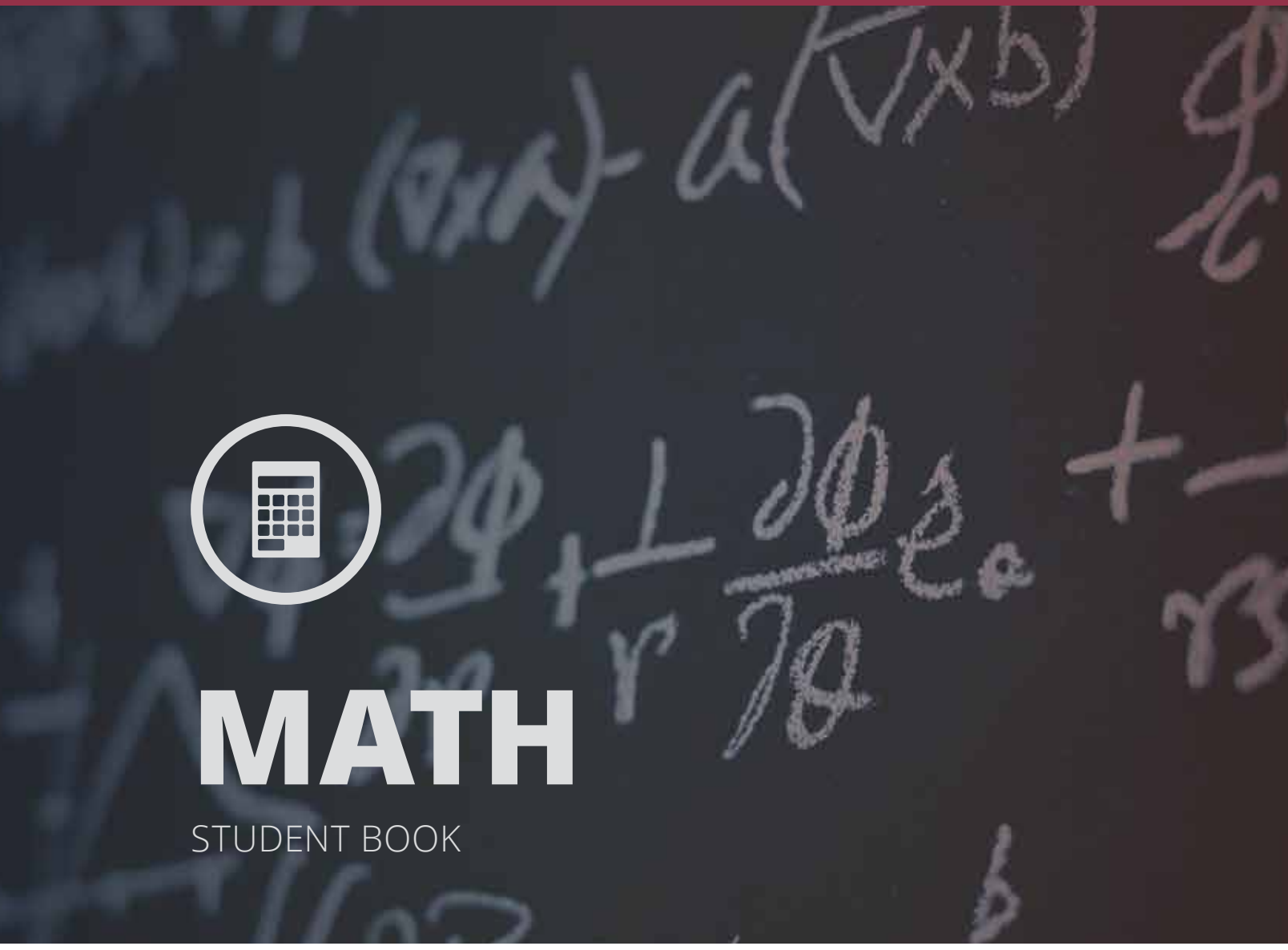
LIFEPAC®



MATH

STUDENT BOOK

7th Grade | Unit 7

Alpha Omega
PUBLICATIONS

Math 707
Data Analysis

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

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.

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.

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

Alpha Omega
PUBLICATIONS

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

800-622-3070
www.aop.com