LIFEPAC®

MATH

STUDENT BOOK

▶ **8th Grade** | Unit 9

Alpha Omega
PUBLICATIONS

Math 809
Data Analysis

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

Author:
Glynlyon Staff

Editor:
Alan Christopherson, M.S.

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

Data Analysis

Introduction

This unit explores methods for gathering and then drawing conclusions about data. Students are introduced to the distinction between categorical and nominal data. Two graphs for categorical data are explained—bar graphs and circle graphs. Five graphs for numerical data are covered—line graphs, stem-and-leaf plots, histograms, box-and-whisker plots, and scatter plots. The unit ends with lessons on misleading graphs and the selection of appropriate graphs. Students are expected to analyze, interpret, and make predictions about data.

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:

- Describe the various types of samples.

- Determine the measures of central tendency and dispersion.

- Identify types of data.

- Construct bar graphs, circle graphs, line graphs, stem-and-leaf plots, histograms, box-and-whisker plots, and scatter plots.

- List ways that graphs can be misleading.

- Determine the appropriate data display for a given set of data.

1. Collecting and Representing Data

COLLECTING DATA

Have you ever seen a TV commercial that claims that eight out of 10 dentists recommend a brand of toothpaste? Or, maybe a laundry detergent commercial that states that the product is preferred by four out of five moms?

There are standards that govern advertising, such as truth in advertising. This means that a company has to be able to prove the data that they claim to be true about their product. How is a company able to claim that four out of five moms prefer their product? Do you think the company was able to ask every single mom? Do you think the toothpaste company was able to get every dentist to answer their questions? That might not be possible. What they do is ask some of the dentists and some of the moms. This is called sampling. Let's find out more.

This lesson will help you understand how companies use *samples* to help them make
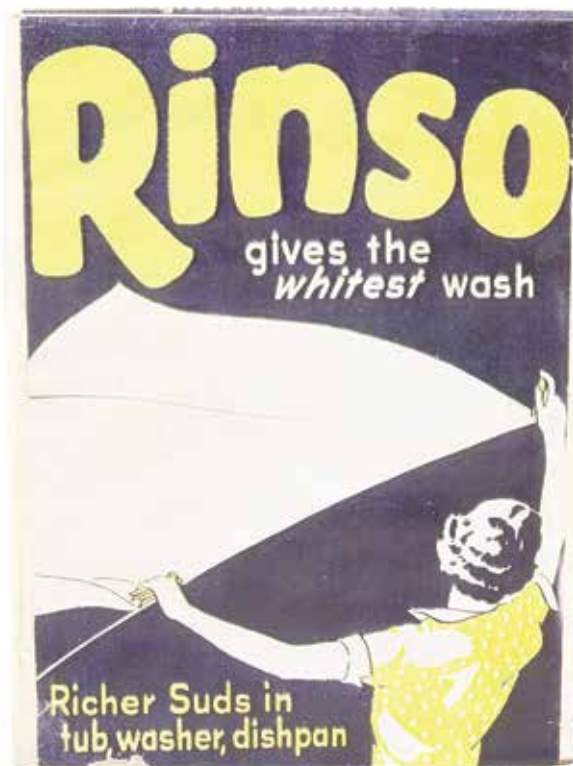
predictions about the general population's preferences. You will also learn about the many different types of sampling that can be used.

Objectives

- Identify a sample as biased or unbiased.
- Make predictions from a sample.
- Interpret a tally chart to identify trends and make predictions about the general population.

Vocabulary

biased sample—sample not representative of the entire population

convenience sample—only members of the population who are easily accessible are selected

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

random sample—every member of the population has an equal chance of being selected (unbiased sample)

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

self-selected sample—members of a population that volunteer for a survey

stratified sample—a population is divided into groups and then members are selected from each group

survey—a sampling of a population used to make predictions

systematic sample—a rule is used to select members from a population

Use of Sampling

Before a company can make a claim that their product is recommended by four out of five dentists, they have to conduct a *survey* to ask dentists their opinions about the product. In this example, the dentists are the *population* for the survey.

How do you think a company is able to get every dentist in the U.S. to answer their survey? They don't. It's extremely difficult to contact every dentist and get them to give their opinions. Because of this, companies use a method called sampling. A *sample* is a small part of a population that is chosen to represent the entire group.

From the results of their samples, companies can make predictions about larger populations. A variety of types of sample groups can be used. The larger the sample used, the more reliable the results and predictions will be.

Types of Samples

Any time you choose to use a sample, you need to be careful to make a good choice of which type to use.

Let's take a closer look at examples of each type of sample, in Figure 1.

A *biased sample* is a sample that is not representative of the entire population. An example of a biased sample happened in the 1948 presidential election. The Chicago Tribune printed a headline that read "Dewey Defeats Truman" based on the results of a telephone survey. In fact, Truman won the election. Their survey was biased because it included only people who owned phones, which in 1948 was not very many. Those that owned phones tended to be better off financially, which was another form of bias.

A *convenience sample* is a sample where the representatives chosen are easily

Type of Sample	Definition
biased sample	sample that isn't representative of the entire population
convenience sample	sample where representatives are easily accessible
random sample (unbiased sample)	every member of the population has an equal chance of being selected
self-selected sample	sample where members of the population volunteer to take part
stratified sample	members are selected from different groups that exist in the population
systematic sample	members are selected through the use of a rule

Figure 1| Types of Samples

accessible. Examples of convenience samples are using shoppers at a local mall or asking your friends their opinions. Convenience samples can be biased.

A *random sample*, or unbiased sample, is where every member of the population has an equal chance of being selected. An example of a random sample is putting all of the names of the students in the class in a hat and then randomly selecting two names from the hat.

A *self-selected sample* is a sample where members of the population volunteer to take part. An example of a self-selected survey is an online survey that appears on a website. People can choose to take the survey or not, making it a self-selected sample. Self-selection can bias a sample.

A *stratified sample* is produced when participants are selected from different groups that exist within a population. If a school cafeteria wanted to identify students' favorite snack foods, they could decide to ask 90 students from each of the grades at the school. Stratified samples are usually designed to produce random samples.

A *systematic sample* is a sample where members are selected through the use of a rule. An example of systematic sampling is asking every tenth person the same question. Systematic samples are designed to produce random samples.

Making Predictions from Samples

Once you have gathered information from a sample, you can use the results to make predictions about the entire population by using a simple proportion. Let's work through some examples.

Example:

▶ Thirteen out of 20 people prefer the taste of coffee B to coffee A. If a company plans to serve both coffees at their next meeting of 440 workers, how many cups of coffee B can they predict they will serve?

Please Check! You can solve a proportion by cross multiplying and then solving for the variable.

Solution:

$$\frac{13}{20} = \frac{x}{440}$$ Set up the proportion.

$(20)(x) = (13)(440)$ Cross multiply.

$20x = 5{,}720$ Multiply on the right side.

$x = 286$ Divide both sides by 20.

▶ The company can predict that they will serve 286 cups of coffee B.

Example:

▶ The local grocery store finds that 43 out of 100 shoppers use their frequent shopper's card. How many customers out of 1,200 can the store predict will use their frequent shopper's card?

Solution:

$$\frac{43}{100} = \frac{x}{1{,}200}$$ Set up the proportion.

$(100)(x) = (43)(1{,}200)$ Cross multiply.

$100x = 51{,}600$ Multiply on the right side.

$x = 516$ Divide both sides by 100.

▶ They can predict that 516 customers out of every 1,200 will use their frequent shopper's card.

Let's Review
Before moving on to the practice problems, make sure you understand the main points of this lesson.

- Companies often use surveys to determine the popularity of their products.

- Because it would be difficult to survey every person in a population, sampling is used.

- There are a number of different types of samples, some of which are more likely to accurately represent their populations.

- From the results of samples, companies can make predictions about entire populations.

Complete the following activities.

1.1 Fifty students were randomly selected to take part in a survey about their favorite school lunch. Each student was allowed to only pick one item. The results are shown in the table.

Food Item Preferred	Number of Students
pizza	27
cheeseburger	14
salad bar	9

Based on the information, if the school served 500 lunches, how many pizzas can they expect to serve?

- ☐ 27
- ☐ 270
- ☐ 140
- ☐ 90

1.2 A group of students wants to determine the favorite snack food at their school. They know they should use a sample because the student population at their high school is 5,000. What is the best way for them to get a random sample of 100 students?

- ☐ Ask the teachers at the school their opinions.
- ☐ Ask every 20th student until they reach 100 students.
- ☐ Ask the first 100 students that get off the bus in the morning.
- ☐ Ask all the students in the cafeteria during one of the four lunch periods.

1.3   Miguel surveyed all the members of the junior high choir about their favorite class. The results from his survey are:

Favorite Class	Number of Students
choir	19
math	7
history	5
English	9
science	4

Miguel concluded, from his results, that the favorite class among the entire student body would also be choir. Which is the *best* explanation of why his conclusion might not be true?

☐ The survey should have been conducted each day for a week.

☐ His sample group does not represent the entire student body.

☐ The survey should have been conducted with the eighth-graders only.

☐ The choir only meets one day a week.

1.4   A cola company wants to test their cola against their biggest competitor's cola. They decide to randomly sample 150 people. The results from their survey show that 98 people out of 150 people prefer their cola to their competitor's. What is the best prediction of the number of people who will prefer their cola if 1,200 people are asked?

☐ 784   ☐ 1,078   ☐ 980   ☐ 686

1.5   A _____ sample is a sample that isn't representative of the entire population.

1.6   A random sample is also called an unbiased sample.

○ True
○ False

1.7   If six out of 15 customers order a chicken sandwich, how many chicken sandwiches should a restaurant predict they will need to serve 270 people?

☐ 108   ☐ 270   ☐ 90   ☐ 120

1.8   A company surveys every tenth person in the phone book. What type of sampling did the company use?

☐ biased sample   ☐ stratified sample

☐ self-selected sample   ☐ systematic sample

MEASURES OF CENTRAL TENDENCY AND DISPERSION

Suppose you threw a dart at a dart board twelve times and recorded each of the numbers. Now, you have a set of numbers. Where do we go from here? We're going to want to find the *mean*, *median*, *mode*, and *range* of your numbers. As you go through this lesson, you will learn how to find these measures.

Objectives
- Identify the mean, median, mode, and range for a set of data.
- Calculate the missing value of a data set when given the mean and the rest of the data set.

Vocabulary
central tendency—ways to describe or summarize the clustering of data

data—information, often numerical

dispersion—how data are distributed

mean—the sum of the data divided by the quantity of data items

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

mode—the most frequently occurring number(s)

range—the difference between the largest and smallest data points

Measures of Central Tendency
Measures of *central tendency* are different ways to summarize *data*. The common measures of central tendency are the mean, median, and mode. Each measure gives different information about the same set of data. Therefore, each one serves a different purpose. Let's take a look at each of the different measures of central tendency.

Median
The *median* is the middle number of a set of data when the data have been put in numerical order.

Connections! The median of a set of data is just like the median on a road. You can find both medians in the middle!

- If there is an odd number of data points, the median will be the number in the middle.

- If there is an even number of data points, you will need to find the mean of the two middle numbers. To do this, add the two middle numbers together, and then divide by two.

Earlier, you threw a dart 12 times to gather a set of data. Let's find the median of this set of data.

Your dart throwing produced the following numbers, not necessarily in this order:

- 4, 17, 50, 11, 20, 18, 5, 19, 4, 16, 14, 15

The first thing we need to do is to put the numbers in order, from smallest to largest.

- 4, 4, 5, 11, 14, 15, 16, 17, 18, 19, 20, 50

The middle of your set of numbers falls between the 15 and 16. This is the middle because it cuts the data into two equal groups. But, we need to know what number represents the middle. So, we need to find out what is in the middle of 15 and 16. To do this, we need to add them together and then divide by two.

- 15 + 16 = 31
- 31 ÷ 2 = 15.5

Your dart throws have a median of 15.5.

Take a look at some more examples.

Example:

▶ Find the median of the data.

▶ 7, 6, 9, 11, 6, 8, 9

Solution:

▶ First, put the data in numerical order from smallest to largest.

▶ 6, 6, 7, 8, 9, 9, 11

▶ There is an odd number of data points so the median is the number in the middle.

▶ The median for this set of data is eight.

Example:

▶ Find the median of the following data.

▶ 11, 9, 14, 8, 12, 15, 8, 13, 7, 19

Solution:

▶ First, put the data in numerical order.

▶ 7, 8, 8, 9, 11, 12, 13, 14, 15, 19

▶ This time, we have an even number of data points. That means that both 11 and 12 are in the middle of the set of data. We now need to find the middle of those two numbers.

- 11 + 12 = 23
- 23 ÷ 2 = 11.5

▶ The median for this set of data is 11.5.

Mode

Another measure of central tendency is the *mode*. The mode is the number or numbers that occur most often. A set of data can have one mode, more than one mode, or no modes.

Let's use the set of data you collected from throwing the darts.

- 4, 17, 50, 11, 20, 18, 5, 19, 4, 16, 14, 15

You are looking for the number or numbers that appear the most often. You don't have to put the data in numerical order, but it can help.

- 4, 4, 5, 11, 14, 15, 16, 17, 18, 19, 20, 50

Now that they are organized, you can see that four is the number that occurs most often. Therefore, the mode for this set of data is four.

Here are a few more examples to take a look at.

Example:

▶ Find the mode of the set of data.

▶ 7, 6, 9, 11, 6, 8, 9

Solution:

▶ The mode is both six and nine, because they each appear twice.

Did you know! When a set of data has two modes, it is bimodal.

Example:

▶ Find the mode of the set of data.

▶ 11, 9, 14, 8, 12, 15, 8, 13, 7, 19

Solution:

▶ The mode for this data set is eight, because it is the only number that occurs more than one time.

Example:

▶ Find the mode of the set of data.

▶ 23, 16, 21, 18, 28, 27

Solution:

▶ There is no mode for this set of data, because each of the values occurs once.

Make note! The mode is "no mode," not zero. Zero implies that the number zero occurs the most often.

Mean

The last measure of central tendency we are going to look at is the *mean*. The mean is the sum of all of the data, divided by the number of data points. You already experience means every day in school. Your grade is determined by finding the mean of all of your assignments, tests, quizzes, etc.

Again, let's use the set of data from the dart throws.

■ 4, 17, 50, 11, 20, 18, 5, 19, 4, 16, 14, 15

The first step is to add all of the numbers together.

■ 4 + 17 + 50 + 11 + 20 + 18 + 5 + 19 + 4 + 16 + 14 + 15 = 193

Next, divide the sum by the number of pieces of data. In this case, we have 12 pieces of data. Round your answer to the nearest tenth, if it isn't a whole number.

■ 193 ÷ 12 = 16.1

■ The mean of the set of data is 16.1.

Let's take a look at finding the mean of some of the sets of data from previous examples.

Example:

▶ Find the mean of the set of data.

▶ 7, 6, 9, 11, 6, 8, 9

Solution:

▶ The first step is to find the sum of the data.

▶ 7 + 6 + 9 + 11 + 6 + 8 + 9 = 56

▶ Next, divide the sum by the number of data points.

▶ 56 ÷ 7 = 8

▶ The mean of the data is eight.

Example:

▶ Find the mean of the set of data.

▶ 11, 9, 14, 8, 12, 15, 8, 13, 7, 19

Solution:

▶ Begin by finding the sum of the data.

▶ 11 + 9 + 14 + 8 + 12 + 15 + 8 + 13 + 7 + 19 = 116

▶ Next, divide the sum of 116 by 10, because there are 10 pieces of data.

▶ 116 ÷ 10 = 11.6

▶ The mean of the data is 11.6.

Example:

▶ Find the mean of the set of data.

▶ 23, 16, 21, 18, 28, 27

Solution:

- ▶ 23 + 16 + 21 + 18 + 28 + 27 = 133
- ▶ 133 ÷ 6 = 22.2
- ▶ The mean is 22.2.

Finding Missing Data Points

Sometimes, you will be given the mean of a set of data and all but one of the data pieces, and then be asked to find the missing datum. Basically, you need to work backward, since you know the mean and want to find out what data produced it. Let's take a look at a few examples of how to find a missing piece of data.

Example:

- ▶ Felix struck out seven players in his first game, eight in his second game, three in his third game, four in his fourth game, and six players in his fifth game. After his sixth game, he had an mean of five strikeouts per game. How many players did he strike out in his sixth game?

Solution:

- ▶ The first step is to set up an equation to find the missing piece of data. We know that his mean was found by adding up all his strikeouts over the six games, and then dividing the sum by six. Let's set it up the same way. Since we don't know how many he struck out in game six, we will use a variable in its place.
- ▶ $7 + 8 + 3 + 4 + 6 + x = 5 \cdot 6$
- ▶ Notice, we set the equation equal to $5 \cdot 6$, which is his mean times the number of games. That gives us the total number of batters that struck out. We can now combine the like terms on the left side to help us solve the equation.

- ▶ $x + 28 = 30$
- ▶ $x = 2$
- ▶ Felix struck out two players in his sixth game.

Example:

- ▶ Keisha has already earned scores of 84, 93, 87, and a 91 in her Spanish class. What score does she have to get on her next test to have an mean test score of 90?

Solution:

- ▶ $84 + 93 + 87 + 91 + x = 90 \cdot 5$
- ▶ $x + 355 = 450$
- ▶ $x = 95$
- ▶ Keisha needs to earn a 95 on her next test in order to have a 90% mean on her tests.

Measure of Dispersion

There are a number of measures of *dispersion*, or how data are distributed. However, we are only going to focus on one in this lesson. The *range* of a set of data is the difference between the highest point of data and the lowest point of data. Simply, all you have to do is subtract the smallest number from the largest number.

Let's use your data from the darts one last time.

- ■ 4, 17, 50, 11, 20, 18, 5, 19, 4, 16, 14, 15

The first step is to identify the smallest and the biggest numbers found in the data. The smallest number you threw was four, and the largest was 50. The next step is to subtract four from 50.

- ■ 50 - 4 = 46
- ■ Your data has a range of 46.

# SELF TEST 1: Collecting and Representing Data

Complete the following activities (6 points, each numbered activity).

1.01 This table lists the number of floors in the 10 tallest buildings in the world. Note: The buildings are organized by height, not by the number of floors. What is the difference between the median and the mode for the number of floors?

| Skyscraper	Number of Floors
Taipei	101
Shanghai World Financial Center	101
Petronas Tower 1	88
Petronas Tower 2	88
Greenland Square Zifeng Tower	69
Sears Tower	108
Jin Mao Tower	88
Two International Financial Centre	88
Citic Plaza	80
Shun Hing Square	69

1.02 The following data needs to be put into a stem-and-leaf plot.
36.7, 38.1, 35.4, 37.4, 39.0, 33.6, 34.7, 36.9, 38.7, 35.5, 36.8, 33.7, 34.8, 39.7, 37.2, 36.8
Which list shows all the stems needed to accurately create a stem-and-leaf plot?

☐ 3 ☐ 0, 1, 2, 4, 5, 6, 7, 8, 9

☐ 3, 4, 5, 6, 7, 8, 9 ☐ 33, 34, 35, 36, 37, 38, 39

1.03 You are going to create a circle graph to represent some data. How many degrees should a section representing 65 out of a hundred be?

☐ 234° ☐ 65° ☐ 100° ☐ 13°

1.04 You are surveying students to find out if they prefer anchovies or vegetables on pizza. Is the following question biased or unbiased? Do you prefer slimy anchovies or fresh vegetables on your pizza?

☐ biased

☐ unbiased

1.05 What is the difference between the number of the students who prefer history to English?
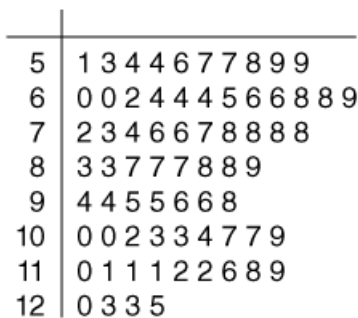
Favorite Class
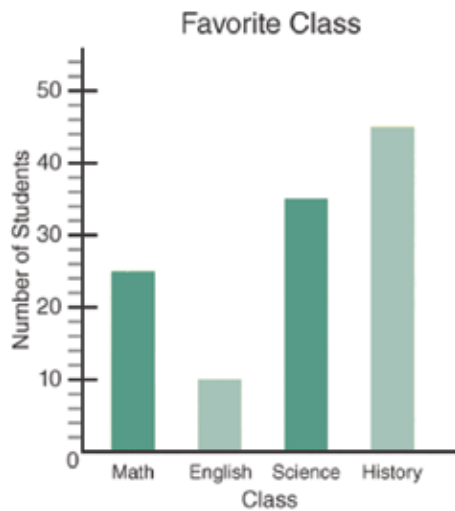
☐ 45

☐ 10

☐ 55

☐ 35

1.06 Between which two months was the largest decrease in profits recorded? What was the amount of the decrease?
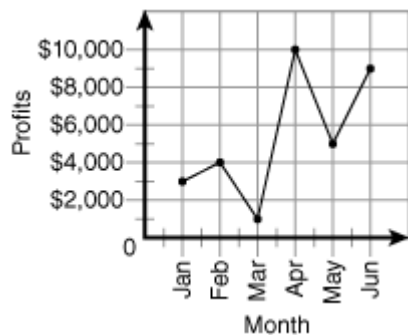
☐ February to March, $4,000

☐ February to March, $5,000

☐ April to May, $4,000

☐ April to May, $5,000

1.07 The stem-and-leaf plot shows the number of people who used each run of the Red Line train. What is the range of the data?

5	1 3 4 4 6 7 7 8 9 9
6	0 0 2 4 4 4 5 6 6 8 8 9
7	2 3 4 6 6 7 8 8 8 8
8	3 3 7 7 7 8 8 9
9	4 4 5 5 6 6 8
10	0 0 2 3 3 4 7 7 9
11	0 1 1 1 2 2 6 8 9
12	0 3 3 5

5|4 = 54

☐ 51

☐ 125

☐ 74

☐ 78

1.08 If this circle graph represents the responses from 350 people, how many people chose silver as their favorite car color?

FAVORITE CAR COLOR

Black 25%
White 10%
Silver 10%
Red 50%
Yellow 5%
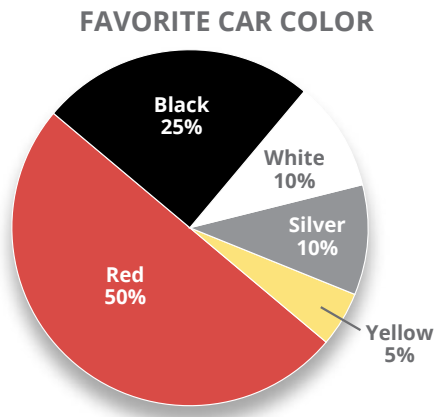
☐ 350
☐ 10
☐ 35
☐ 100

1.09 Match each word to its definition.

_____ sample not representative of the entire population

_____ members are selected from the different subgroups within a population

_____ sample where members volunteer to take part

_____ sample members are selected through the use of a rule

_____ sample where every member of the population has an equal chance of being selected

_____ sample where representatives are easily accessible

biased sample
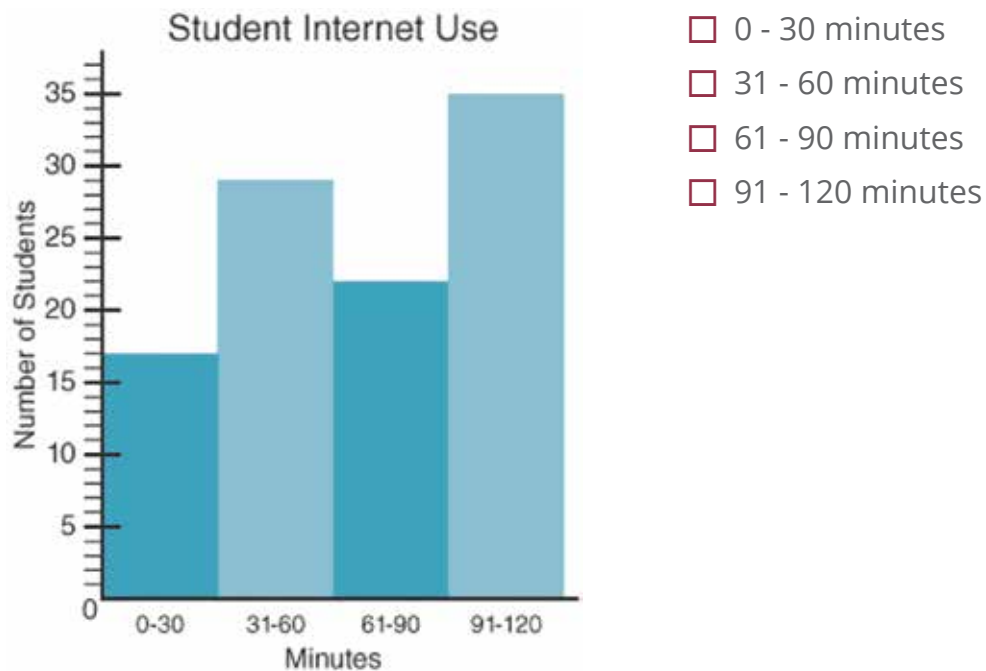convenience sample
unbiased sample
self-selected sample
stratified sample
systematic sample

1.010 The bear was chosen _____ more times than the elephant as the favorite animal.

□ 11

□ 10

□ 9

□ 8
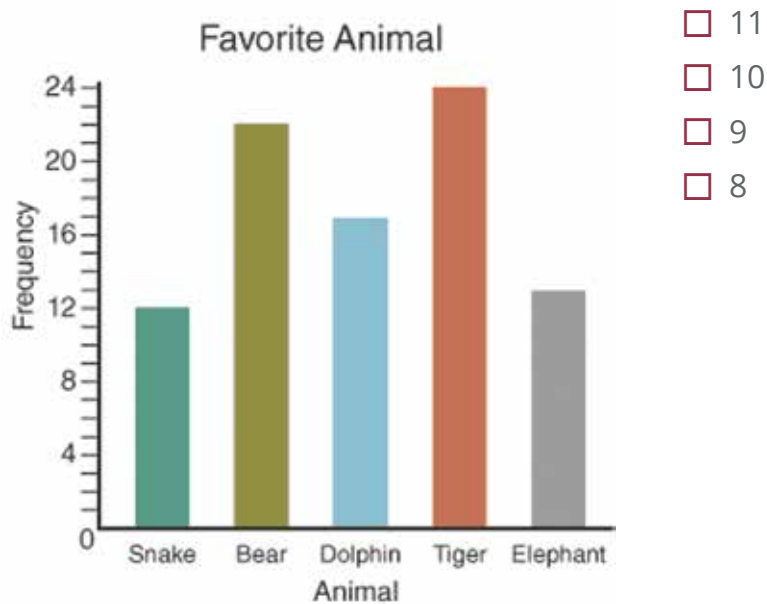
1.011 According to the histogram, which interval of time do most students spend on the Internet?

□ 0 - 30 minutes

□ 31 - 60 minutes

□ 61 - 90 minutes

□ 91 - 120 minutes

1.012 What was the difference in profits between the two companies for the year 2000?

Profits
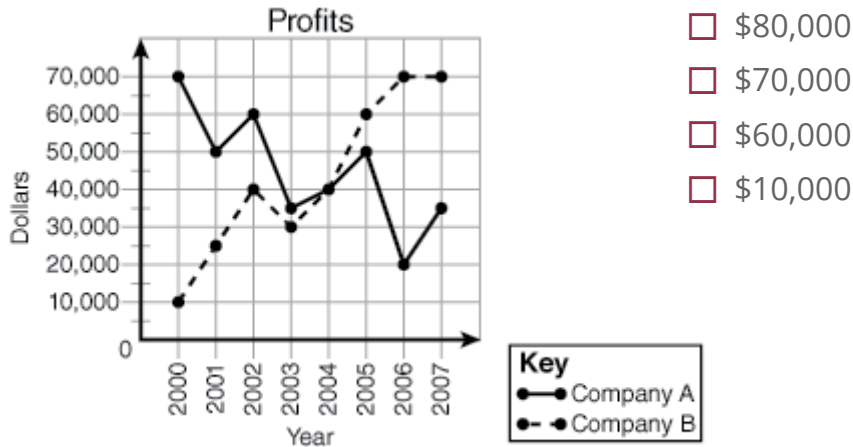
☐ $80,000

☐ $70,000

☐ $60,000

☐ $10,000

1.013 Which list shows the mean, median, mode, and range in ascending order for this set of data?   23, 15, 35, 31, 45, 19, 15, 22, 27, 41, 30, 24, 44, 15, 19

☐ mean, median, mode, range

☐ median, mode, mean, range

☐ mode, median, mean, range

☐ mode, mean. median, range

1.014 If the circle graph represents the responses from 500 people, how many more people prefer burgers than cheeseburgers?

FAVORITE FOOD
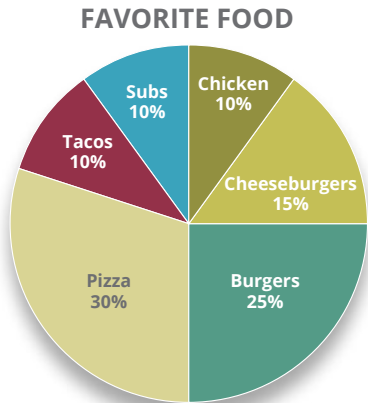
☐ 125

☐ 75

☐ 50

☐ 25

1.015 Malina has to bowl four games in her league. She bowled a 149, 162, and a 152 in her first three games. She wants to have an average score of 154. What score does she need in her fourth game?

☐ 152   ☐ 153   ☐ 154   ☐ 155

72 / 90		SCORE _____	TEACHER _____ _____
			initials date

Alpha Omega
PUBLICATIONS