Advanced Math 7 Period _____ Name________________________
Statistics Date_________________________

6.6 Table of Contents

7.5 Table of Contents

OVERVIEW

Students focus on using random sampling to draw informal inferences about a population. They investigate sampling from a population. They learn to estimate a population mean using numerical data from a random sample. They also learn how to estimate a population proportion using categorical data from a random sample. Students learn to compare two populations with similar variability. They learn to consider sampling variability when deciding if there is evidence that the means of two populations are actually different.

CHALLENGE 34
Can you develop a PROBABILITY model/experiment?

CHALLENGE 65
Can you find the measures of center of data: MEAN, MEDIAN, and MODE?

CHALLENGE 83
I can compare EXPERIMENTAL and THEORETICAL PROBABILITY.

CHALLENGE 96
I can express data in a HISTOGRAM, LINE PLOT/DOT PLOT, BOX PLOT, FREQUENCY TABLE?

CHALLENGE 97
Can you calculate measures of variation of data: RANGE, IQR, MAD?

Focus Standards for Mathematical Practice

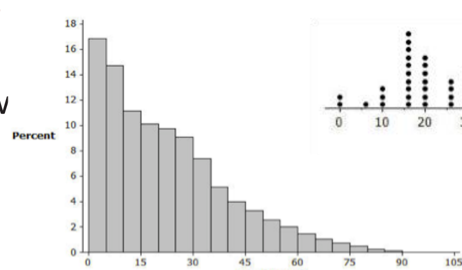
MP.2 **Reason abstractly and quantitatively.**
MP.3 **Construct viable arguments and critique the reasoning of others.**
MP.4 **Model with mathematics.**
MP.5 **Use appropriate tools strategically.**

Terminology

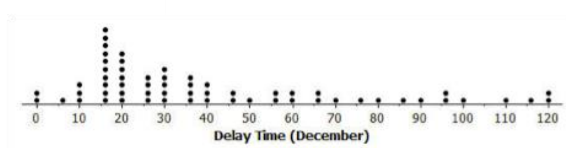
- **Statistical Question** - A question that anticipates variability in the data that would be collected in order to answer the question.
- **Median** - A measure of center appropriate for skewed data distributions. It is the middle value when the data are ordered from smallest to largest if there are an odd number of observations and half way between the middle two observations if the number of observations is even.
- **Mean** - A measure of center appropriate for data distributions that are approximately symmetric. It is the average of the values in the data set. Two common interpretations of the mean are as a “fair share” and as the balance point of the data distribution.
- **Dot Plot** - A plot of numerical data along a number line.
- **Histogram** - A graphical representation of a numerical data set that has been grouped into intervals. Each interval is represented by a bar drawn above that interval that has a height corresponding to the number of observations in that interval.
- **Box Plot** - A graph of five numerical summary measures: the minimum, lower quartile, median, upper quartile, and the maximum. It conveys information about center and variability in a data set.
- **Variability** - when the observations in the data set are not all the same.
- **Deviations from the Mean** - The differences calculated by subtracting the mean from the observations in a data set.
- **Mean Absolute Deviation (MAD)** - A measure of variability appropriate for data distributions that are approximately symmetric. It is the average of the absolute value of the deviations from the mean.
- **Range –** A measure of spread of the data. The greatest value minus the least value.
- **Interquartile Range (IQR)** - A measure of variability appropriate for data distributions that are skewed. It is the difference between the upper quartile and the lower quartile of a data set and describes how spread out the middle 50% of the data are.
- **Random sample** - A sample selected in a way that gives every different possible sample of the same size an equal chance of being selected.
- **Inference** - Using data from a sample to draw conclusions about a population.
- **Sampling Variability** – variation from sample to sample in the values of the sample statistics (mean, median, range, …)
- **Absolute Deviation** - how far away that data point is from the mean

Suggested Tools and Representations

- Dot plots (see example below)
- Histograms (see example below



Histogram



Dot Plot

6.6.10 Lesson Date_________________________

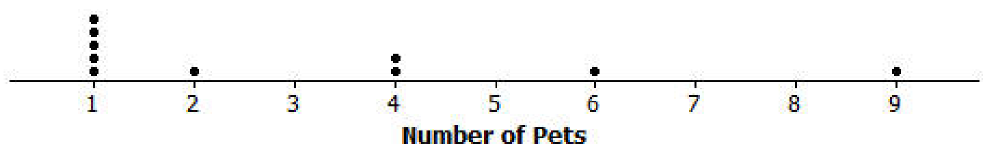
Describing Distributions Using the Mean and MAD

Student Objectives

- I can calculate the mean and MAD for a data distribution.
- I can use the mean and MAD to describe a data distribution in terms of center and variability.

Example 1: Describing Distributions

Using the mean and MAD with a dot plot allows you to describe the center, spread, and shape of a data distribution. For example, suppose that data on the number of pets for ten students is shown in the dot plot below.



There are several ways to describe the data distribution. The mean number of pets each student has is three, which is a measure of a typical value. There is variability in the number of pets the students have, which is an average of 2.2 pets from the mean (the MAD). The shape of the distribution is heavy on the left and it thins out to the right (call skewed ___________________).

Exercises 1–4

Absolute Deviation = $|Data\ point\ -\ Mean|$

1. Suppose that the weights of seven middle-school students' backpacks are given below.
   a. Draw a dot plot for these data and calculate the mean and MAD.

Student	Alan	Beth	Char	Damon	Elisha	Fred	Georgia
Weight (lbs.)	18	18	18	18	18	18	18
Absolute Deviations							

Mean = Mean Absolute Deviation =

b. Describe this distribution of weights of backpacks by discussing the

Center:

Spread:

Shape:

2. Suppose that the weight of Elisha's backpack is 17 pounds, rather than 18.

a. Draw a dot plot for the new distribution.

b. Without doing any calculation, how is the mean affected by the lighter weight? Would the new mean be the same, smaller, or larger?

c. Without doing any calculation, how is the MAD affected by the lighter weight? Would the new MAD be the same, smaller, or larger?

3. Suppose that in addition to Elisha's backpack weight having changed from 18 to 17 lb., Fred's backpack weight is changed from 18 to 19 lb.

a. Draw a dot plot for the new distribution.

b. Without doing any calculation, what would be the value of the new mean compared to the original mean?

c. Without doing any calculation, would the MAD for the new distribution be the same, smaller, or larger than the original MAD?

d. Without doing any calculation, how would the MAD for the new distribution compare to the one in Exercise 2?

4. Suppose that seven second-graders' backpack weights were:

Student	Alice	Bob	Carol	Damon	Ed	Felipe	Gale
Weight (lbs.)	5	5	5	5	5	5	5

a. How is the distribution of backpack weights for the second-graders similar to the original distribution for sixth-graders given in Exercise 1? How are the distributions different?

Lesson Summary

A data distribution can be described in terms of its center, spread, and shape.

- The center can be measured by the mean.
- The spread can be measured by the mean absolute deviation (MAD).
- A dot plot shows the shape of the distribution.

6.6.11 Lesson Date________________________

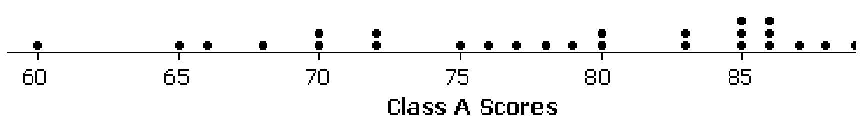
Describing Distributions Using the Mean and MAD

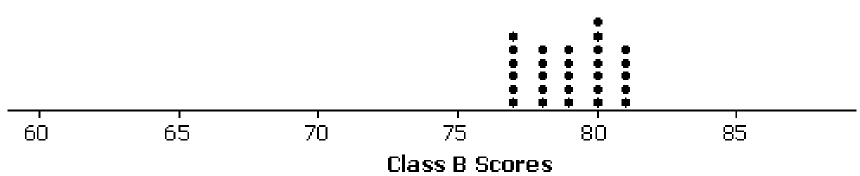
Student Objectives

- I can use the mean and MAD to describe a data distribution in terms of center and variability.
- I can use the mean and MAD to describe similarities and differences between two distributions.

Example 1: Comparing Distributions with the Same Mean

In Lesson 10, a data distribution was characterized mainly by its center (mean) and variability (MAD). How these measures help us make a decision often depends on the context of the situation. For example, suppose that two classes of students took the same test and their grades (based on 100 points) are shown in the following dot plots. The mean score for each distribution is 79 points.





Exercises 1–3

1. Looking at the dot plots, which class has the greater MAD? Explain without actually calculating the MAD.

2. If Liz had one of the highest scores in her class, in which class would she rather be? Explain your reasoning.

3. If Logan scored below the average of the class, in which class would he rather be? Explain your reasoning.

If two data distributions have **different means**, how does a measure of variability play a part in making decisions?

Exercises 4–9

Suppose that you wanted to answer the following question: Are field crickets better predictors of atmospheric temperature than katydids are? Both species of insect make chirping sounds by rubbing their front wings together.

The following data are the number of chirps (per minute) for 10 insects each. All the data were taken on the same evening at the same time.

Insect	A	B	C	D	E	F	G	H	I	J
Crickets	35	32	35	37	34	34	38	35	36	34
Katydids	66	62	61	64	63	62	68	67	66	61

4. Draw dot plots for these two data distributions using the same scale, going from 30 to 70. Visually, what conclusions can you draw from the dot plots?

5. The mean number of chirps for the crickets is 35 and the mean for the katydids is 64. Calculate the MAD for each distribution.

Crickets	35	32	35	37	34	34	38	35	36	34
Absolute Deviation										

Katydids	66	62	61	64	63	62	68	67	66	61
Absolute Deviation										

6. The outside temperature T can be predicted by counting the number of chirps made by these insects.

   a. For crickets, T is found by adding 40 to its mean number of chirps per minute. What value of T is being predicted by the crickets?

   b. For katydids, T is found by adding 11 to its mean number of chirps per minute. What value of T is being predicted by the katydids?

   c. The temperature was 75 degrees when these data were recorded, so using the mean from each data set gave an **accurate** prediction of temperature. If you were going to use the number of chirps from a single cricket or a single katydid to predict the temperature, would you use a cricket or a katydid? Explain how variability in the distributions of number of chirps played a role in your decision.

Lesson Summary

This lesson focused on comparing two data distributions based on center and variability. It is important to consider the context when comparing distributions. In decision-making, drawing dot plots and calculating means and MADs can help you make informed decisions.

6.6.12 Lesson Date________________________

Describing a Distribution Using the Median and Range

Student Objectives

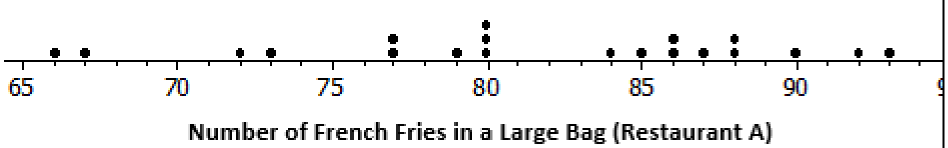
- Given a data set, I can calculate the median and the range of the data.
- I can estimate the percent of values above and below the median value.

How do we summarize a data distribution? What provides us with a good description of the data? The following exercises help us to understand how a numerical summary answers these questions.

The_________________ is the middle value in a data set or the mean of the middle two values when the data are ordered from _____________________________.

Example 1: The Median—A Typical Number

Suppose a chain restaurant (Restaurant A) advertises that a typical number of french-fries in a large bag is 82. The graph shows the number of french fries in selected samples of large bags from Restaurant A.



The ___________________ is the largest value minus the least value.

Exercises 1–4

1. You just bought a large bag of fries from the restaurant. Do you think you have 82 french fries? Why or why not?

2. How many bags were in the sample? ________

3. What is the range of numbers of fries in a bag?_____________________________

4. Which of the following statements would seem to be true given the data? Discuss your reasoning.

   a. Half of the bags had more than 82 fries in them.
   b. Half of the bags had fewer than 82 fries in them.
   c. More than half of the bags had more than 82 fries in them.
   d. More than half of the bags had fewer than 82 fries in them.
   e. If you got a random bag of fries, you could get as many as 93 fries.

Exercises 5–6

5. Use the information from the dot plot in Example 1. The median number of fries was 82.

a. What percent of the bags have more fries than the median? Less than the median?

b. Suppose the bag with 93 fries was miscounted and there were only 85 fries. Would the median change? Why or why not?

c. Suppose the bag with 93 fries really only had 80 fries. Would the median change? Why or why not?

6. The owner of the chain decided to check the number of French-fries at another restaurant in the chain. Here is the data for Restaurant B:
$84, 83, 84, 82, 83, 78, 85, 83, 83, 83, 85, 79, 80, 81, 81, 79, 79, 79, 80$.

a. How many bags of fries were counted?

b. Sallee claims the median is 83 as she sees that 83 is the middle number in the data set listed above. She thinks half of the bags had fewer than 83fries. Do you think she would change her mind if the data were plotted in a dot plot? Why or why not?

c. What is the median? the range?

Exercises 7–8: Finding Medians from Frequency Tables

7. A third restaurant (Restaurant C) tallied a sample of bags of french-fries and found the results below.

Number of fries	Frequency
75	||
76	
77	|||
78	|||
79	|||||
80	||||
81	|
82	|
83	
84	|||
85	|||
86	|

a. How many bags of fries did they count?

b. What is the median number of fries for the sample of bags from this restaurant?

c. What is the range?

8. Which of the three restaurants seems most likely to really have 82 fries in a typical bag? Explain your thinking.

Lesson Summary

In this lesson, you learned about a summary measure for a set of data called the median. To find a median you first have to order the data. The **median** is the midpoint of a set of ordered data; it separates the data into two parts with the same number of values below as above that point. For an even number of data values, you find the average of the two middle numbers; for an odd number of data values, you use the middle value. It is important to note that the median might not be a data value and that the median has nothing to do with a measure of distance. Medians are sometimes called a measure of the center of a frequency distribution but do not have to be the middle of the spread or range (maximum-minimum) of the data.

6.6.13 Lesson Date__________________________

Describing Variability using the Interquartile Range (IQR)

Student Objectives

- I can calculate the median and upper and lower quartiles of the data.
- I can describe the variability in the data by calculating the interquartile range.

Classwork

When the mean is not a very good measure of center (there are **outliers** that skew the **data**), the median can be used to describe the typical value of data. We need a good way to indicate how the data vary when we use a **median** as our typical value. The median of the top half of the data is called the upper quartile, _________; the median of the bottom half of the data is called the lower quartile, _______. The interquartile range, _______ is the measure of spread between the ______ and the _______.

Example 1: Finding the IQR

Consider the data:

Creating an IQR:

I. Order the data: The data is already ordered.

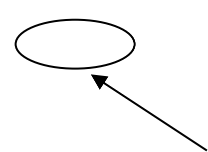
II. Find the median: There are $\mathbf{26}$ data points so the mean of the 13th and the 14th tallies from the smallest or from the largest will be the median.

III. Find the lower quartile and upper quartile: The lower half of the data has ________data points, so the lower quartile (LQ or Q1) will be the ________data point from the minimum value, or _______.

The upper quartile (UQ or Q3) will be the _______ data point from the maximum value, or ________.

IV. Find the difference between UQ and LQ: The $\boldsymbol{IQR} =$.

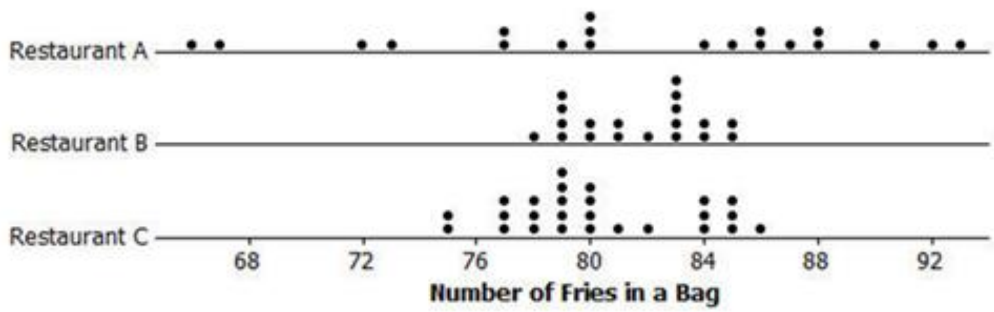
Number of fries	Frequency
75	||
76	
77	|||
78	|||
79	|||||
80	||||
81	|
82	|
83	
84	|||
85	|||
86	|



Exercises 1–2

1. In Exercise 9 of Lesson 12, you found the median of the top half and the median of the bottom half of the counts for Restaurant C. The numbers for each restaurant are: Restaurant A: 87.5 and 77; Restaurant B: 83 and 79; Restaurant C: 84 and 78.

   a. Mark the quartiles for each restaurant on the graphs below.



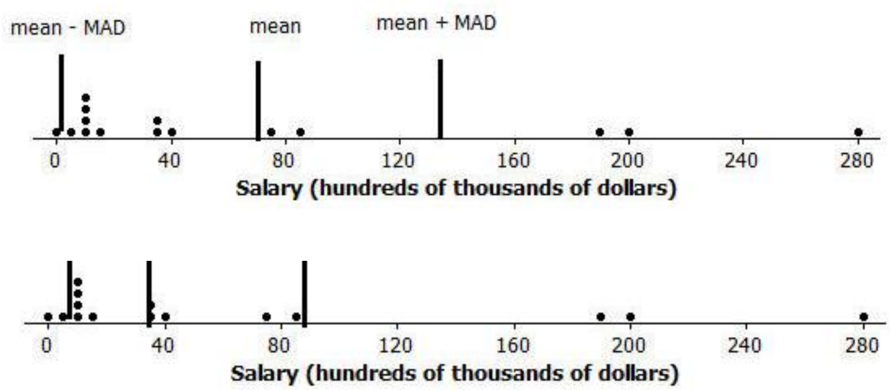
   b. The difference between the medians of the two halves is called the **interquartile range** or **IQR**.

   What is the IQR for each of the three restaurants? Formula: **IQR =**___________

   a. Which of the restaurants had the smallest IQR, and what does that tell you?

   b. About what fraction of the counts would be between the upper and lower quartiles? Explain your thinking.

2. When should you use the IQR? The data for the 2012 salaries for the Lakers basketball team are given in the two plots below (see problem 5 in the Problem Set from Lesson 12).



a. The data are given in hundreds of thousands of dollars. What would a salary of 40 hundred thousand dollars be?

b. The vertical lines on the top plot show the mean and the mean ± the MAD. The bottom plot shows the median and the IQR. Which interval is a better picture of the typical salaries?

Lesson Summary

One of our goals in statistics is to summarize a whole set of data in a short concise way. We do this by thinking about some measure of what is typical and how the data are spread relative to what is typical.

In earlier lessons, you learned about the **MAD** as a way to measure the spread of data about the **mean**. In this lesson, you learned about the **IQR** as a way to measure the spread of data around the **median**.

To find the **IQR**, you order the data, find the **median** of the data, and then find the **median of the lower half** of the data (the lower quartile) and the **median of the upper half** of the data (the upper quartile). The IQR is the difference between the upper quartile and the lower quartile, which is the length of the interval that includes the middle half of the data, because the median and the two quartiles divide the data into four sections, with about $\frac{1}{4}$ of the data in each section. Two of the sections are between the quartiles, so the interval between the quartiles would contain about 50% of the data.

6.6.14 Summarizing a Distribution Using a Box Plot Date______________________

Student Objective

- I can construct a box plot from a given set of data.

A box plot is a graph that is used to summarize a data distribution.

Example 1: Making a Box Plot

To make a box plot

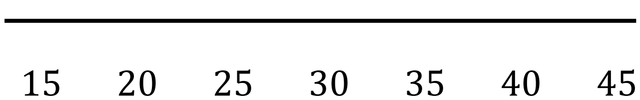
- Find the median of all of the data
- Find Q1, the median of the bottom half of the data, and Q3, the median of the top half of the data.
- Draw a box that goes from Q1 to Q3.
- Draw a line inside the box to show the median.
- Draw a line segment connecting the minimum value to the box and one that connects the maximum value to the box.

Now use the given number line to make a box plot of the data below.

$20, 21, 25, 31, 35, 38, 40, 42, 44$

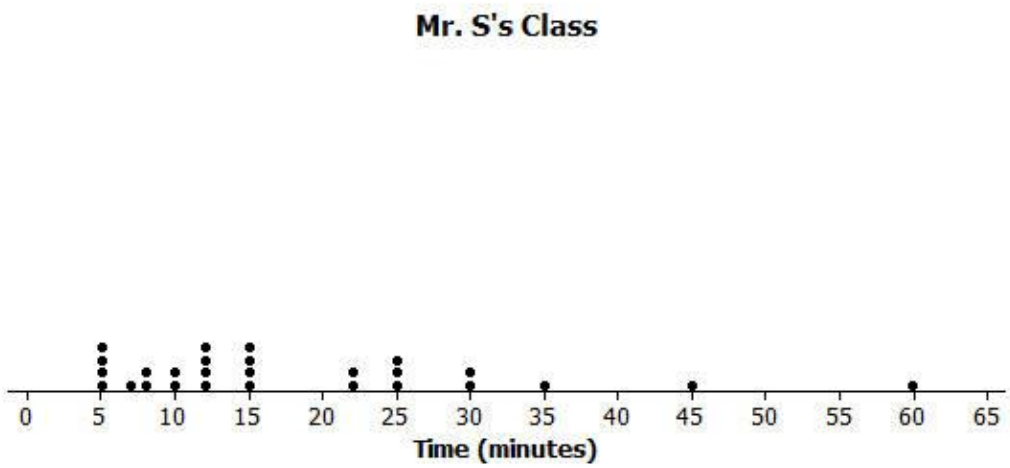
The 5-number summary is as follows:

Min = 20
Q1 = 23
Median = 35
Q3 = 41
Max = 44



Exercises 1-2

Here is a dot plot of the estimates of the times it took students in Mr. S's class to get to school one morning.



1. Make a box plot of the times it took students in Mr. S's class to get to school one morning.

2. What don't the pictures tell you about the length of time it takes the students to get to school?

Lesson Summary

The focus of this lesson is moving from a plot that shows all of the data values (dot plot) to one that summarizes the data with five points (box plot).

You learned how to make a box plot by doing the following:

- Finding the median of all of the data
- Finding LQ (Q1), the median of the bottom half of the data, and UQ (Q3), the median of the top half of the data.
- Drawing a box that goes from LQ (Q1) to UQ (Q3), the two middle sections.
- Drawing a line segment connecting the minimum value to the box and one that connects the maximum value to the box.

You also learned important characteristics of box plots:

- $\frac{1}{4}$ of the data are in each of the sections of the plot.
- The length of the interval for a section does not indicate either how the data are grouped in that interval or how many values are in the interval.

6.6.15 More Practice with Box Plots Date__________________

Student Objectives

- Given a box plot, I can summarize the data by the 5-number summary (Min, Q1, Median, Q3, Max) and the interquartile range.
- I can construct a box plot from a 5-number summary.

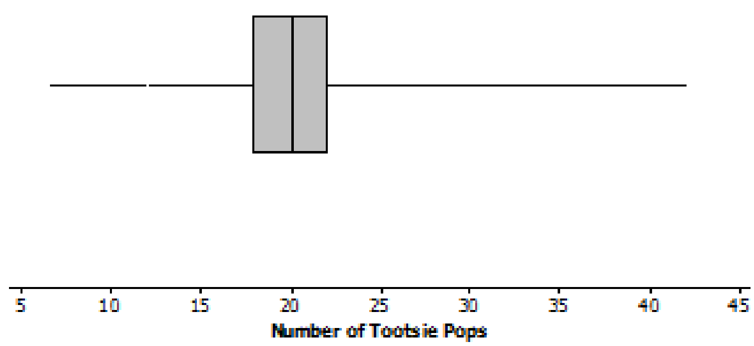
You reach into a jar of Tootsie Pops. How many Tootsie Pops do you think you could hold in one hand? Do you think the number you could hold is greater than or less than what other students can hold? Is the number you could hold a typical number of Tootsie Pops? This lesson examines these questions.

Classwork

Example 1: Tootsie Pops

As you learned earlier, the five numbers that you need to make a box plot are the minimum, the lower quartile, the median, the upper quartile, and the maximum. These numbers are called the 5-number summary of the data.

Ninety-four people were asked to grab as many Tootsie Pops as they could hold. Here is a box plot for these data. Are you surprised?



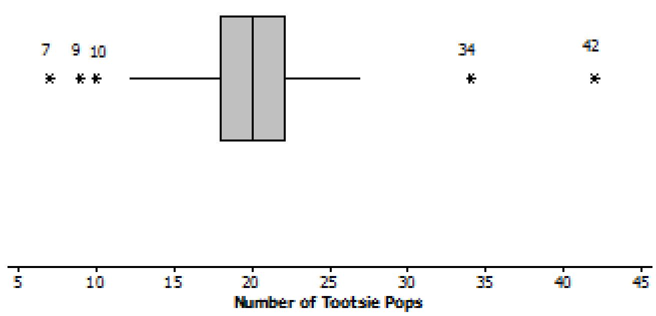
Exercises 1–4

1. What might explain the variability in how many Tootsie Pops those 94 people were able to hold?

2. Estimate the values in the 5-number summary from the box plot.

3. Here is Jayne's description of what she sees in the plot. Do you agree or disagree with her description?
   * One person could hold as many as 42 Tootsie Pops.____________
   - The number of Tootsie Pops people could hold was really different and spread about equally from 7 to 42._____________
   - About one half of the people could hold more than 20 Tootsie Pops.____________

4. Here is a different plot of the same data on the number of Tootsie Pops 94 people could hold. The 5 values are separated and labeled because they are _______________________ (not typical of the rest of the data or lie outside the common results).



Does knowing these data values change anything about your responses to Exercise 3?

Exercises 5–8: Maximum Speeds

The maximum speeds of selected birds and land animals are given in the tables below.

Bird	Speed (mph)
Peregrine falcon	242
Swift bird	120
Spine-tailed swift	106
White-throated needletail	105
Eurasian hobby	100
Pigeon	100
Frigate bird	95
Spur-winged goose	88
Red-breasted merganser	80
Canvasback duck	72
Anna's Hummingbird	61.06
Ostrich	60

Land animal	Speed (mph)
Cheetah	75
Free-tailed bat (in flight)	60
Pronghorn antelope	55
Lion	50
Wildebeest	50
Kangaroo	45
African wild dog	44
Jackrabbit	44
Horse	44
Thomson's gazelle	43
Greyhound	43
Coyote	40
Mule deer	35
Grizzly bear	30
Cat	30
Elephant	25
Pig	9

Data 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

5. Do birds or land animals seem to have the greatest variability in speeds? Explain your reasoning.

6. Find the 5-number summary for the speeds in each data set. What do the 5-number summaries tell you about the distribution of speeds for each data set?

7. Use the 5-number summaries to make a box plot for each of the two data sets.

0 25 50 75 100 125 150 175 200 225 250

Maximum speed of land animals (mph)

0 25 50 75 100 125 150 175 200 225 250

Maximum speed of birds (mph)

8. Write several sentences to tell someone about the speeds of birds and land animals.

Lesson Summary

In this lesson, you learned about the 5-number summary for a set of data: minimum, lower quartile, median, upper quartile, and maximum. You made box plots after finding the 5-number summary for two sets of data (speeds of birds and land animals), and you estimated the 5-number summary from box plots (number of Tootsie Pops people can hold, class scores). You also found the interquartile range (IQR), which is the difference between the upper quartile and lower quartile. The IQR, the length of the box in the box plot, indicates how closely the middle half of the data is bunched around the median. (Note that because sometimes data values repeat and the same numerical value may fall in two sections of the plot, it is not always exactly half. This happened with the two speeds of 50 mph – one went into the top quarter of the data and the other into the third quarter – the upper quartile was 50.)

You also practiced describing a set of data using the 5-number summary, making sure to be as precise as possible- avoiding words like “a lot” and “most” and instead saying about one half or three fourths.

7.5.15 Lesson Part 1 Date_________________________

Random Sampling

Student Objectives

- I can select a random sample from a population.
- I begin to develop an understanding of sampling variability.

Classwork

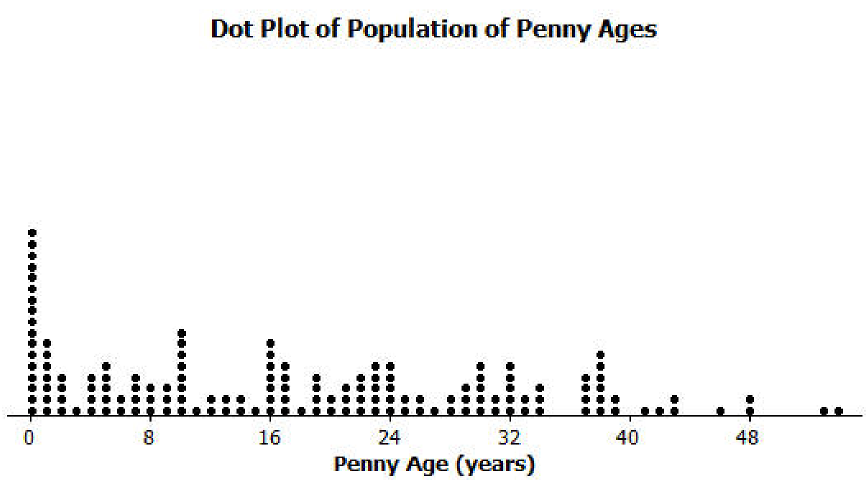
In this lesson, you will investigate taking random samples and how random samples from the same population vary.

Exercises 1–9: Sampling Pennies

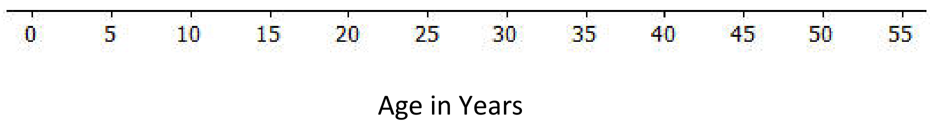
1. Do you think different random samples from the same population will be fairly similar? Explain your reasoning.

2. The plot below shows the number of years since being minted (the penny age) for 150 pennies that JJ had collected over the past year. Describe the shape, center, and spread of the distribution.

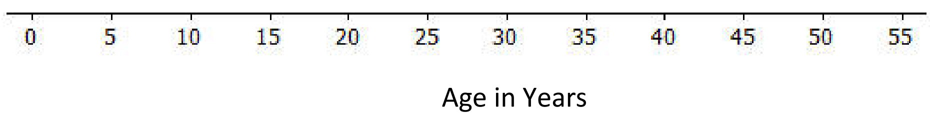
Example 1 – Refer to this plot for your first homework problem.



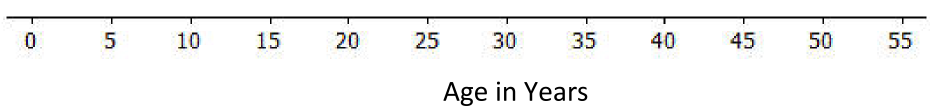
3. Make a dot plot of the ages of the ten pennies from ____________________sample of pennies from the jar.



4. Make a dot plot of the ages of the ten pennies from ____________________sample of pennies from the jar.

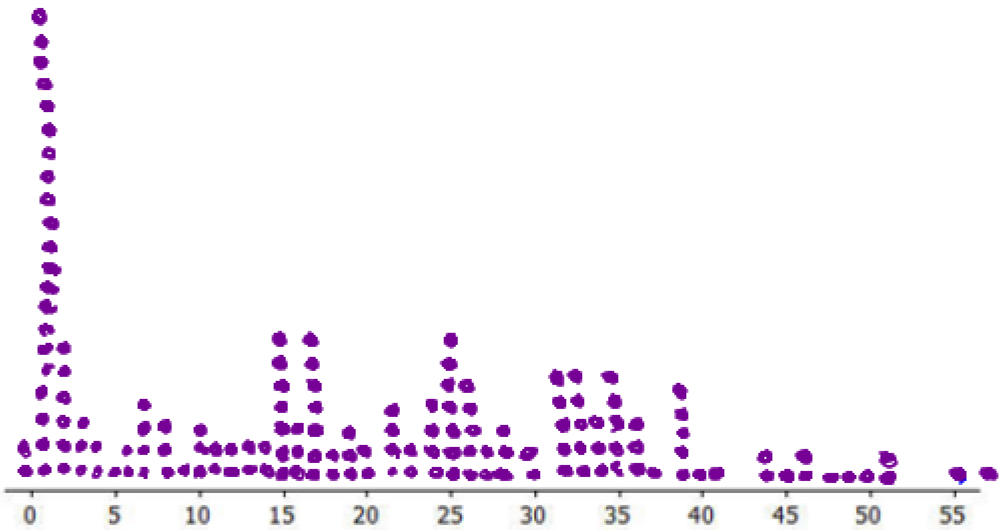


5. Make a dot plot of the ages of the ten pennies from ____________________sample of pennies from the jar.



6. Compare the distributions of the penny ages in these sample.

Ages of Pennies from Mr. Garrison's 2015 Collection



Age in Years

7. What do you observe about the population distribution?

a. How does your sample distribution compare to those of other students? To the population distribution?

b. Whose sample distribution was the most similar to the population distribution?

7.5.15 Lesson Part 2 Date_________________________

Random Sampling

Exercises 8–11: Grocery Prices and Rounding

8. Consider the following statistical question, "Do the store owners price the merchandise with cents that are closer to a higher dollar value or a lower dollar value?" Do you think more grocery prices will round up or round to the lower dollar value? Explain your thinking.

9. To investigate the question, "Do the store the store owners price the merchandise with cents that are closer to a higher dollar value or a lower dollar value?", you will look at some grocery prices in weekly flyers and advertising for local grocery stores.

   a. How would you round $\$3.49$, $\$4.99$, $\$2.50$, and $\$2.00$ to the nearest dollar?

   b. If the advertised price was three for $\$4.35$, how much would you expect to pay for one item?

10. Follow your teacher's instructions to select a random sample of 25 items without replacement, and record the items and their prices in the table below.

Item	Price	rounded
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		

item	price	rounded
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		

11. Count the number of times you rounded up and the number of times you rounded down.

Group	Number of times prices were rounded to the higher value.	Number of times the prices were rounded to the lower value.	Percent of prices rounded up.

a. Given the results of the different samples, how would you answer the question: Are grocery prices in the weekly ads at the local grocery closer to a higher dollar value or a lower dollar value?

b. Identify the population, sample, and sample statistic used to answer the statistical question.

Population:

Sample:

Sample Statistic:

Lesson Summary

In this lesson, you took random samples in two different scenarios. In the first scenario, you physically reached into a jar and drew a random sample from a population of pennies. In the second scenario, you drew items from a bag and recorded the prices. In both activities, you investigated how random samples of the same size from the same population varied. Even with sample sizes of 10, the sample distributions of pennies were somewhat similar to the population (the distribution of penny ages was skewed right, and the samples all had 0 as an element). In both cases, the samples tended to have similar characteristics. For example, the samples of prices from the same store had about the same percent of prices that rounded to the higher dollar value.

7.5.17 Sampling Variability Date____________________

Student Objectives

- I can use data from a random sample to estimate a population mean.
- I can understand the term *sampling variability* in the context of estimating a population mean.

Example 1: Estimating a Population Mean

The owners of a gym have been keeping track of how long each person spends at the gym. Two hundred of these times (in minutes) are shown in the population tables located at the end of the lesson. These 200 times will form the population that you will investigate in this lesson.

Look at the values in the population. Can you find the longest time spent in the gym in the population? _______Can you find the shortest?_______

On average, roughly how long do you think people spend at the gym? In other words, by just looking at the numbers in the table, make an estimate of the population mean.

____

Instead of doing a calculation using every value in the population, we will use a random sample to find the mean of the sample. The sample mean will then be used as an estimate of the population mean.

Exercises 1–4

Initially, you will select just five values from the population to form your sample. This is a very small sample size, but it is a good place to start to understand the ideas of this lesson.

Select a random sample of five.

1. Use the die to select five values from the population of times. What are the five observations in your sample?

2. For the sample that you selected, calculate the sample mean.

3. You selected a random sample and calculated the sample mean in order to estimate the population mean. Do you think that the mean of these five observations is exactly correct for the population mean? Explain.

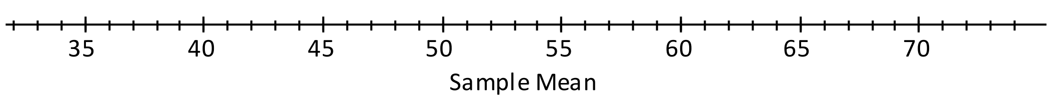
4. In practice, you only take one sample in order to estimate a population characteristic. But, for the purposes of this lesson, your classmates took other random samples from the same population of times at the gym. Could the other sample means be closer to the population mean than the mean of these five observations? Could they be further from the population mean?

As a class, you will now investigate sampling variability by looking at several samples from the same population. Each sample will have a different sample mean. This variation provides an example of sampling variability.

Exercises 5–8

5. Write the sample means for everyone in the class in the space below.

6. Use all the sample means to make a dot plot using the axis given below. (Remember, if you have repeated or close values, stack the dots one above the other.)



7. What do you see in the dot plot that demonstrates sampling variability?

8. Remember that in practice you only take one sample. (In this lesson, many samples were taken in order to demonstrate the concept of sampling variability.) Suppose that a statistician plans to take a random sample of size 5 from the population of times spent at the gym and that he or she will use the sample mean as an estimate of the population mean. Approximately how far can the statistician expect the sample mean to be from the population mean?

Population

	0	1	2	3	4	5	6	7	8	9
01	50	45	45	66	71	55	65	33	60	51
02	53	83	40	51	83	57	75	38	43	77
03	49	49	81	57	42	36	22	66	68	52
04	60	67	43	60	55	63	56	44	50	58
05	64	41	67	73	55	69	63	46	50	65
06	54	58	53	55	51	74	53	55	64	16
07	28	48	62	24	82	51	64	45	41	47
08	70	50	38	16	39	83	62	50	37	58
09	79	62	45	48	42	51	67	68	56	78
10	61	56	71	55	57	77	48	65	61	62
11	65	40	56	47	44	51	38	68	64	40
12	53	22	73	62	82	78	84	50	43	43
13	81	42	72	49	55	65	41	92	50	60
14	56	44	40	70	52	47	30	9	58	53
15	84	64	64	34	37	69	57	75	62	67
16	45	58	49	78	59	36	52	39	70	51
17	50	45	45	66	71	55	65	33	60	51
18	53	83	40	51	83	57	75	38	43	77
19	49	49	81	57	42	36	22	66	68	52
20	60	67	43	60	55	63	56	44	50	58

Lesson Summary

A population characteristic is estimated by taking a random sample from the population and calculating the value of a statistic for the sample. For example, a population mean is estimated by selecting a random sample from the population and calculating the sample mean.

The value of the sample statistic (e.g., the sample mean) will vary based on the random sample that is selected. This variation from sample to sample in the values of the sample statistic is called **sampling variability**.

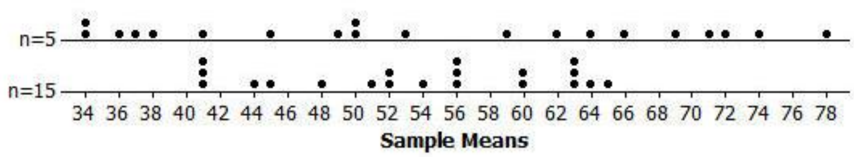
7.5.18 Sampling Variability and the Effect of Sample Size Date________________

Student Objectives

- I can use data from a random sample to estimate a population mean.
- I know that increasing the sample size decreases the sampling variability of the sample mean.

Example 1: Sampling Variability

The previous lesson investigated the statistical question, "What is the typical time spent at the gym?" by selecting random samples from the population of 200 gym members. Two different dot plots of sample means calculated from random samples from the population are displayed below. The first dot plot represents the means of 20 samples with each sample having 5 data points. The second dot plot represents the means of 20 samples with each sample having 15 data points.



Based on the first dot plot, Jill answered the statistical question by indicating the mean time people spent at the gym was between 34 and 78 minutes. She decided that a time approximately in the middle of that interval would be her estimate of the mean time the 200 people spent at the gym. She estimated 52 minutes. Scott answered the question using the second dot plot. He indicated that the mean time people spent at the gym was between 41 and 65 minutes. He also selected a time of 52 minutes to answer the question.

- Describe the differences in the two dot plots.

- Which dot plot do you feel more confident in using to answer the statistical question? Explain your answer.

- In general, do you want sampling variability to be large or small? Explain.

Exercises 1–2

In the previous lesson, you saw a population of 200 times spent at the gym. You will now select a random sample of size 15 from that population. You will then calculate the sample mean.

1. Write the values of all 15 observations that you have selected randomly in the space below.

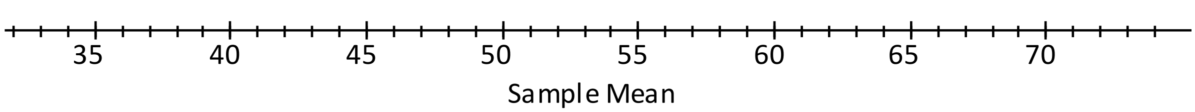
2. Calculate the mean of your 15 sample values. Write the value of your sample mean below. Round your answer to the nearest tenth. (Be sure to show your work.)

Exercises 3–5

You will now use the sample means from Exercise 2 for the entire class to make a dot plot.

3. Write the sample means for everyone in the class in the space below.

4. Use all the sample means to make a dot plot using the axis given below. (Remember, if you have repeated values or values close to each other, stack the dots one above the other.)



5. In the previous lesson, you drew a dot plot of sample means for samples of size 5. How does the dot plot of sample means for samples of size 15 compare to the dot plot of sample means for samples of size 5? For which sample size (5 or 15) does the sample mean have the greater sampling variability?

Exercises 6–8

6. Remember that in practice you only take one sample. Suppose that a statistician plans to take a random sample of size 15 from the population of times spent at the gym and will use the sample mean as an estimate of the population mean. Based on the dot plot of sample means that your class collected from the population, approximately how far can the statistician expect the sample mean to be from the population mean? (The actual population mean is 53.9 minutes.)

7. How would your answer in Exercise 6 compare to the equivalent mean of the distances for a sample of size 5?

8. Suppose you have a choice of using a sample of size 5, 15, or 25. Which of the three makes the sampling variability of the sample mean the smallest? Why would you choose the sample size that makes the sampling variability of the sample mean as small as possible?

Lesson Summary

The greater the sample size, the smaller the sampling variability of the sample mean. In other words, the more data points that you use to estimate a mean, the better chances that your estimate is accurate.

7.5.21 Different Sample Means Date________________________

Student Objective

- I understand that a *meaningful* difference between two sample means is one that is greater than would have been expected due to just sampling variability.

Classwork

There are three bags, Bag A, Bag B, and Bag C, with 100 numbers in each bag. You and your classmates will investigate the population mean (the mean of all 100 numbers) in each bag. Each set of numbers has the same range. However, the population means of each set may or may not be the same. We will see who can uncover the mystery of the bags!

Exercises 1–4

1. To begin your investigation, start by selecting a random sample of ten numbers from Bag A. Remember to mix the numbers in the bag first. Then, select one number from the bag. Do not put it back into the bag. Write the number in the chart below. Continue selecting one number at a time until you have selected ten numbers. Mix up the numbers in the bag between each selection.

Selection	1	2	3	4	5	6	7	8	9	10
Bag A										

a. **Create a dot plot** of your sample of ten numbers. Use a dot to represent each number in the sample.

b. Based on the dot plot, what would you **estimate the mean** of the numbers in Bag A to be? ________

2. Repeat the process by selecting a random sample of ten numbers from Bag B.

Selection	1	2	3	4	5	6	7	8	9	10
Bag B										

a. **Create a dot plot** of your sample of ten numbers. Use a dot to represent each of the numbers in the sample. Use the dot plot to **estimate the mean** of Bag B.

____

b. Based on your dot plot, do you think the mean of the numbers in Bag B is the same or different than the mean of the numbers in Bag A? Explain your thinking.

3. Repeat the process once more by selecting a random sample of ten numbers from Bag C.

Selection	1	2	3	4	5	6	7	8	9	10
Bag C										

a. Create a dot plot of your sample of ten numbers. Use a dot to represent each of the numbers in the sample. Use the dot plot to **estimate the mean** of Bag C. __________

b. Based on your dot plot, do you think the mean of the numbers in Bag C is the same or different than the mean of the numbers in Bag A? Explain your thinking.

4. Calculate the mean of the numbers for each of the samples from Bag A, Bag, B, and Bag C.

	Mean of the sample of numbers
Bag A	
Bag B	
Bag C	

a. Are the sample means you calculated the same as the sample means of other members of your class?

	Mean of the sample of numbers
Bag A	
Bag B	
Bag C	

b. After creating dot plots for the sample means of the class, compare the sample means for Bag A and for Bag B. Compare the sample means for Bag A and for Bag C. Can you be sure which bag has the larger population mean?

____

____

____

Exercises 5–13

5. Find the difference between sample mean of Bag A and the sample mean of Bag B, $(Sample\ Mean_A - Sample\ Mean_B)$, for your samples. Do you think that the two populations (Bags A and B) have different means or do you think that the two population means might be the same?

6. Plot your difference of the means, $(Sample\ Mean_A - \text{Sample}\ Mean_B)$, on a class dot plot.

7. Describe the center of differences plotted on the graph.

8. Calculate the sample mean of Bag A minus the sample mean of Bag C, $(Sample\ Mean_A - Sample\ Mean_C)$.

9. Plot your difference $(Sample\ Mean_A - Sample\ Mean_C)$ on a class dot plot.

10. Describe the center of differences plotted on the graph.

11. Each bag has a population mean that is either 10.5 or 14.5. State what you think the population mean is for each bag. Explain your choice for each bag.

12. What is $(Mean_A - Mean_B)$? ______ Why are the differences in the sample means of Bag A and Bag B not always 0?

13. What is $(Mean_A - Mean_C)$? ______What did your plot predict it is?_______ Why don't they always match?

Lesson Summary:

The differences in the class dot plots occur because of sampling variability—the chance variability from one sample to another. You were asked about how great the difference in sample means would need to be before you have convincing evidence that one population mean is larger than another population mean. A "*meaningful*" difference between two sample means is one that is unlikely to have occurred by chance if the population means are equal. In other words, the difference is one that is greater than would have been expected just due to sampling variability.

7.5.22 Lesson Date________________________

Using Sample Data to Decide if Two Population Means are Different

Student Objectives

- I can know if the mean or the median is a more appropriate measure of center.
- I know when a difference in sample means provides evidence that the population means are different.

1: Collecting a Sample – Class Discussion

Your teacher will give you a paper that lists a data set with 100 numbers in it. Place a star by each method of obtaining a sample of size 20 that would produce a random sample.

Option 1: A spinner has 10 equal sections on it. Spin once to get the row number and again to get the column number for each member of your sample. Repeat this 20 times.

Option 2: Since the data looks random already, use the first two rows.

Option 3: Cut up the data and put them into a bag. Shake the bag to mix up the papers, and take out 20 values.

Option 4: Close your eyes and point to one of the numbers to use as your first value in your sample. Then, keep moving one square from where your finger is to get a path of 20 values for your sample.

2: Sample Probabilities

Continue working with the data set your teacher gave you in the previous activity. The data marked with a star all came from students at Springfield Middle School.

1. Select a random sample of size 10. Record your data here.

3: Estimating a Measure of Center for the Population

1. Decide which **measure of center** (mean or median) makes the most sense to use based on the distribution of your sample. Discuss your thinking with your partner. If you disagree, work to reach an agreement. ______________
2. Calculate the **measure of center** for your sample.

Data Set 1 Center	

Data Set 2 Center	

Difference in Centers	

3. Calculate the **measure of variability** (MAD or IQR) for your sample that goes with the measure of center that you found.

Data Set 1										

Data Set 2										

Typical Spread Set 1 _________ Typical Spread Set 2 _________

4: Comparing Populations

Is it reasonable to conclude that the measures of center for each of your populations are meaningfully different? Explain or show your reasoning

7.5.23 Lesson Date________________________

Using Sample Data to Decide if Two Population Means are Different

Student Objective

- I can use data from random samples to draw informal inferences about the difference in population means.

In the previous lesson, you described how far apart the means of two data sets are in terms of the MAD (mean absolute deviation), a measure of variability. In this lesson, you will extend that idea to informally determine when two sample means computed from random samples are far enough apart from each other so to imply that the population means also differ in a "meaningful" way. Recall that a "meaningful" difference between two means is a difference that is greater than would have been expected just due to sampling variability.

Example 1: Texting

With texting becoming so popular, Linda wanted to determine if middle school students memorize *real* words more or less easily than *fake* words. For example, real words are "food," "car," "study," "swim;" whereas fake words are "stk," "fonw," "cqur," "ttnsp." She randomly selected 28 students from all middle school students in her district and gave half of them a list of 20 real words and the other half a list of 20 fake words.

Exercises 1–3

1. How do you think Linda might have randomly selected 28 students from all middle school students in her district?

2. Why do you think Linda selected the students for her study randomly? Explain.

3. She gave the selected students one minute to memorize their list after which they were to turn the list over and after two minutes write down all the words that they could remember. Afterwards, they calculated the number of correct "words" that they were able to write down. Do you think the lists of real and fake words need to be the same length? Explain your reasoning.

Exercises 4–7

Suppose the data (number of correct words recalled) she collected were:

For students given the real words list: $8, 11, 12, 8, 4, 7, 9, 12, 12, 9, 14, 11, 5, 10$.

For students given the fake words list: $3, 5, 4, 4, 4, 7, 11, 9, 7, 7, 1, 3, 3, 7$.

4. On the same scale, draw dot plots for the two data sets.

5. From looking at the dot plots, compare the distribution of the number of correctly recalled real words with the distribution of number of correctly recalled fake words.

   Center:

   Shape:

   Spread:

6. Linda made the following calculations for the two data sets:

	Mean	MAD
Real words recalled	9.43	2.29
Fake words recalled	5.36	2.27

What is the difference in the means?___________

What is 2 times the larger MAD?________

If the difference in the means is less than 2 times the greater MAD, then you can conclude that the difference in the sample means **might** just be sampling variability and that there may **not** be a meaningful difference in the population means. Using these criteria, what can Linda conclude about the difference in population means based on the sample data that she collected?

Lesson Summary

To determine if the mean value of some numerical variable differs for two populations, take random samples from each population. It is very important that the samples be random samples. If the number of MADs that separate the two sample means is 2 or more, then it is reasonable to think that the populations have different means. Otherwise, the population means are considered to be the same.