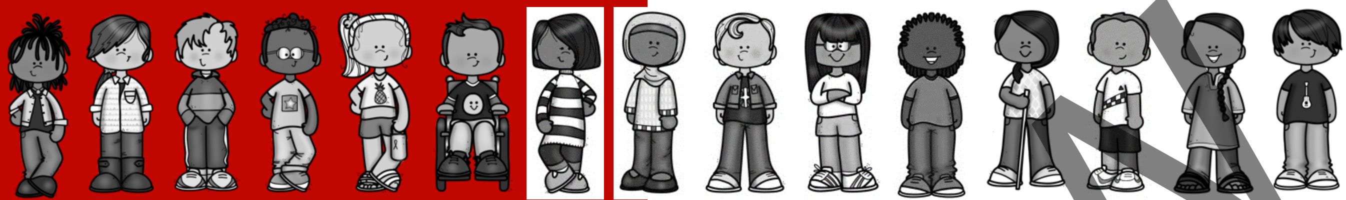
Name ______________________________ Period __________ Date ______________

UNIT 10
STUDENT PACKET

MathLinks

GRADE 7



SAMPLING

		Monitor Your Progress	Page
	My Word Bank		0
10.0	**Opening Problem: Screen Time**		1
10.1	**Introduction to Sampling** • Differentiate between theoretical probability and experimental probability • Identify populations and samples • Use random sampling to make valid inferences about populations	 3 2 1 0 3 2 1 0 3 2 1 0	2
10.2	**Comparing Samples** • Use measures of center and spread to compare numerical data sets • Use dot and box plots to visually compare data sets	 3 2 1 0 3 2 1 0	7
10.3	**Fish in a Lake** • Use a mathematical model with random sampling and proportional reasoning to estimate an entire population • Use data displays and measures of center and spread to make inferences about populations	 3 2 1 0 3 2 1 0	13
	Review		17
	Student Resources		25

Parent (or Guardian) signature ___________________________________

MY WORD BANK

Explain the mathematical meaning of each word or phrase, using pictures and examples when possible. See **Student Resources** for mathematical vocabulary.

measure of center mean, median, mode	Measure of spread (variability) range, interquartile range, mean absolute deviation
percent error	population
sample	simulation

OPENING PROBLEM: SCREEN TIME

Follow your teacher's directions.

(1)

(2)

(3)

INTRODUCTION TO SAMPLING

We will use examples to explore the difference between theoretical probability and experimental probability. We will learn about populations and samples, and we will explore biased and random sampling.

[7.SP.1, 7.SP.5, 7.SP.7a; SMP1, 3, 4, 6, 7]

GETTING STARTED

Recall that a statistical question is a question where numerical data that has potential for variability can be collected and analyzed for the purpose of answering the question.

Example: "How much TV do students in my class watch on average?"
Non-example: "How many hours of TV did you watch last week?"

For each problem 1 – 5, put a check next to the one that is a statistical question.

1. How tall are you?
   How tall is the average student at your school?

2. How much time per night do you spend on homework during a typical week?
   What did you have for homework last night?

3. How many texts did you send yesterday?
   How many texts does the average 13-year old send in a day?

4. What is the high temperature for today?
   What is the average high temperature for the city for today's date?

5. What is your favorite after-school club?
   What is the favorite after-school club at your school?

6. Suppose you want to know what is considered the favorite after-school club at your school. Select the methods that you think will best let you know.

   a. Asking the person sitting next to you.
   b. Asking all of the boys in your school.
   c. Publishing a survey in the school paper.
   d. Asking all of the teachers in your school.
   e. Asking each member of the math club.
   f. Asking every fifth person who enters the school.

REVISITING PROBABILITY

Follow your teacher's directions for (1) – (3).

(1)

(2)

Trial	Result	Cumulative Totals for Blue			Prediction
		Total Drawn	As a fraction	As a percent	# of blue in bag
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

(3)

4. Determine if each situation below describes a theoretical probability or experimental probability situation.

a. The XYZ insurance company determines that a 25-year-old male must pay a higher automobile insurance premium than his 56-year-old mother.	b. You have a full deck of shuffled playing cards and predict that you have a 25% chance of drawing a card that is clubs.

POPULATIONS AND SAMPLING

Follow your teacher's directions for (1).

(1)

Situation			
a. Of the 371 teenagers surveyed, 29.9% said they never owned a pet due to allergies.			
b. A reporter surveyed 100 freshmen at UCLA and found that 95% of them lived on campus.			
c. A news station reported that of the 1,385 American families asked, 92% of them had internet service.			
d. The average late fee for 500 credit card holders was found to be \$52.24.			

2. Consider the question, "What is the likelihood that a U.S. high school athlete gets a concussion?" Explain why each sample is not random (is biased) regarding this question.

a. Ask your sister about her tennis team.	b. Ask the football coach at the local high school about his players.	c. Randomly choose 3 players to ask from each sport at the local high school.

SAMPLING SORT

1. Your group will be given a set of cards.
   - Read one card at a time and discuss what are the population and sample
   - Determine whether the sample is biased or random
   - Record each card letter in the appropriate column

Examples of Biased Sampling	Examples of Random Sampling

2. Choose three of the biased sampling examples from above. Explain why they are biased and suggest a way to get a ,ore random sample.

Card	Creating a Random Sampling Situation

3. What is the same about every sample?

4. Why do we want to avoid choosing a biased sample for statistical analysis?

5. Record the meanings of <u>population</u> and <u>sample</u> in **My Word Bank**.

PRACTICE 1

1. For which populations would taking a sample be more efficient than surveying the entire population?
   a. Determining the average score on a class test.
   b. Determining the favorite ice cream flavor in Vermont.
   c. Determining the most common injury cared for in an emergency room.
   d. Determining your family's favorite fast-food restaurant.

2. You want to estimate the number of students who walk to school. Which sample is most representative of the school population?
   a. The first 10 people you see during passing period.
   b. Selecting 50 students randomly by ID numbers from each grade.
   c. Asking every other student in the lunch line during seventh grade lunch time.

3. The student council wants to know which assembly students liked best this year. Each student council member asked 10 students to vote.

   a. How might the council choose the students so that it is representative of the school's population?

   b. How might the council choose the students so that it does **not** represent the entire school population?

4. In a poll taken in Mr. Fernandez's math class, 70% of the students said he was their favorite teacher. The school newspaper wants to run an article stating that Mr. Fernandez is the favorite math teacher in the school. Explain why this is not a valid conclusion, and suggest a better way to determine who is the favorite math teacher at the school.

5. Refer back to the opening problem, **Screen Time**.
   a. What is the population being considered?

   b. Why would it be inefficient to survey the entire population (like we do for the Census)?

   c. Suggest a method that the news organization could use to appropriately sample the population.

COMPARING SAMPLES

We will use statistical measures and data displays to compare samples.

[7.EE.3, 7.RP.3, 7.SP.1, 7.SP.2, 7.SP.3, 7.SP.4; SMP1, 2, 4, 5, 6, 7]

GETTING STARTED

Below is a random sample of 10 teens' estimated daily screen times rounded to the nearest full hour.

Liam	Mia	Xander	Noah	Emma	Ava	William	Olivia	Quentin	Evelyn
5	6	9	7	10	7	4	10	7	5

1. Rewrite the times in order from least to greatest.

Match the letters A – I to each description below. Then write the values of each measure based upon the screen time sample above. Refer to **Student Resources** as needed.

A. mean	B. 1st Quartile (Q_1)	C. number of observations (*n*)
D. range	E. median (*med* = Q_2)	F. five-number summary
G. minimum (*min*)	H. 3rd Quartile (Q_3)	I. interquartile range (IQR)

	A – I	Description	Value(s)
2.		The number of pieces of data collected.	
3.		The middle number, when the observations are ordered from least to greatest.	
4.		The arithmetic average.	
5.		The lowest number of times watching.	
6.		The difference between the highest and lowest number of times watching.	
7.		The middle value of the lower half of the data set.	
8.		The middle value of the upper half of the data set.	
9.		The following five data points: (*min*, Q_1, *med*, Q_3, *max*).	
10.		The difference between Q_3 and Q_1.	

GETTING STARTED

Continued

11. Make a line plot and box-plot of the screen time data.

Line Plot	**Box-Plot**

12. The mean number of times watching a screen: is _____. Compute the mean absolute deviation (MAD) for the screen time sampling. See **Student Resources** if needed.

Screen Time data (number of times)	**Distance from data point to mean**
5	$\lvert 5 - \_\_\_\_ \rvert = \_\_\_\_$
6	
9	
7	
10	
7	
4	
10	
7	
5	

Sum of distances = _____; MAD = _____

13. Name the measure of spread associated with each measure of center.

Measure of center: mean Measure of spread: ______________________

Measure of center: median Measure of spread: ______________________

MATH SCORE SAMPLES

All seventh-grade students in Shermer School District took math benchmark assessments. Here are samples of student scores selected randomly for two of the middle schools.

1. Rewrite the lists of scores in order from least to greatest.

Burkhart MS	13	18	13	15	13	16	14	20	13	15	17	22

Starke MS	14	22	21	14	12	15	20	21	17	25	14	15

Calculate each statistic below.

Statistic	Burkhart MS	Starke MS
2. number of observations (*n*)		
3. minimum (*min*)		
4. maximum (*max*)		
5. range		
6. mean		
7. median (*med* = Q_2)		
8. 1st quartile (Q_1)		
9. 3rd quartile (Q_3)		
10. interquartile range (IQR)		
11. five-number summary		

12. Based on the data above, what is the typical score on the benchmark assessment for each school? Explain.

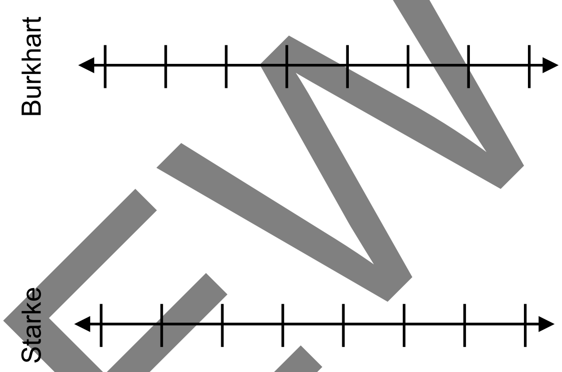
13. Record the meanings of <u>measure of center</u> and <u>measure of spread</u> in **My Word Bank**.

MATH SCORE SAMPLES
Continued

Use the test score data to create the following data displays.

14. Create dot plots for each sample.

15. Create box plots for each sample.



Using the displays above, circle the best answer for each of the following statements.

16. The center of the Burkhart distribution is ____ the center of the Starke distribution.

greater than less than about the same as

17. The variability of the Burkhart distribution is ____ the variability of the Starke distribution.

greater than less than about the same as

18. Three weeks later, the district reported actual mean test scores for each school. Find each school's sample mean error, called, <u>percent error</u>, by using this formula.

$$\text{Percent error} = \left| \frac{\text{actual} - \text{estimate}}{\text{actual}} \right|, \text{ written as a percent}$$

School	Actual Mean	Sample Mean (estimate)	Percent Error
Burkhart	15		
Starke	16		

19. Which school's data sample represented their seventh-grade school population better? How do you know?

PRACTICE 2

Below are the heights (in inches) of athletes from two sports at UCLA in 2014. They were selected randomly from the rosters and organized in order from least to greatest.

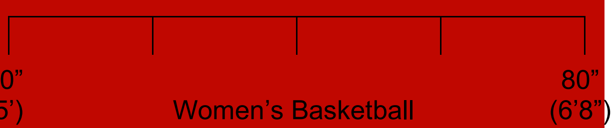
Women's basketball team (B)

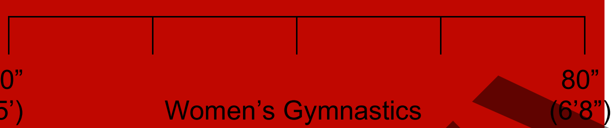
66	68	69	69	72	73	73	73	75	78

Women's gymnastics team (G)

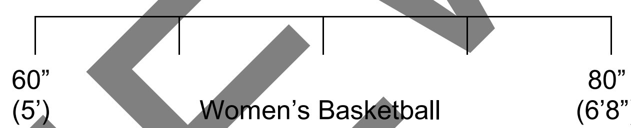
61	61	62	62	63	63	64	64	64	65	65

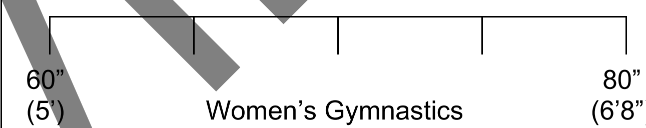
1. Create dot plots for each sample.





2. Create box plots for each sample.





Use the displays above to answer the following questions.

3. The __________________ team has the greater measure of center. What does this tell you about the heights of this team as compared to the heights of the other team?

4. The ____________________ team has the lesser measure of spread. What does this tell you about the heights of this team as compared to the heights on the other team?

5. What can you conclude about the typical height of a gymnast versus the typical height of a basketball player from these teams?

6. Another gymnast joined the team and is 6’4. How would the new gymnast’s height affect the center and variability of team data distributions?

PRACTICE 3

Rachel selected small random samples of housing prices for Texas (TX) and California (CA). The data is in the table below.

	Cost (in thousands of dollars; for example, 100 means $100,000)									
Texas	190	178	235	298	109	275	123	280	122	200
California	320	296	502	277	548	665	410	750		

Rachel used an online statistics calculator to compute the mean and the MAD for each data set, but she didn't organize her work or label it. Here are the six numbers she wrote down.

471,000 10 56.8 201,000 8 145.25

1. Use the data samples and Rachel's numbers to complete the table below.

	Texas	**California**
number of observations		
Mean		
MAD		

2. On average, which state do you think has more expensive homes? By how much?

3. Which state do you think has more consistent housing prices? Explain why you think this.

4. The variability of the CA sample is about _____ times the variability of the TX sample.

5. According to McMath's Realty, the average house in Texas costs $195,000 and the average house in California costs $445,000. Find each data sample's error as a percent.

   Average cost in TX percent error: _____ Average cost in CA percent error: _____

6. The percent error of the CA sample is about _____ times greater than the TX sample.

7. Which sample is closer to the actual average house cost in the state? How do you know?

8. Record the meaning of percent error in **My Word Bank**.

FISH IN A LAKE

We will do a sampling experiment to estimate the fish population in a lake. We will make inferences about fish populations using data displays and measures of center and spread.

[7.EE.3, 7.RP.3, 7.SP.1, 7.SP.2, 7.SP.4, 7.SP.8c, SMP1, 2, 3, 4, 5, 6, 7, 8]

GETTING STARTED

Solve each equation using any method.

1. $\frac{2}{3} = \frac{x}{27}$	2. $\frac{x}{12} = \frac{3}{15}$
3. $\frac{6}{x} = \frac{4}{7}$	4. $\frac{5}{14} = \frac{12}{x}$

ESTIMATING FISH POPULATIONS

Follow your teacher's directions for (1) – (3).

(1) – (2)

Total number of marked fish:	Marked fish sample:
Total fish population:	Total fish sample:

(3)

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
S_{marked}						
S_{total}						
P_{marked}						

4. Create and solve equations to estimate the number of fish in the lake for each sample.

Sample 1	Sample 2	Sample 3
Sample 4	**Sample 5**	**Sample 6**

5. Based on your experiment, estimate the number of fish in the population. __________
   Explain how you arrived at this estimate.

6. Count all of the fish. The actual number of fish in the population is ___________.

7. Find the error in your estimate as a percent of the actual population. ___________

PRACTICE 4: FISH LENGTHS

1. Record the meaning of simulation in **My Word Bank**.

In **Estimating Fish Populations**, suppose that when fish were marked, they were also measured. Here are fish lengths in centimeters from two different random samples from two different lakes:

Sample A: 75, 32, 38, 42, 47, 68, 51, 51, 61, 31, 51, 62
Sample B: 49, 45, 51, 49, 63, 56, 51, 48, 52, 42, 51, 52

2. Rewrite each list in order from least to greatest.

Sample A												
Sample B												

3. Calculate statistics for the two data sets. Use additional paper if necessary.

Statistic	Sample A	Sample B
number of observations		
minimum		
maximum		
range		
mean		
median		
mode		
five-number summary		
interquartile range (IQR)		
mean absolute deviation (MAD), rounded to the hundredths		

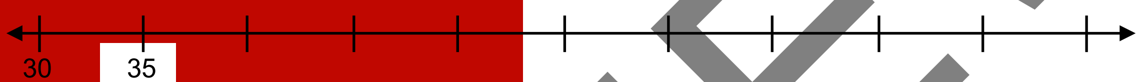
PRACTICE 4: FISH LENGTHS
Continued

4. Compare the measures of center for each sample. What do you notice?

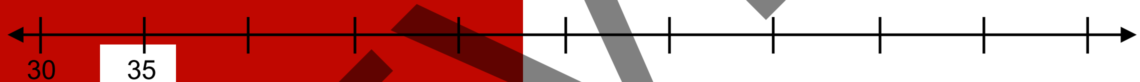
5. Which sample had more consistent lengths of fish? Justify with the data above.

6. Create box plots from the data on the previous page for Sample A and Sample B using the scale provided below.

Sample A:



Sample B:



7. Why is it important to use the same scale when comparing the data?

8. About what percent of the data is included between Q_1 and Q_3?

9. What do the box plot and interquartile range tell us about fish lengths in the lakes?

10. What is a statistic NOT included in the box plot that you think is relevant or useful? Explain.

REVIEW

POSTER PROBLEMS: SAMPLING

Part 1: Your teacher will divide you into groups.

- Identify members of your group as A, B, C, or D
- Each group will start at a numbered poster. Our group start poster is ______
- Each group will have a different colored marker. Our group marker is ________

Part 2: Follow your teacher's directions. Complete problems on the posters.

Poster 1 (or 5)	Two stores sold the following shoe sizes during the last hour: Store X: 9, 7, 8, 8, 10, 8, 6, 5, 9, 8 Store Y: 9, 6, 7, 6, 8, 6, 7, 5, 6, 8
Poster 2 (or 6)	Housing prices (in thousands) for the most recent sales in two housing areas of Mathville: Area X: \$350, \$400, \$800, \$370, \$425, \$320, \$350, \$360, \$365, \$300 Area Y: \$475, \$470, \$460, \$375, \$500, \$450, \$650, \$480, \$500, \$410
Poster 3 (or 7)	Teens were surveyed on the number of hours per week they spend on their phone from two different math classes: Class X: 63, 50, 40, 15, 35, 45, 54, 29, 25, 37 Class Y: 28, 35, 30, 42, 40, 32, 25, 28, 21, 29
Poster 4 (or 8)	The number of family pets from two random samples: Sample X: 5, 3, 2, 0, 1, 1, 4, 5, 3, 1, 0, 2, 1, 0, 2 Sample Y: 3, 1, 2, 0, 1, 2, 7, 2, 2, 3, 0, 1, 0, 2, 4

A. Copy the data in numerical order and determine the mean for each data set.
B. Find the five-number summary for each data set.
C. Make an appropriate data display for each data set. Be sure to label each graph.
D. Write a statistical question you could ask that would compare the data sets.

Part 3:

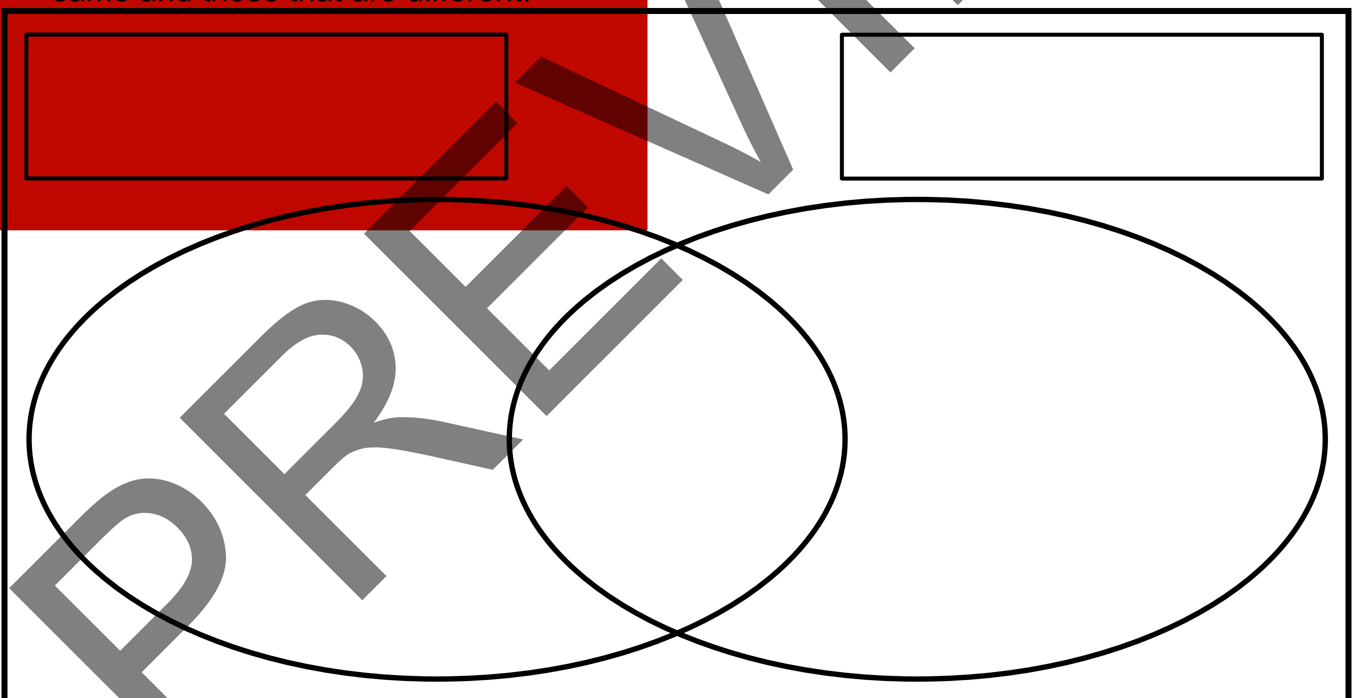
Return to your seats with your original poster. Work with your group. Answer the statistical question from problem D.

MATCH AND COMPARE SORT: SAMPLING

1. Individually, match words with descriptions. Record results.

Card set △			Card set ○		
Card number	**word**	**Card letter**	**Card number**	**word**	**Card letter**
I			**I**		
II			**II**		
III			**III**		
IV			**IV**		

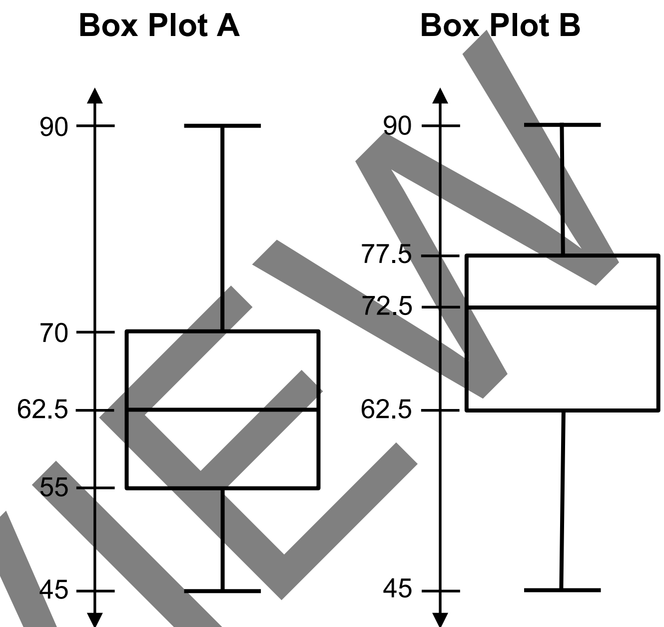
2. Partners, choose a pair of numbered matched cards and record the attributes that are the same and those that are different.



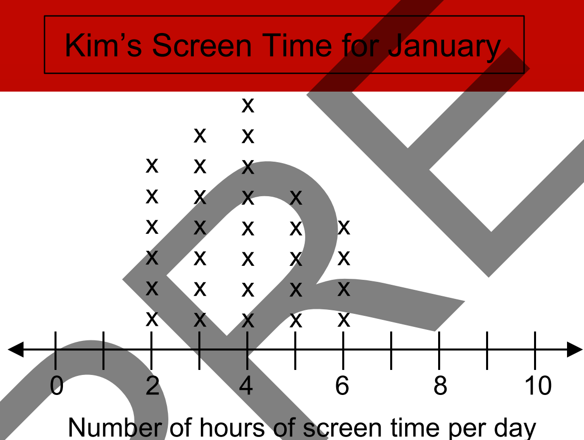
3. Partners, choose another pair of numbered matched cards and discuss the attributes that are the same and those that are different.

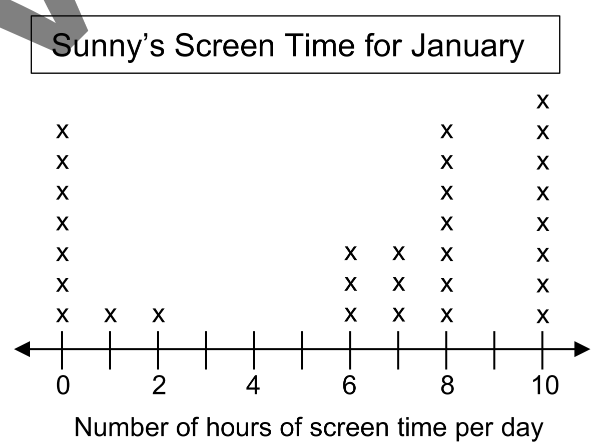
WOULD YOU RATHER?

1. Would your teacher rather have the class data from an exam look like box plot A or B? Explain. The five-number-summaries are clearly marked for each.

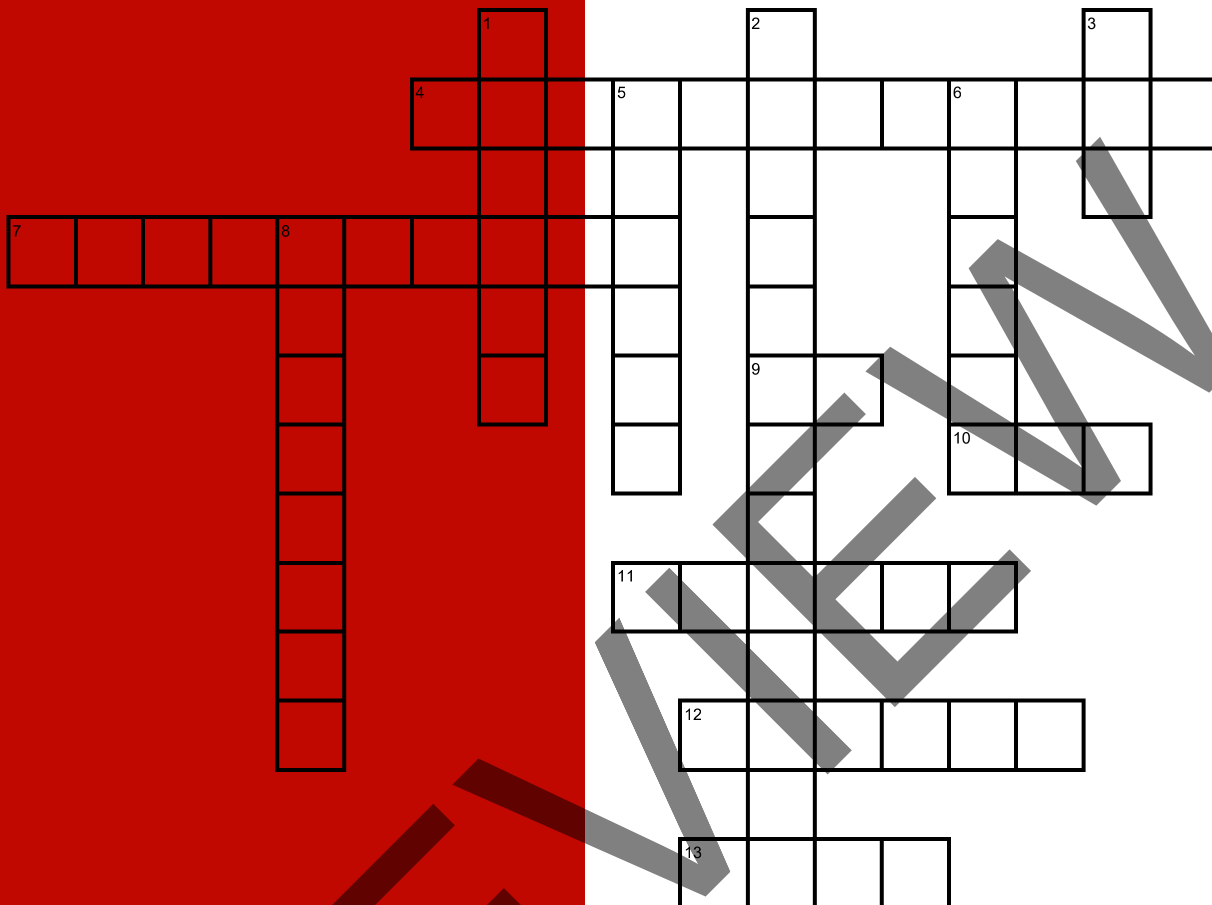


2. Would you rather have Kim's or Sunny's screen watching habits? Explain.





VOCABULARY REVIEW



Across

4 a measure that includes the difference between an actual measure and its estimate

7 an entire group of people or objects

9 First quarter of a sample (abbreviation)

10 measure of spread associated with the mean (abbreviation)

11 another word for variability

12 sample where every person does not have an equal chance to be selected

13 a measure of center

Down

1 measure of center associated with IQR

2 middle 50% of a data set

3 plot showing the five-number summary

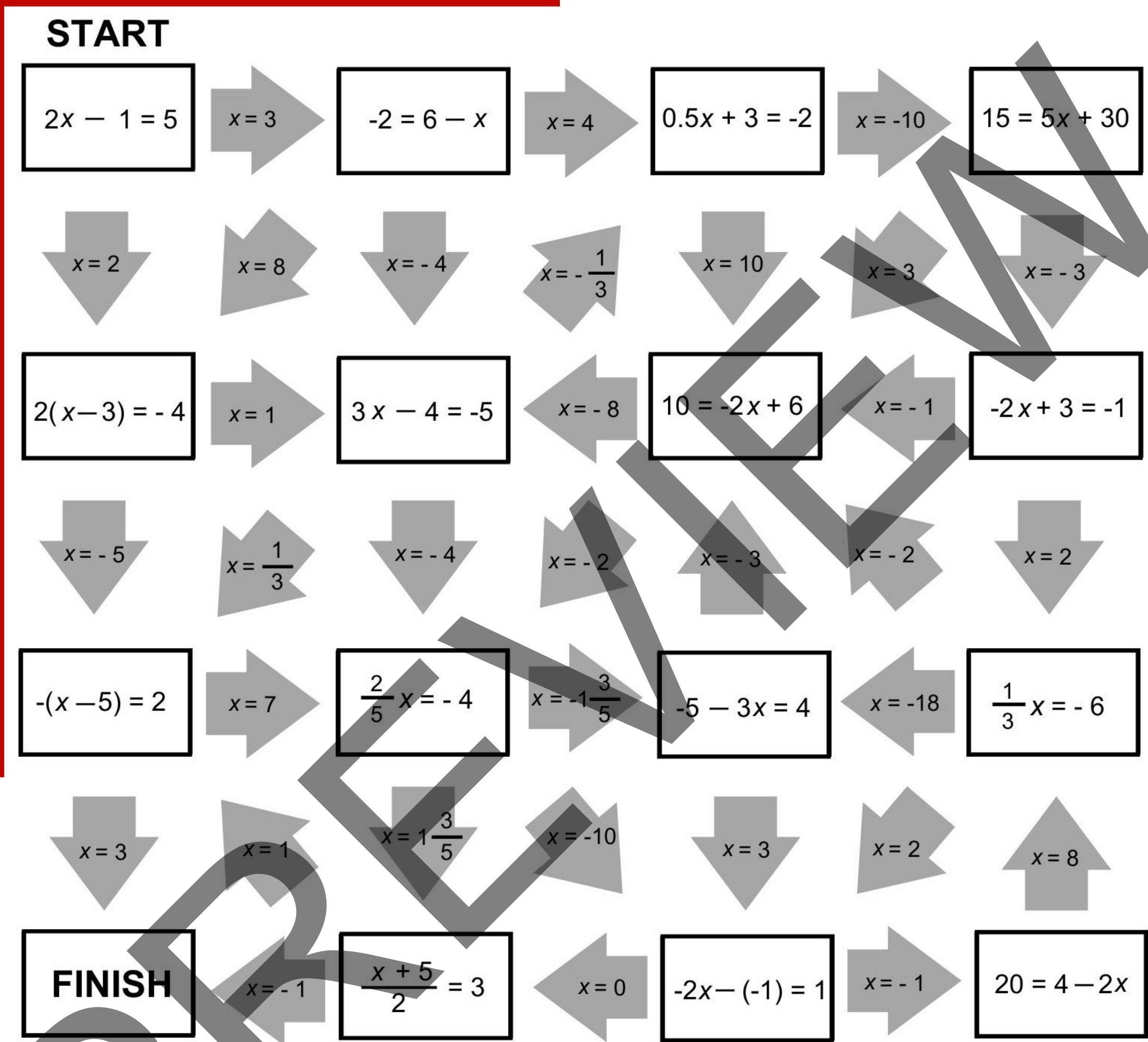
5 Mean, median, and mode are measures of ___.

6 sample where every object has equal chance to be selected

8 Another name for dot plot (two words)

SPIRAL REVIEW

1. **Math Path Fluency Challenge**: Use what you know about solving equations to find the correct path from Start to Finish.

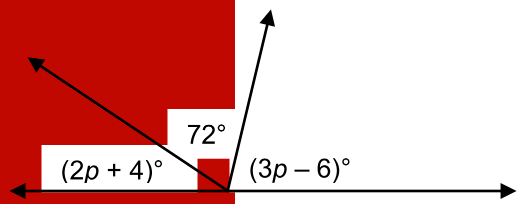


2. Complete the table below. Round to the nearest cent.

	1%	5%	4%	120%	0.5%
	$1.50				
		$4.80			
$28.25					

SPIRAL REVIEW
Continued

3. Find the value of p and the measure of each angle in the diagram below.



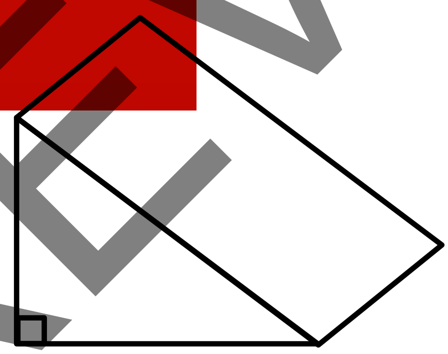
4. Donavon wants to build a kicker ramp based on the sketch below. Donavon wants to use a scale of 1 cm : 0.75 ft.

    a. What is the shape of the kicker ramp?

    b. Measure the drawing below and label side lengths.

    c. Write the actual dimensions of the ramp.

    d. Find the amount of wood (round up to the nearest square foot) that will be needed to build the entire ramp (with five faces).



SPIRAL REVIEW
Continued

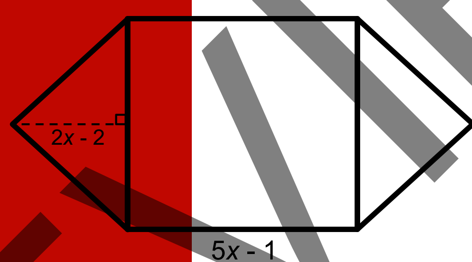
5. JP has $4\frac{3}{8}$ yards of wire. He cuts it into pieces that are each $\frac{1}{4}$ yard and uses each piece to tie up plants in the garden.

   a. How many pieces does he cut?

   b. How much wire will be left over?

   c. How much of a piece of wire will be left over?

6. The perimeter of the square below is 96 units. The height of each triangle is a third of the height of the side of the square. Find x and the total area.



7. Draw $\overline{BA}$ 6 cm long. Label it BA. From point A, draw $\overline{AD}$ so that it is 4 cm long and $\angle BAD$ formed is 25°. Finally, connect points B and D to create $\overline{BD}$.

   a. What shape did you create?

   b. What kind of triangle does it appear to be?

   c. What else can you conclude about triangle BAD from measuring?

REFLECTION

1. **Big Ideas**. Shade all circles that describe big ideas in this unit. Draw lines to show connections that you noticed.

- Find the likelihood of events with probability.
- Sample to understand populations with statistics.
- Apply proportional reasoning to ratios, rates, percent and scale.
- Solve problems involving measurements of geometric figures.
- Operate with rational numbers and solve problems.
- Develop spatial reasoning in two- and three-dimensions.
- Use algebra as a problem-solving tool.

Give an example from this unit of one of the connections above.

2. **Unit Progress.** Go back to **Monitor Your Progress** on the cover and complete or update your responses. Explain something you understand better now than before.

3. **Mathematical Practices.** Describe an instance in a problem where your computations were not precise and how you interpreted them [SMP 6]. Then circle one more SMP on the back of this packet that you think was addressed in this unit and be prepared to share an example.

4. **Making Connections.** Daniel Keys Moran said, “You can have data without information, but you cannot have information without data.” What do you think he meant?

STUDENT RESOURCES

Word or Phrase	Definition
experimental probability	In a repeated probability experiment, the experimental probability of an event is the number of times the event occurs divided by the number of trials. This is also called empirical probability. If, in 25 rolls of a number cube, we obtain an even number 11 times, we say that the experimental probability of rolling an even number is $\frac{11}{25} = 0.44 = 44\%$.
measure of center	A measure of center is a statistic describing the middle of a numerical data set. The mean, the median, and the mode are three commonly used measures of center. For the data set {3, 3, 5, 6, 6}, the mean (average) is $\frac{(3+3+5+6+6)}{5}$ = 4.6, and the median is 5. There are two modes, 3 and 6. Each of these numbers can be viewed as the "center" of the data set in some way.
measure of spread	A measure of spread (or a measure of variability) is a statistic describing the variability of a numerical data set. It describes how far the values in a data set are from the mean or median. The standard deviation (SD or σ), the mean absolute deviation (MAD), and the interquartile range (IQR) are three measures of spread.
population	The population is the entire group of individuals (objects or people) to which a statistical question refers. If a survey is taken to investigate how many pets the students at Seaside School own, the population under study is the entire student body of Seaside School.
sample	A sample is a subset of the population that is examined in order to make inferences about the entire population. The sample size is the number of elements in the sample. In order to estimate how many phones coming off the production line were defective, the plant manager randomly selected a sample of 50 phones and tested them to see if they worked properly.
simulation	Simulation is the imitation of one process by means of another process. We may simulate rolling a number cube by drawing a card blindfold from a group of six identical cards labeled one through six. We may simulate the weather by means of computer models.
theoretical probability	The theoretical probability of an event is a measure of the likelihood of the event occurring. In the probability experiment of rolling a (fair) number cube, there are six equally likely outcomes, each with probability $\frac{1}{6}$. Since the event of rolling an even number corresponds to 3 of the outcomes, the theoretical probability of rolling an even number is 3 out of 6, or $3 \bullet \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$.

Dot Plots (Line Plots)

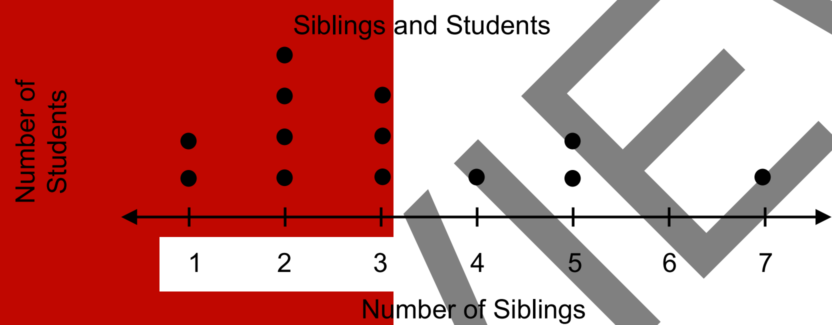
A dot plot (also called a line plot) displays data on a number line with a dot (•) or an X to show the frequency of data values.

Here are the number of siblings (brothers and sisters) for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To make a dot plot of this data set:

- Make a number line that extends from the minimum data value to the maximum data value
- Mark a dot or an X for every data value
- Write a title and add vertical and horizontal labels.



Measures of Center

Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the mean (average) of a data set, add all the values in the data set and divide the total by the number of values (number of observations, n).

Number of observations: $n = 13$
To find the mean: $3 + 4 + 5 + 2 + 2 + 3 + 3 + 2 + 2 + 5 + 7 + 1 + 1 = 40$
$40 \div 13 \approx 3.08$

To find the median (M), order the values from least to greatest and find the middle number. If there is an even number of values in the data set, the median is the mean (average) of the two middle numbers.

For the siblings data set: $\{1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 5, 5, 7\}$

median

To find the mode, find the value that occurs most often. (Some data sets may have more than one mode.)

For the siblings data set, the mode is 2. It is the value 2 occurs most often.

The Range, the Quartiles, and the Five-Number Summary

Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the range of a data set, find the difference between the greatest value and the least value in the data set.

For the siblings data set, the range is 6, since 7 – 1 = 6.

To find quartiles, first put the numbers in numerical order. Then locate the points that divide the data set into four equal parts.

1	1	2	2	2	2	3	3	3	4	5	5	7

min Q_1 $Q_2 = M$ median Q_3 *max*

For the siblings data set:
$Q_1 = 2$ (the 1st quartile)
$Q_2 = 3$ (the 2nd quartile; also the median)
$Q_3 = 4.5$ (the 3rd quartile)

Q_1 is the median of the first half of the data set, and Q_3 is the median of the second half.

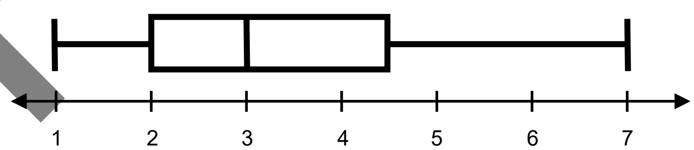
The <u>five-number summary</u> is (min, Q_1, Q_2, Q_3, max) = (1 , 2 , 3 , 4.5 , 7).

Box Plots (Box-and-Whisker Plots)

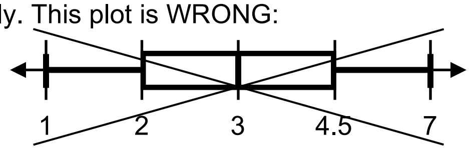
A box plot (or box-and-whisker plot) provides a visual representation of the center and spread of a data set. The display is based on the five-number summary.

For the sibling data, the five-number summary is (1 , 2 , 3 , 4.5 , 7).

To make a box plot:
- Locate the five-number summary values on a number line, and indicate each value with a vertical segment
- Create a "box" to highlight the interval from the first to the third quartile, and draw "whiskers" that extend to the minimum and maximum

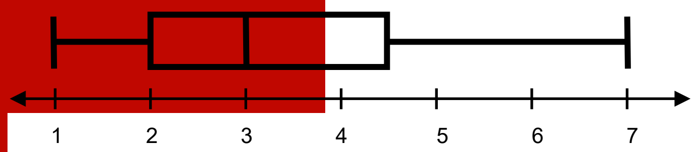


Be sure to scale the box plot properly. This plot is WRONG:



Interpreting a Box Plot

This is a box plot.



- Each of the four "sections" (the two whiskers and the two rectangular parts of the box) contains (close to) one-fourth of the data points. Be careful: If one section appears larger than another, we cannot say it has more data points, but only that the data points are spread out over a wider range.

- Sometimes we use the word "quartile" to refer to specific data points. Sometimes the word "quartile" is also used to refer to one of the four quarters, or sections, of the data set. For example, data points that lie within the farthest left section may be referred to as "in the first quartile."

Mean Absolute Deviation

The mean absolute deviation (MAD) is a measure of spread of a numerical data set. It is the arithmetic average of the distance (absolute value) of each data point to the mean. To calculate the MAD statistic:

For the sibling data, there are 13 data points:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the MAD statistic:

- Find the mean of the sample.

  The mean is 3.08.

- Find the distance (absolute value) from each data point to the mean.

  See the table entries to the right.

- Find the sum of the distances.

  See the bottom row of the table.

- Divide the sum of the distances by the number of data points to find the average distance from the mean.

  See the calculation below.

Sibling Data	**Distance from data point to mean**
3	$\lvert 3.08 - 3 \rvert = 0.08$
4	$\lvert 3.08 - 4 \rvert = 0.92$
5	$\lvert 3.08 - 5 \rvert = 1.92$
2	$\lvert 3.08 - 2 \rvert = 1.08$
2	$\lvert 3.08 - 2 \rvert = 1.08$
3	$\lvert 3.08 - 3 \rvert = 0.08$
3	$\lvert 3.08 - 3 \rvert = 0.08$
2	$\lvert 3.08 - 2 \rvert = 1.08$
2	$\lvert 3.08 - 2 \rvert = 1.08$
5	$\lvert 3.08 - 5 \rvert = 1.92$
7	$\lvert 3.08 - 7 \rvert = 3.92$
1	$\lvert 3.08 - 1 \rvert = 2.08$
1	$\lvert 3.08 - 1 \rvert = 2.08$
Sum of distances from mean	17.4

$$\text{MAD} = \frac{\text{sum of distances from mean}}{\text{number of data points}} = \frac{17.4}{13} = 1.34$$

Sampling
Sampling refers to selecting a subset of a population to be examined for the purpose of drawing statistical inferences about the entire population. If the sample is representative of the entire population, we may make valid inferences about the entire population based on properties of the sample. Suppose you want to know how many hours per week students in our school spend watching television. From the population of all students, you select a sample and you ask the students in the sample how many hours they watch television. You would like to infer that the average time spent watching TV for all students is about the same as for students in the sample. An easy way to select a sample might be to ask your friends how many hours they watch TV. Such a sample is called a convenience sample. However, your friends may not be representative of all students. To select a more representative sample, you might assign a number to the name of each student in the school, and then use a computerized "random number generator" to select a certain number of the students' numbers. This sort of sample is referred to as a random sample. Its mathematical properties allow us to draw inferences about the population.

The Percent Error Formula

A percent error computation is used to quantify the difference between an estimated value and an actual value as a percentage of the actual value. The formula for percent error (as a percent) is:

$$\text{Percent Error}_{(\text{as a percent})} = \frac{|\text{actual} - \text{estimate}|}{\text{actual}}, \text{ written as a percent}$$

If the estimated number of fish in a lake is 45, and the actual number is 50, then the percent error is

$$\frac{|50 - 45|}{50} = \frac{|5|}{50} = \frac{1}{10} = 10\%$$

If the estimated number of fish in a lake is 50, and the actual number is 45, then the percent error is

$$\frac{|45 - 50|}{50} = \frac{|-5|}{50} = \frac{1}{10} = 10\%$$

COMMON CORE STATE STANDARDS

STANDARDS FOR MATHEMATICAL CONTENT	
7.RP.A	**Analyze proportional relationships and use them to solve real-world and mathematical problems.**
7.RP.3	Use proportional relationships to solve multistep ratio and percent problems. *Examples: simple interest, tax, markups and markdowns, gratuities and commissions, fees, percent increase and decrease, percent error.*
7.EE.A	**Use properties of operations to generate equivalent expressions.**
7.EE.3	Solve multi-step real-life and mathematical problems posed with positive and negative rational numbers in any form (whole numbers, fractions, and decimals), using tools strategically. Apply properties of operations to calculate with numbers in any form; convert between forms as appropriate; and assess the reasonableness of answers using mental computation and estimation strategies. *For example: If a woman making $25 an hour gets a 10% raise, she will make an additional 1/10 of her salary an hour, or $2.50, for a new salary of $27.50. If you want to place a towel bar 9 3/4 inches long in the center of a door that is 27 1/2 inches wide, you will need to place the bar about 9 inches from each edge; this estimate can be used as a check on the exact computation.*
7.SP.A	**Use random sampling to draw inferences about a population.**
7.SP.1	Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
7.SP.2	Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.
7.SP.B	**Draw informal comparative inferences about two populations.**
7.SP.3	Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviatfion) on either team; on a dot plot, the separation between the two distributions of heights is noticeable.*
7.SP.4	Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*
7.SP.C	**Investigate chance processes and develop, use, and evaluate probability models.**
7.SP.5	Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.
7.SP.7	Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy:
a.	Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. For example, if a student is selected at random from a class, find the probability that Jane will be selected and the probability that a girl will be selected.
7.SP.8	Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation:
c.	Design and use a simulation to generate frequencies for compound events. For example, use random digits as a simulation tol to approximate the answer to the question: *If 40% of donors have type A blood, what is the probability that it will take at least 4 donors to find one type A blood?*

STANDARDS FOR MATHEMATICAL PRACTICE	
SMP1	Make sense of problems and persevere in solving them.
SMP2	Reason abstractly and quantitatively.
SMP3	Construct viable arguments and critique the reasoning of others.
SMP4	Model with mathematics.
SMP5	Use appropriate tools strategically.
SMP6	Attend to precision.
SMP7	Look for and make use of structure.
SMP8	Look for and express regularity in repeated reasoning.

9781614454342