the following statements? Use examples from the TNS activity to support your thinking.

   a. A sample will always have the same shape as the population.

   b. A distribution of sample means will have the same shape as the population.

   c. A distribution of sample means will be fairly mound-shaped and symmetric.

   d. A simulated distribution of sample means will be symmetric around the mean of the population.

   Answers will vary. a) is not true because a sample might not have any of the large numbers or any of the small numbers, like a sample of animal speeds that never had any speeds over 40 mph. b) is not true because we found that the distribution of the speeds of all animals was skewed but the simulated distribution of the sample means was mound shaped and symmetric. c) and d) are both true because that is what happened both with the distribution of the means of samples of the maximum speeds of all animals and the speeds of land animals.

Activity 5 [Page 2.2]

1. Go to page 2.2. Choose **Population> all animals yrs**.

   a. Change the sample size to 5. Generate a simulated distribution of the sample mean life spans and describe the spread of the mean life spans represented by the bar on the right.

      Answers may vary. One simulated distribution of mean life spans had a spread of 58 years, from 10 to 68 years.

   b. Change the sample size to 10 and generate a sample. Repeat for samples of size 15 and samples of size 20. What observations can you make about sample size and the spread of the mean life spans in the simulated sampling distributions?

      Answers may vary. One set of simulated distributions of the mean number of years the animals live for different sample sizes had bars representing the spreads for the different sample sizes that decreased from the spread of 58 for sample size 5 to a spread of 22 years, from 15 to 37 years for sample size 20.

2. Match the Sentence Starters to the Sentence Enders to make at least four true sentences. Explain your reasoning in each case.

Sentence Starters	Sentence Enders
a. For a simulated distribution of sample means	1. the variability decreases.
b. As the sample size increases,	2. the variability increases.
c. As the sample size decreases	3. the distribution will be skewed.
d. If the population is skewed	4. the variability does not change.
e. If the population is mound shaped and symmetric	5. the distribution of sample means will be mound shaped and symmetric.
	6. the sampling distribution will become more symmetric around the population mean as the sample size increases.

Answers will vary. True statements come from matching a to 5; b to 1; c to 2; d to 6; e to 5