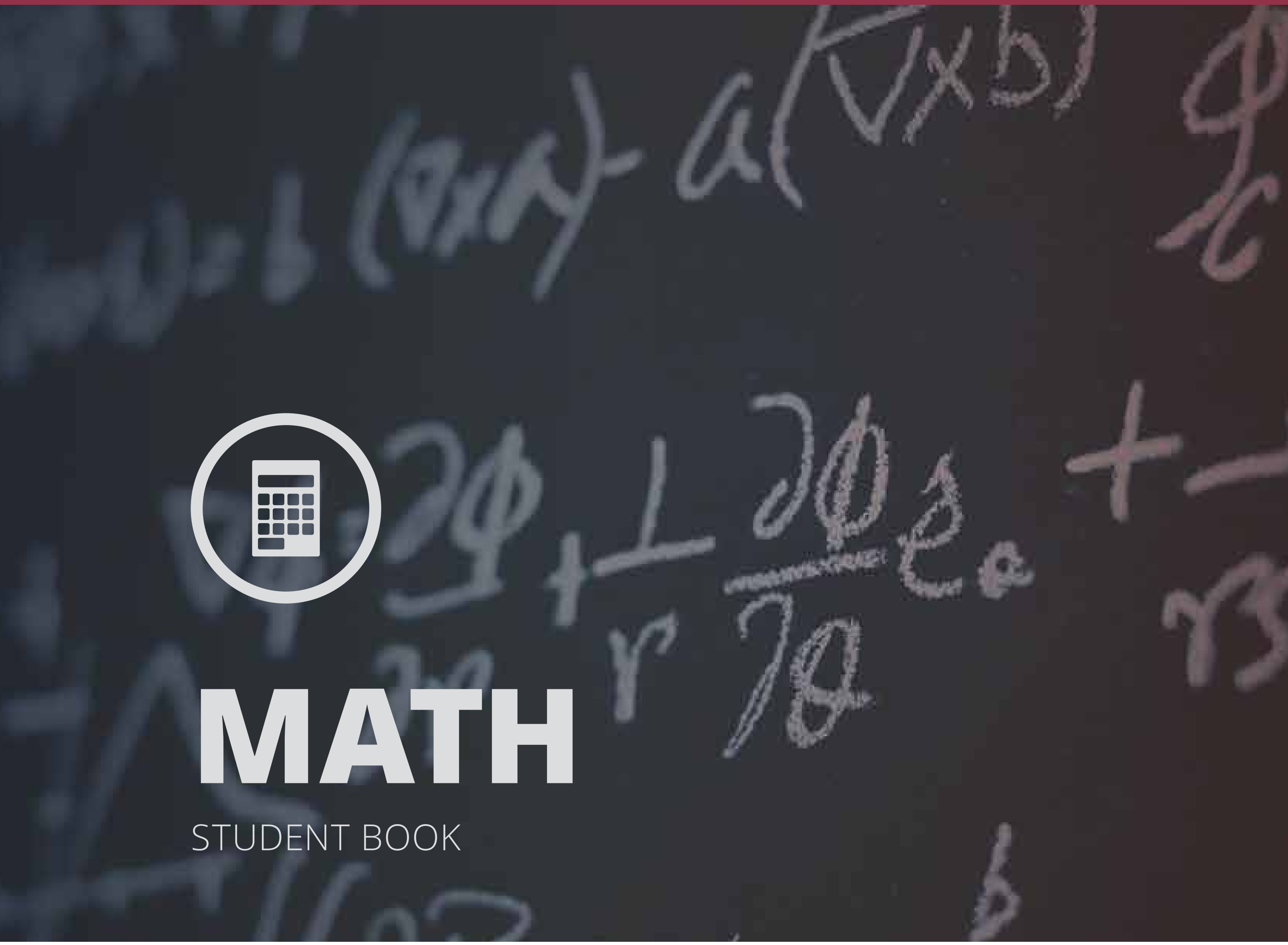
LIFEPAC®



MATH

STUDENT BOOK

7th Grade | Unit 7

Alpha Omega PUBLICATIONS

Math 707
Data Analysis

Introduction |3

1. Describing Data 5

Collecting Data |**5**
Determining Mean, Median, and Mode |**11**
Using Mean, Median, and Mode |**16**
Using Range |**23**
Self Test 1: Describing Data |**30**

2. Organizing Data 33

Box-and-Whisker Plots |**33**
Stem-and-Leaf Plots |**40**
Histograms |**49**
Other Graphs |**58**
Self Test 2: Organizing Data |**66**

3. Graphing Data and Making Predictions 71

Line Graphs |**71**
Bar Graphs |**81**
Circle Graphs |**89**
Scatter Plots |**97**
Self Test 3: Graphing Data and Making Predictions |**105**

4. Review 111

LIFEPAC Test is located in the center of the booklet. Please remove before starting the unit.

Author:
Glynlyon Staff

Editors:
Alan Christopherson, M.S.
Michelle Chittam

Westover Studios Design Team:
Phillip Pettet, Creative Lead
Teresa Davis, DTP Lead
Nick Castro
Andi Graham
Jerry Wingo

Alpha Omega PUBLICATIONS

804 N. 2nd Ave. E.
Rock Rapids, IA 51246-1759

Data Analysis

Introduction

In this unit, students will study statistics through organizing, analyzing, interpreting, and displaying numerical and categorical data. The types of graphs they will look at include box-and-whisker plots, stem-and-leaf plots, histograms, Venn diagrams, pictographs, bar graphs, line graphs, circle graphs, and scatter plots. Also included is a discussion of how statistics can be biased or misleading.

Objectives

Read these objectives. The objectives tell you what you will be able to do when you have successfully completed this LIFEPAC. When you have finished this LIFEPAC, you should be able to:

- Determine whether a sample is biased or random.
- Determine whether a question is biased or unbiased.
- Make predictions from a random sample, line graph, or scatter plot.
- Define and find the measures of central tendency and dispersion.
- Construct, interpret, and use the following graphs: box-and-whisker plots, stem-and-leaf plots, histograms, pictographs, line graphs, bar graphs, circle graphs, and scatter plots.
- Use Venn diagrams to solve problems.

Survey the LIFEPAC. Ask yourself some questions about this study and write your questions here.

1. Describing Data

Collecting Data

In this lesson, you'll be exploring how to collect information, which is just one aspect of *statistics*. You'll look at how to choose good questions, how to get accurate results from your questioning, and how to use those results to make predictions.

Objectives

- Determine whether a sample is biased or random.
- Determine whether a question is biased or unbiased.
- Make predictions from a sample.

Vocabulary

biased question—a question that leads individuals towards a certain answer

biased sample—a sample not representative of the entire population

population—the group of individuals or items from which samples are taken

random sample—a sample in which every member of the population has an equal chance of being selected; unbiased sample

sample—a small part of a population chosen to represent the entire group

statistics—the collection, organization, and analysis of numerical information

survey—a sampling of a population used to make predictions

Have you ever taken a *survey* and felt like every question was leading you towards a specific answer? And that you'd be a fool to not pick the "correct" answer? For example, in the cartoon, the way Ondi words her question to Malik leaves you knowing exactly which option she thinks is better!

Reminder! A survey is a process of questioning a group in order to make generalizations and predictions.

This type of question is called a *biased question*. If a question is biased, then it seems to have a right or wrong answer. When asking people for their opinions, to get honest answers, there shouldn't be one correct answer. Here are some more examples of biased questions:

Examples:

- On a warm, sunny day, would you rather be outside enjoying the beautiful weather or just sitting inside reading a book?
- Are you happy with the ridiculous way in which your city mayor has been doing his job?
- In each of the above examples, you know exactly what answer the questioner is looking for—or at least which answer the questioner would choose. An unbiased question should give no indication as to what the surveyor's answer would be. Look at how these questions could be phrased differently in order to become unbiased questions.

Think about it! If the surveyor is questioning people in person, then tone of voice also plays a role in whether the question is biased or unbiased.

Examples:

- On a warm, sunny day, would you rather be outside playing or inside reading a book?
- Are you happy with how your city mayor has been doing his job?

Questions aren't the only way in which a survey can be biased. Another factor that affects survey results is *who* is being questioned. For example, suppose you want to find out how kids in your school would rather spend a sunny afternoon—inside reading or outside playing. If you go to a large school, it wouldn't be very easy to ask every student in the school. Instead, you would want to ask a small group that could represent all the kids in your school.

A *sample* is a small group that is used to represent the opinions of the entire group. The entire group is called the *population*. In this case, the population is all the kids in your school. The sample is the group used to represent the kids in your school. Choosing a sample is very important in getting accurate results from a survey.

For example, it wouldn't be a good idea to only ask the kids already playing outside whether they would prefer to play outside or read inside. Those kids have already made the choice of playing outside, so they wouldn't provide a good representation of what all the kids would choose. This group of kids would be a *biased sample* because it doesn't accurately represent the entire population.

A better representation would be to choose a group of students randomly. In a *random sample*, every member of the population has an equal chance of being in the sample. For example, a random sample could be chosen by putting all the student's names into a box and drawing out twenty names.

Example:

- A school is planning clubs for the upcoming school year. The teachers surveyed the boys in each class to determine what their interests are. Was this a well-chosen sample?

Solution:

- If the school were all boys, there would not be a problem with just surveying boys. If you surveyed all of the boys in every class, you would not be taking a sample, but gathering information from the whole group. If there are girls in the school, this survey would exclude their interests and potentially eliminate clubs the girls would enjoy. This would not be an appropriate sample.

Make note! A larger sample will yield more accurate results than a smaller one. So surveying twenty students would be better than surveying only five students. However, if the sample is too large, then sampling isn't an efficient way of gathering information anymore.

You may be wondering why sampling is helpful or necessary. Results from surveying a sample can be used to make predictions about the entire population. Many businesses and manufacturers use sampling in order to predict how well a product will do or to determine which direction their company should go. Sampling can also be used to keep costs down. Here's an example of making predictions about an entire population.

Think back to Ondi's question to Malik. Suppose your school wants to have an ice cream party on the last day of school. They'll be serving vanilla and chocolate ice cream to all students. In order to predict how much of each kind of ice cream to buy, they could poll a random sample of the student body. For example, suppose the results of the survey show that 12 out of 20 students prefer vanilla ice cream over chocolate. If there are 425 students in the student body, you could set up a proportion to predict how many total students will prefer vanilla:

$$\frac{\text{students that prefer vanilla}}{\text{total students}} = \frac{12}{20} = \frac{v}{425}$$

Keep in mind! Making predictions does not guarantee the results. However, it's helpful in making good estimates about what will happen in the future.

Remember that to solve proportions, you can use cross multiplication. Take a look:

- $(12)(425) = 20v$
- $5{,}100 = 20v$
- $\frac{5{,}100}{20} = \frac{20v}{20}$
- $255 = v$

From the results of the poll, you could expect that approximately 255 of the students would prefer vanilla ice cream.

Example:

- Susie surveyed several randomly selected students at school to find out whether they prefer walking, running, swimming, or biking. The results of her survey showed that 8 students prefer walking, 11 prefer running, 14 prefer swimming, and 7 prefer biking. What fractional part of the sample prefers to swim?

Solution:

- First, you need to find the total number of students in the sample. To do that, add the students from each of the four groups:
  - $8 + 11 + 14 + 7 = 40$

- So there are 40 students in Susie's sample and 14 of them prefer to swim. As a fraction, $\frac{14}{40}$ or $\frac{7}{20}$ would rather swim.

S-T-R-E-T-C-H! Can you think of another way to express this fractional part? Remember that a fraction can be converted to a percent by dividing the numerator by the denominator. In this case, 7 divided by 20 is 0.35, so 35% of the population prefers to swim.

Let's Review

Before going on to the practice problems, make sure you understand the main points of this lesson:

- Sampling is used to make generalizations about a population.
- Biased questions and samples can affect the accuracy of a survey's results.
- Proportions are used to apply the results of a sample in making numerical predictions about a population.

Complete the following activities.

1.1 Any sample from a group of people will be representative of the entire group.
- ○ True
- ○ False

1.2 A biased sample is one in which every member of the group has an equal chance of being chosen.
- ○ True
- ○ False

1.3 Biased questioning will probably result in inaccurate survey results.
- ○ True
- ○ False

1.4 Which of the following questions represents an unbiased question?
- ☐ Many students have said that they prefer to bring a sack lunch rather than eat the school lunch. Do you agree?
- ☐ Would you prefer to bring a sack lunch or eat the school lunch?
- ☐ Which would you prefer: a sack lunch from home or the tasty school lunch?

1.5 Which of the following situations represents choosing a random sample?

- ☐ Assign each person of the population a number. Put all the numbers into a bowl and choose ten numbers.
- ☐ Make a list of everyone in the population and put their names in alphabetical order. Choose every twelfth name.
- ☐ Make a list of everyone in the population and put their names in order of age. Choose the ten youngest and ten oldest people.

1.6 A certain population has 1,000 people in it. Which of the following numbers would be an appropriate number for a random sample?

☐ 1,000 ☐ 900 ☐ 200 ☐ 10

Use this info to complete activities 1.7-1.8. A random sample about people's favorite primary color yielded the following results.

Color	Number
blue	9
yellow	2
red	4

1.7 Find the percentage of people in the sample who prefer blue.

☐ 67% ☐ 27% ☐ 13% ☐ 60%

1.8 If there are 90 people in the population, how many would you expect to prefer yellow?

☐ 12 ☐ 24 ☐ 30 ☐ 36

1.9 Write your own example of a biased question.

A church youth group took a survey to determine what food the students in the youth group wanted to eat. After surveying 25 students, they discovered that 14 wanted pizza, 5 wanted hamburgers, 3 wanted hotdogs, and 3 wanted chicken sandwiches.

1.10 If there are 100 students in the youth group, how many would you expect to want pizza?

1.11 If you served pizza and hamburgers to a group of 150 students, how many students would not have the food they wanted as a choice?

1.12 How many students from a group of 75 would you expect to want hotdogs?

1.13 If you are planning to serve only chicken sandwiches to a group of 500, how many students would have the food they chose?

1.14 If you are serving hamburgers and hotdogs to a group of 350 students, how many students would not have their first choice of food as an option?

Determining Mean, Median, and Mode

Carlton's right! The *mean* of a group of numbers is often called their average. This is just one way to describe *data*, though. You're also going to look at two other ways to describe a group of numbers, or *numerical data*: the *median* and the *mode*. The mean, median, and mode are all measures of *central tendency*. They're called this because they describe the "center" of the data.

Vocabulary! Data is simply information. In statistics, information is often expressed as numbers, or quantities. Then it's called numerical data.

Objectives

- Determine the mean, median, and mode of a set of data.

Vocabulary

bimodal—having two modes

central tendency—ways to describe or summarize data

data—information (often numerical)

mean—the sum of a set of data divided by the number of items in the set

median—the middle value of a set of data arranged in numerical order

mode—the most frequently occurring number(s) in a set of data

numerical data—data represented by quantities

Mean

The mean is probably the most commonly used measure of central tendency. It is found by adding all of the numbers in the set and then dividing by the number of items in the set:

- $\text{mean} = \frac{\text{sum of numbers}}{\text{number of items}}$

For example, suppose all of Ondi's homework assignments were worth ten points. She's turned in nine assignments and received the following scores for them:

- 10, 7, 9, 9, 8, 10, 4, 9, 7

To find the mean of her assignment scores, simply add up the nine scores and then divide that sum by nine:

$$\frac{10+7+9+9+8+10+4+9+7}{9} = \frac{73}{9} = 8.\overline{1}$$

So the mean of Ondi's scores is about 8.1.

Remember! The mean is the sum of a set of data divided by the number of items in the set.

Median

The median is the measure of central tendency that tells you what the middle value of the data is. The best way to find the median is to line up the data from the smallest value to the largest value. Then find the value that cuts the data into two equal parts. Try finding the median of Ondi's homework scores.

This might help! To help you remember what the median of a set of data is, remember that the median in a street is in the middle of the road, dividing the lanes of traffic.

Remember, Ondi's quiz scores were as follows:

- 10, 7, 9, 9, 8, 10, 4, 9, 7

Line up the data from smallest to largest. Make sure that every value in the list is accounted for. You may want to count the numbers in the original set and the ordered set in order to check that you have the same number of items in each:

- 4, 7, 7, 8, 9, 9, 9, 10, 10

Now find the middle number in the list. You may find it helpful to cross off a number on each side of the list until you get to the middle.

So the median of Ondi's homework scores is 9.

Remember! The median is the middle value of a set of data or the value that cuts the data into two parts.

When there is an odd number of items in the set of data, there will only be one value in the middle of the list. That middle value is the median. When there is an even number of items in the set of data, there will be two values in the middle of the list. There can't be two medians in a set of data, so in this case, the mean of the two middle numbers is the median. Look at this next example.

Example:

- Find the median of 120, 142, 83, 211, 187, and 99.

Solution:

- First, put the numbers in numerical order from smallest to largest, making sure that all values in the set are accounted for:
  - 83, 99, 120, 142, 187, 211
- Then find the middle value. Since there is an even number of items in this set, the median will be the mean of the middle two numbers.
- The middle of the list is between 120 and 142. In order to find the median, find the mean of 120 and 142:

$$\frac{120+142}{2} + \frac{262}{2} = 131$$

- So the median of this set is 131.

Mode

The final measure of central tendency is the mode. The mode is the easiest of the three measures to find. It represents the value that occurs most frequently in the set of data. Just like with the median, the easiest way to find the mode is to put the data in order from smallest to largest. Find the mode of Ondi's homework scores. Since we already put them in order to find the median, you can use that same list:

- 4, 7, 7, 8, 9, 9, 9, 10, 10

Look for items that are repeated in the list. In this list, 7 occurs twice, 9 occurs three times, and 10 occurs twice. Since 9 occurs the most times, it is the mode of this set.

This might help! There is no mode when none of the numbers appears more frequently than the others. The set is bimodal, or has two modes, when there are two numbers that appear the same amount of times and more frequently than the others. Although rare, it's even possible to have three or more modes in a set.

Unlike the median, it is possible to have more than one mode. In fact, it's even possible to have no mode. When there are two modes, the list is said to be *bimodal*. Here are a couple of examples.

Examples:

- ▶ The set 83, 99, 120, 142, 187, 211 has no mode because there is no number that occurs more frequently than the others.
- ▶ The set 19, 21, 21, 24, 25, 26, 26 is bimodal because there are two numbers that appear more frequently than the others. The modes of this set are 21 and 26.
- ▶ The set 32, 32, 32, 38, 39, 40, 40, 41, 41, 41, 43, 44, 44, 45, 45, 45, has three modes. The numbers 32, 41, and 45 each appear three times, so the modes are 32, 41, and 45. Notice that 40 and 44 each appear twice. These are not part of the mode because there are numbers that appear more often.

Remember: The mode is the number that occurs most frequently in a set of data.

The measures of central tendency for Ondi's homework scores are as follows: The mean is 8.1, and both the median and mode are 9. Either of these numbers can be used to describe her scores. In many sets of data, the three measures of central tendency each yield a different value. So make sure to be specific about which measure of central tendency you're using to describe a set!

Let's Review

Before going on to the practice problems, make sure you understand the main points of this lesson:

- The three measures of central tendency are the mean, median, and mode.
- The mean of a set of data is found by dividing the sum of the data by the number of items in the set.
- The median is the middle value of a set of data after the set has been listed from smallest to largest.
- The mode is the value that occurs most frequently in a set of data.

Complete the following activities.

1.15 The mean, the median, and the mode of a set of numbers can sometimes be equal to each other.

- ○ True
- ○ False

Use this information to complete activities 1.16-1.18. Linda received the following scores on her essay tests. 64, 65, 63, 66, 60, 65, 66, 68, 69, 66, 63

1.16 What is the median of her scores?

☐ 64.5 ☐ 65 ☐ 65.5 ☐ 66

1.17 What is the mean of her scores?

☐ 64.5 ☐ 65 ☐ 65.5 ☐ 66

1.18 How many modes does this set of scores have?

☐ none ☐ one ☐ two ☐ three

1.19 Which of the following lists has a mode of 213?

☐ 111, 108, 213, 198, 205 ☐ 213, 278, 108, 213, 157

☐ 212, 215, 213, 211, 220 ☐ 210, 200, 213, 221, 221

1.20 A class of 25 students is trying to find the mean weight of the students in the class. If the sum of all the students' weights is 2,946 pounds, what is the mean weight rounded to the nearest pound?

☐ 110 pounds ☐ 117 pounds ☐ 118 pounds ☐ 125 pounds

1.21 What is the median of the following set of numbers?
121, 347, 45, 21, 300, 614, 312, 333, 421

☐ 312 ☐ 306 ☐ 322.5 ☐ 300

1.22 What is the mean of the following set of numbers?
121, 347, 45, 21, 300, 614, 312, 333, 421

☐ 312 ☐ no mean ☐ 251.4 ☐ 279.3

1.23 What is the mode of the following set of numbers?
121, 347, 45, 21, 300, 614, 312, 333, 421

☐ no mode ☐ 279.3 ☐ 312 ☐ 21

1.24 What is the median of the following set of numbers?
15, 12, 18, 17, 13, 12, 12, 14, 15, 19

- ☐ 12
- ☐ 14
- ☐ 15
- ☐ 14.5

1.25 Select all the statements that describe the following set of numbers.
3, 9, 8, 6, 3, 4, 9, 2, 5, 10, 8, 1

- ☐ This set has three modes.
- ☐ The mode is 3.
- ☐ The median is the mean of 5 and 6.
- ☐ The mean is approximately 5.67.
- ☐ The mean is smaller than the median.

1.26 What is the mean of the following set of numbers? 19, 22, 25, 28, 31

1.27 What is the median of the following set of numbers? 19, 22, 25, 28, 31

1.28 What is the mode of the following set of numbers? 19, 22, 25, 28, 31

1.29 What is the median of the following set of numbers? 18, 32, 26, 14, 15, 31

1.30 What is the mode of the following set of numbers? 54, 61, 61, 32, 14, 27

Using Mean, Median, and Mode

Mia is looking for a summer job, and she's excited about an ad that she just read in the paper. A landscaping company is looking for ten teenagers to help them over the summer with their clients' yard maintenance. The best part is that the ad says the mean wage in the company is $12 per hour! When Mia tells her mom about the ad, though, her mom isn't nearly as excited about the pay as Mia is. Can you think of a reason why Mia's mom isn't as excited?

In this lesson, you're going to look at how the measures of central tendency can sometimes be misleading or draw you to make the wrong conclusions. You'll see how the mean, median, and mode are used in real life and how extreme values can affect these measures.

Objectives

- Determine the effect of an outlier on an average.
- Determine which measure of central tendency should be used in a situation.
- Use the mean to find a missing value.

VOCABULARY

outlier—a value that is far removed from the rest of the values in a set of data

Representing Data

So why isn't Mia's mom excited about the pay for the landscaping job? Most parents would be happy at the prospect of their teenager making $12 an hour. The reason is that Mia's mom knows that statistics can often be misleading. Remember that any of the three measures of central tendency can be used to describe the "center" of the data. Businesses often use the measure that makes their data look the best.

Mia automatically assumed that if the mean wage is $12 an hour, she'll probably be making that much, too. Unfortunately, the following data shows that Mia's mom is correct. Figure 1 is a table showing the wages of all the positions at this landscaping company.

Position	#of Employees	Wage $ per hour
lawn maintenance (summer)	12	$5
administrative assistant	1	$18
general labor (all year)	4	$21
supervisor/owner	1	$54

Figure 1| Landscaping Company Wages

Now use the data to figure out all three of the measures of central tendency. First, list the data in numerical order from smallest to largest. Notice that there are twelve employees who make $5 and four employees who make $21, so each of those dollar amounts will need to be listed multiple times:

Keep in mind! Do you remember what each measure of central tendency represents? The mean is the sum of the values in the set divided by the number of items in the set. The median is the middle value in the set after it has been listed in numerical order. And the mode is the value that is represented the most in the set. Remember that it is possible to have more than one mode in a set.

- 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 18, 21, 21, 21, 21, 54

To find the mean of a set, you need to add all the values together and then divide that sum by the number of values. In this case, add all the numbers from the list above and divide by 18. There are a lot of numbers in this set, so make sure you account for each value!

- $\frac{5+5+5+5+5+5+5+5+5+5+5+5+18+21+21+21+21+54}{18} = \frac{216}{18} = 12$

So the mean wage at this landscaping company *is* $12 per hour. The ad wasn't lying. However, when you get to the other two measures, you'll see a different story. Find the median next.

Recall that the median of a set of data is the middle value. In this case, there are 18 items in the list. Since the amount is even, there will be two numbers in the middle of the set. You'll need to find the mean of those two numbers.

5, 5, 5, 5, 5, 5, 5, 5, 5, | 5, 5, 5, 18, 21, 21, 21, 21, 54

nine numbers | nine numbers

So the median, or middle, of this set lies between two 5's. The mean of 5 and 5 is 5 because $\frac{5+5}{2} = \frac{10}{2} = 5$. Now find the mode.

The mode is the easiest value to find. It is the value that occurs the most in the set of data. In this case, the value of 5 occurs twelve times, so the mode is 5.

So you have the results from this set of data. The mean is $12, and the median and mode are both $5. How did this happen? How can one measure of central tendency be so different from the other two? It's because this set of data contains an *outlier*.

An outlier is a value in the data that is far removed from the rest of the data. In this case, the amount of money that the supervisor makes, or $54, is far higher than the amount that everyone else in the company makes. The value of 54 is the outlier in this set of data. How do outliers affect the measures of central tendency?

This might help! Many sets of data don't contain an outlier. An outlier can either be an extremely high value or an extremely low value compared to the other values in the data.

The only measure that is directly affected by an outlier is the mean. That's because only the mean takes into account the exact value of every item in the set. The mode is rarely affected by outliers because it represents the value that occurs the most, and most sets don't have repeated outliers. The median is really only affected by what is going on in the middle of the data and not by what is going on at the ends of the data.

The landscape problem is a good example of how outliers affect the measures of central tendency. The outlier (the supervisor's wage) dramatically affected the mean of the data. However it had no affect on the values of the median and the mode. To prove this, remove the supervisor's wage and recalculate the mean, median, and mode of this set. Notice that there are only 17 numbers in the set now. Figure 2 shows the calculation of the mean, median, and mode for this example.

Notice that removing the outlier only changed the value of the mean. The median and the mode were not affected by it in the first place.

Key point! Of the three measures of central tendency, the mean is usually the most affected by outliers. In the previous example, it was the only measure affected. This isn't always the case, though. Depending on the data, the median or mode may be affected as well.

So why did the landscaping company choose to use the mean wage in its ad when it wasn't the most accurate representation of the data? Because it was the measure that was most likely to get people to respond to the ad. Of the three measures, it made the company look the most enticing to applicants. If the ad had said something like, "Most employees at our company make about $5 an hour," or "The median wage at our company is $5 an hour," many people probably wouldn't have given the ad a second glance. The company used the description that would best enable it to achieve its goal—lots of excited applicants!

Although the landscaping company didn't lie about its data, it did use misleading information. The company chose the

Set of Data with Outlier removed. 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 18, 21, 21, 21, 21

Mean. $\frac{5+5+5+5+5+5+5+5+5+5+5+5+18+21+21+21+21}{17} = \frac{162}{17} = 9.53$

Median. $\underbrace{5, 5, 5, 5, 5, 5, 5, 5,}_{\text{eight numbers}}\ \textcircled{5}\ \underbrace{5, 5, 5, 18, 21, 21, 21, 21}_{\text{eight numbers}}$

Mode. The value with the highest frequency is still 5.

Figure 2 | Landscape Problem with Outlier Remove measure that best suited its own goals, even though it didn't accurately represent the wages in the company. A better representation would have been either the median or the mode of the data. This is the tricky thing about statistics: They can be misleading! You must carefully analyze each situation, or set of numbers, to determine the most accurate way to describe or represent it.

Example:

- A school has 6 students in its senior class. The SAT scores for those 6 students are 790, 860, 860, 910, 1120, and 1250. If the school wants the SAT scores to appear as high as possible, should the school report the mean, the median, or the mode?

Solution:

- Calculate each of the measures of central tendency and compare the results. Figure 3 shows the calculations for this solution.
- Out of these three numbers, 965 is the highest. The school should report the mean SAT score in this case.

Think about it! Even though a measure may be calculated correctly, it may not accurately represent the data. Also, the context of the situation or the goals of the presenter help determine which measure of central tendency will be the most effective.

Finding Missing Values

So far, you've used measures of central tendency to represent or describe a set of given numbers, but there's another way you can use these measures. You can also use a desired mean to find a missing part of the set.

For example, Jackson has taken four math tests so far. His scores have been 79, 88, 91, and 85. In order to keep his after school job, Jackson's parents have said that he needs to maintain an average of at least 85. Has Jackson maintained a high enough grade so far? Check by finding the mean:

- $\frac{79 + 88 + 91 + 85}{4} = \frac{343}{4} = 85.75$

Yep! His grade is on the edge, but he's still above an 85. Jackson has a test coming up at the end of the week. He wants to know how well he has to do in order to keep his job. How can you figure that out?

To find his new mean, you would have to add up the now *five* test scores in order to find the sum of the scores. If you divide

Mean. $\frac{790+860+860+910+1120+1250}{6} = \frac{5790}{6} = 965$

Median. $\frac{860+910}{2} = \frac{1770}{2} = 885$

Mode. 860.

Figure 3 |SAT Problem

that total by 5, you want the score to be at least 85. Here's what that looks like in mathematical form:

- $\frac{\text{Sum of five scores}}{5} \geq 85$

Reminder! The phrase "at least" can be expressed as an inequality. It translates to the "greater than or equal to" symbol, or ≥.

The part you don't know, because of the unknown test score, is the sum of the five scores. Using your skills to solve algebraic equations, you want to get that sum by itself. The sum is currently divided by 5. So to get it by itself, multiply both sides of the inequality by 5 and simplify:

- $5 \cdot \frac{\text{Sum of five scores}}{5} \geq 85 \cdot 5$

  sum of five scores ≥ 425

In the equation above, replace the words "sum of five scores" with the values of the five scores. Use *x* to represent the score of the upcoming test since you don't know it yet:

$79 + 88 + 91 + 85 + x \geq 425$	Substitute in the five values.
$343 + x \geq 425$	Add the four known scores together.
$343 - 343 + x \geq 425 - 343$	Subtract 343 from each side of the equation.
$x \geq 82$	Simplify.

So Jackson needs to score at least an 82 in order to maintain a mean score of 85.

Example:

- Aaron's P.E. class is completing a unit on bowling. So every Friday afternoon for the last three weeks, they have bowled one game at a nearby bowling alley. His scores have been 121, 147, and 96. What would he have to bowl next time in order to have a mean score of 128?

Solution:

- Let *x* represent Aaron's fourth and unknown score. To have a mean of 128, the sum of the four scores would have to be divided by 4. To get that sum by itself, multiply it by 4. You also have to multiply the desired mean of 128 by 4. So the sum of the four scores is equal to 128 times 4.

$121 + 147 + 96 + x = 128 \cdot 4$	Substitute in the values of the four scores.
$364 + x = 512$	Simplify.
$364 - 364 + x = 512 - 364$	Subtract 364 from each side of the equation.
$x = 148$	Simplify.

- Aaron must score a 148 in order to have a mean score of 128.

Let's Review

Before going on to the practice problems, make sure you understand the main points of this lesson:

- Outliers are values in a data set that are far removed from the rest of the data.
- The mean is most affected by outliers.
- The most effective measure of central tendency for describing a data set is dependent on the situation and the goals of the person presenting the information.
- The mean can be used to find an unknown value in a set of data.

Complete the following activities.

1.31 Which of the following values would be an outlier in the following set?
1, 14, 17, 18, 19, 23

☐ 1 ☐ 14 ☐ 18 ☐ 23

1.32 What would be the mean if the value 3 was added to the following set?
14, 14, 17, 19, 25

☐ 17.8 ☐ 20.8 ☐ 18.4 ☐ 15.3

1.33 Find the mean, median, and mode of the following set. 27, 20, 34, 37, 21, 42, 39

☐ mean: 31.4; median: 35.5; mode: 37
☐ mean: 36.7; median: 34; no mode
☐ mean: 31.4; median: 34; no mode
☐ mean: 36.7; median: 37; no mode

1.34 Which measures of central tendency would be affected if the outlier 11 was added to the following set? 27, 20, 34, 37, 21, 42, 39

☐ mean only
☐ mean and median
☐ mean, median, and mode
☐ mean and mode

1.35 Aidan works a different number of hours each week, depending on whether he is needed or not. For the last nine weeks, he has worked the following number of hours. 1, 3, 1, 1, 4, 5, 8, 6, 5

Which measure of central tendency would be the most effective for Aidan to use if he wants to demonstrate to his supervisor that he hasn't worked very many hours recently?

☐ mean ☐ median ☐ mode

1.36 In the following set, which measure of central tendency would probably be the most accurate representation of the data? 11, 11, 18, 32, 34, 115

☐ mean ☐ median ☐ mode

1.37 On this week's homework assignments, Lydia had scores of 85, 78, 92, 98, and 85. Which measure of central tendency would give Lydia the highest score so she could show her teacher that she really understood her assignments?

☐ mean ☐ median ☐ mode

1.38 A local restaurant advertises that the mode cost of their most popular meals is $8. If the costs of their most popular meals are $7, $8, $8, $12, $13, $15, $17, $18, and $20, which word or phrase best describes this kind of advertising?

☐ accurate ☐ accurate, but misleading

☐ inaccurate ☐ inaccurate and misleading

1.39 A set of data already has the values 11, 14, 23, and 16. What value would have to be added to the set for the mean of the five numbers to be 18?

☐ 8 ☐ 12 ☐ 18 ☐ 26

1.40 Tasha's scores on her first five tests were as follows. 74, 90, 92, 83, 81 What would Tasha have to score on the sixth test to have a mean of 85?

☐ 5 ☐ 85 ☐ 90 ☐ 95

Use the following set of numbers to answer the questions below.

21, 22, 24, 25, 26, 26, 26, 27, 28, 45

1.41 What is the mean of the set of numbers?

1.42 Identify the outlier in above set of numbers. What is the mean of the remaining numbers?

1.43 Does removing the outlier change the median of the set of numbers?

1.44 Does removing the outlier change the mode of the set of numbers?

1.45 To have a more accurate representation of the data above, should you report the mean with the outlier or not? Why?

Using Range

What do you think of when you hear the word *range*? Maybe you think of a mountain range or a shooting range. Maybe you enjoy golf and think of a driving range. Or maybe what comes to mind is a range where cattle graze. The word *range* has lots of meanings and uses. In statistics, it has a special meaning that you'll explore in this lesson.

Objectives

- Find the range of a set of data.
- Determine the effect of outliers on the range and the interquartile range.
- Find the interquartile range of a set of data.

Vocabulary

dispersion—how data is distributed

interquartile range—the difference between the upper quartile and the lower quartile

lower quartile—median of the lower half of a set of data

quartile—division of data into four equal parts

range—the difference between the largest and smallest data points

upper quartile—median of the upper half of a set of data

So what does the range stand for in statistics? It will help if you think of the word *range* as a distance. The range of a set of data is the distance, or difference, between the highest value in the set and the lowest value in the set. It shows how far apart the data lie. This type of measure, which shows how the data are distributed, is called a measure of *dispersion*. Take a look.

This might help! Dispersion and central tendency measure two different characteristics of a set of data. However, both are important! A measure of dispersion looks at how far the data are spread out, or the outside values. A measure of central tendency looks more at the middle values of a set. But even though the two measure different characteristics, knowing how far the data are dispersed can give you a better understanding of what the middle values look like!

One of the seventh grade homeroom classes at Lincoln Middle School is selling baked goods before school in order to raise money for a class trip. They have the following six items for sale each morning.

Item	Price
cookie	$0.50
muffin	$2.00
doughnut	$1.25
scone	$1.25
bagel	$1.75
coffee cake	$2.25

To find the range of the data, you first want to order the values from smallest to largest. Just as with the median and the mode, this is the easiest way to organize the data and determine the smallest and largest values:

- $0.50, $1.25, $1.25, $1.75, $2.00, $2.25

Remember that the range is the difference between the largest and smallest values in the data:

- range = largest value - smallest value
- range = $2.25 - $0.50
- range = $1.75

In this case, the range in prices is $1.75. Now see what happens to the range when an outlier is added to the group. In addition to their other items, suppose that the seventh grade class begins selling whole pies for $12. Now the list of prices is as follows:

- $0.50, $1.25, $1.25, $1.75, $2.00, $2.25, $12.00

Take a look at what happens to the range:

- range = largest value - smallest value
- range = $12.00 - $0.50
- range = $11.50

Is the range affected by outliers? Yes! When an outlier was added, the range went from $1.75 to $11.50. That's a big jump! Remember that the mean is also directly affected by outliers. In fact, the range of the data can tell you something about how accurate of a representation the mean will be. Look at the mean of each of these sets:

Keep in mind! There are three types of measures of central tendency: the mean, the median, and the mode. Determining which measure most accurately describes a set is dependent on the situation.

- mean (small range) = $\frac{0.5 + 1.25 + 1.25 + 1.75 + 2 + 2.25}{6} = \frac{9}{6} = 1.5$, or $1.50
- mean (large range) = $\frac{0.5 + 1.25 + 1.25 + 1.75 + 2 + 2.25 + 12}{7} = \frac{21}{7} = 3$, or $3.00

Notice that the mean of the first set of data, where the range is small, provides a very good representation of the data. The value of $1.50 falls in the middle of the values in the set. When the range is large, however, as in the second set, the mean is not a very good representation of the data. The value of $3.00 falls above six of the seven values in the set. What's the connection? When the range is small, the mean is a more accurate description. When the range is large, the mean is probably not the best description of the data.

Think about it! Whether or not the range is large or small depends on the situation. For example, if you were dealing with thousands of dollars, a range of $11.50 would be very small. Make sure you analyze the situation before deciding whether the range is large or small.

Example:

- Find the range of the following set of data.
  - 27, 82, 14, 9, 11, 37, 29, 41

Solution:

- Begin by putting the values in numerical order:
  - 9, 11, 14, 27, 29, 37, 41, 82
  - range = highest value - lowest value
  - range = 82 - 9
  - range = 73
- The range of this set is 73.

S-T-R-E-T-C-H! Do you think the mean of this set will provide the most accurate representation of the data? From the range, would you say that the values in the set are close together or far apart?

The range can not only help you determine whether the mean is an accurate description of the data, but it can also help you understand more about the measures of central tendency. It has to do with the median. You already know that the median is the middle value of a set of data. Once ordered from smallest to largest, the median cuts the data in half. You can also cut the data into fourths. Each of the four parts is called a *quartile*. Look at the following set:

- 1, 1, 2, 5, 6, 9, 11, 15, 18, 21

Notice that it's already in numerical order.

The median of the upper half of the data is called the *upper quartile*. The median of the lower half of the data is called the *lower quartile*. Not only can you find the range of the entire set of data, but you can also find the range of the upper and lower quartiles. This range is called the *interquartile range*, and it's the difference between the upper quartile and the lower quartile.

The great thing about the interquartile range is that it is affected very little by outliers. Since outliers are extremely high or low values in the data, they usually have little effect on the middle of the data. So the interquartile range measures how far apart the data lie in the middle of the set. Find the interquartile range of the set from above:

Step by Step! To find the interquartile range:
1. put the set in numerical order;
2. find the median of the data;
3. find the median of both the upper and lower halves of the data;
4. subtract the lower quartile from the upper quartile.

- interquartile range = upper quartile - lower quartile
- interquartile range = 15 - 2
- interquartile range = 13

You can also compare the interquartile range to the range of the set. In this case, the range is the difference between 21 and 1, or 20. Since 13 and 20 are not too far apart, you know that the data in the set are pretty evenly distributed and there probably aren't any outliers.

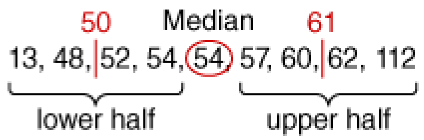
Example:

- Find the range and the interquartile range of the following set.
  - 52, 112, 60, 54, 54, 57, 62, 13, 48

Solution:

- Begin by putting the data in numerical order:
  - 13, 48, 52, 54, 54, 57, 60, 62, 112
  - range = highest value - lowest value
  - range = 112 - 13
  - range = 99
- Now find the interquartile range. Begin by finding the median. Since there is an odd number of items in the data, the middle value is the median. So 54 is the median of the set.

- Next, find the upper and lower quartiles. The upper quartile is the mean of 60 and 62, or 61. The lower quartile is the mean of 48 and 52, or 50:



Be Careful! Notice that in this case, the upper and lower quartiles are not actually part of the set. That's okay. It sometimes happens with the median, too. Make sure that you do not count them as members of the set if they weren't members to start with. There are still only the original nine numbers in the set.

- Last, calculate the interquartile range by subtracting the lower quartile from the upper quartile:
    - interquartile range = upper quartile - lower quartile
    - interquartile range = 61 - 50
    - interquartile range = 11
- So the range of the data is 99 and the interquartile range of the data is 11.

Notice in the above example how far apart the range is from the interquartile range. The interquartile range is only 11, and the range is 99. Even if you were given just those two measures, you would know that there is probably at least one outlier in the set. Towards the middle of the set, the values lie pretty close together. In the case where the range and interquartile range are far apart, the median is probably the most accurate representation of the data.

Example:

- Identify the upper and lower quartiles and find the range and interquartile range of the following set. Would the mean or the median be a more accurate representation of the data?
    - 31, 32, 32, 34, 25, 6, 26, 27

Solution:

- Begin by putting the data in numerical order:
    - 6, 24, 26, 27, 31, 32, 32, 34
- Find the median of the set.

$$\frac{27+31}{2}=\frac{58}{2}=29$$

- Find the upper and lower quartiles by finding the median of each half.
    - The median of 6, 24, 26, 27 is

$$\frac{24+26}{2}=\frac{50}{2}=25.$$

    - The median of 31, 32, 32, 34 is 32.
    - range = highest value - lowest value
    - range = 34 – 6
    - range = 28
- Now find the interquartile range.
    - 29 – 25 = 4.
- The median would be a more accurate representation of the data because the is a big difference between the range (28) and the interquartile range (4) in a set of 8 values.

Let's Review

Before going on to the practice problems, make sure you understand the main points of this lesson:

- The range is the difference between the highest and lowest values in a set of data.
- The interquartile range is the difference between the upper and lower quartiles in a set of data.
- The range is directly affected by outliers.
- The measures of dispersion can give you a better understanding of the measures of central tendency in a set of data by helping you determine which measure of central tendency is the most accurate description of the data.

Complete the following activities.

1.46 What is the range of the following set? 28, 45, 12, 34, 36, 45, 19, 20

☐ 8 ☐ 33 ☐ 45 ☐ 57

1.47 If 12 is replaced with 3 in the following set, what will happen to the value of the range? 28, 45, 12, 34, 36, 45, 19, 20

☐ It will increase. ☐ It will stay the same.

☐ It will decrease.

1.48 What is the median of the following set? 28, 45, 12, 34, 36, 45, 19, 20

☐ 28 ☐ 34 ☐ 31 ☐ 35

1.49 What is the upper quartile of the following set? 28, 45, 12, 34, 36, 45, 19, 20

☐ 36 ☐ 45 ☐ 32 ☐ 40.5

1.50 What is the lower quartile of the following set? 28, 45, 12, 34, 36, 45, 19, 20

☐ 28.5 ☐ 19.5 ☐ 19 ☐ 28

1.51 What is the interquartile range of the following set? 28, 45, 12, 34, 36, 45, 19, 20

☐ 21 ☐ 33 ☐ 16 ☐ 26

1.52 If 12 is replaced with 3 in the following set, what will happen to the value of the interquartile range? 28, 45, 12, 34, 36, 45, 19, 20

☐ It will increase. ☐ It will stay the same.

☐ It will decrease.

1.53 Martin earned the following scores on his last five tests. 98, 78, 84, 75, 91
What is the range of his scores?

☐ 7 ☐ 20 ☐ 23 ☐ 84

1.54 Martin earned the following scores on his last five tests. 98, 78, 84, 75, 91
What is the interquartile range of his scores?

☐ 23 ☐ 13 ☐ 16 ☐ 18

1.55 Trisha had the following seven distances (in feet) in the triple jump.
29, 30, 28, 32, 28, 31, 27
What was the range of her distances?

☐ 1 ☐ 3 ☐ 4 ☐ 5

1.56 Trisha had the following seven distances (in feet) in the triple jump.
29, 30, 28, 32, 28, 31, 27
What was the interquartile range of her distances?

☐ 3 ☐ 5 ☐ 2 ☐ 1

1.57 The range of a set of data is 15, and the interquartile range of the same set is 12.
Which of the following statements is probably true about the set?

☐ The two ranges are close together, so there is probably an outlier in this set.

☐ The two ranges are close together, so there are probably no outliers in this set.

☐ The two ranges are far apart, so there is probably an outlier in this set.

☐ The ranges are far apart, so there are probably no outliers in this set.

Use the following information to complete 1.58 to 1.62.

A college football team had the following scores during one season: 41, 35, 52, 40, 33, 42, 44, 38, 21, 29, 49, 49, 32, and 42. Use these numbers to answer the following questions.

1.58 What is the mean of the scores? Round your answer to the nearest whole number.

1.59 What is the range of the scores?

1.60 What is the interquartile range of the scores?

1.61 What is the median of the scores?

1.62 Would you choose to use the mean or the median to represent the data in this set? Why?

Review the material in this section in preparation for the Self Test. The Self Test will check your mastery of this particular section. The items missed on this Self Test will indicate specific areas where restudy is needed for mastery.

Self Test 1: Describing Data

Complete the following activities (5 points, each numbered activity).

1.01 The following is an example of a biased question. Would you prefer to watch a comedy or an action movie?

- ○ True
- ○ False

1.02 The mode of a set is the number that occurs the most in a set.

- ○ True
- ○ False

1.03 The upper quartile of a set of data is the mode of the upper half of the data.

- ○ True
- ○ False

1.04 An outlier always affects the mean.

- ○ True
- ○ False

1.05 A random sample was taken to determine whether students from a certain classroom prefer to shop at Store A, Store B, or Store C. Which of the following would represent a random sample?

- ☐ selecting all the students that have a last name that begins with R
- ☐ selecting all the girls
- ☐ putting all the names in a hat and selecting six students
- ☐ putting all the names in alphabetical order and selecting every fourth student

Use the following information to solve 1.06-1.07.

A random sample was taken to determine whether students from a certain classroom prefer to shop at Store A, Store B, or Store C. There were six students in the sample. Five students preferred Store A, one student preferred Store B, and none of the students preferred Store C.

1.06 Which of the following represents the percentage of students that preferred Store B?

☐ 20% ☐ 16.7% ☐ 0% ☐ 83.3%

1.07 If there are twenty-four total students in the classroom, about how many of them would you expect to prefer Store A?

☐ 20 ☐ 30 ☐ 29 ☐ 5

Use the following information to solve 1.08-1.09.

The boys' basketball team at Washington High School has been doing really well so far this year. In the past six games, they've scored the following points.
84, 67, 74, 87, 67, 73

1.08 What is the mean of their scores?

☐ 67 ☐ 75.3 ☐ 73.5 ☐ 20

1.09 If the coach is discussing their year with the local newspaper, which measure of central tendency is he *least* likely to use?

☐ mean ☐ median ☐ mode ☐ range

Use the following information to solve 1.010-1.013.

Danielle has taken five math tests so far this year. The tests are out of twenty points, and she has gotten the following scores. 17, 19, 20, 14, 16

1.010 What is the median of her scores?

☐ 17.2 ☐ no mode ☐ 6 ☐ 17

1.011 What is the mode of her scores?

☐ 17.2 ☐ no mode ☐ 6 ☐ 17

1.012 Which of the measures of central tendency would Danielle want her teacher to use in order to describe her test scores?

☐ mode ☐ range ☐ median ☐ mean

1.013 What must Danielle score on her sixth test in order to have a mean of 17.5?

- ☐ 17
- ☐ 18
- ☐ 19
- ☐ 20

1.014 What is the range of the following set of numbers? 114, 90, 83, 101, 97, 142, 117, 87, 72

- ☐ 70
- ☐ 42
- ☐ 45
- ☐ 59

1.015 Which numbers represent the lower quartile, the median, and the upper quartile of the following set of numbers? 114, 90, 83, 101, 97, 142, 117, 87, 72

- ☐ lower quartile: 87; median: 97; upper quartile: 117
- ☐ lower quartile: 85; median: 93.5; upper quartile: 115.5
- ☐ lower quartile: 83; median: 93.5; upper quartile: 114
- ☐ lower quartile: 85; median: 97; upper quartile: 115.5

1.016 What is the interquartile range of the following set of numbers?
114, 90, 83, 101, 97, 142, 117, 87, 72

- ☐ 12
- ☐ 18.5
- ☐ 30
- ☐ 30.5

Use the set of numbers to answer 1.017–1.021.
8, 9, 9, 10, 11, 13, 24

1.017 What is the mean of this set of numbers?

1.018 What is the mean of this set of numbers without the outlier?

1.019 What is the median of this set of numbers?

1.020 What is the interquartile range of this set of numbers?

1.021 Which is a more accurate representation of the data: the mean or the median?

84/105 **SCORE** ________ **TEACHER** ________ ________
initials date

MAT0707 – May ‘14 Printing
ISBN 978-0-7403-3172-5
9 780740 331725

Alpha Omega
PUBLICATIONS

804 N. 2nd Ave. E.
Rock Rapids, IA 51246-1759

800-622-3070
www.aop.com