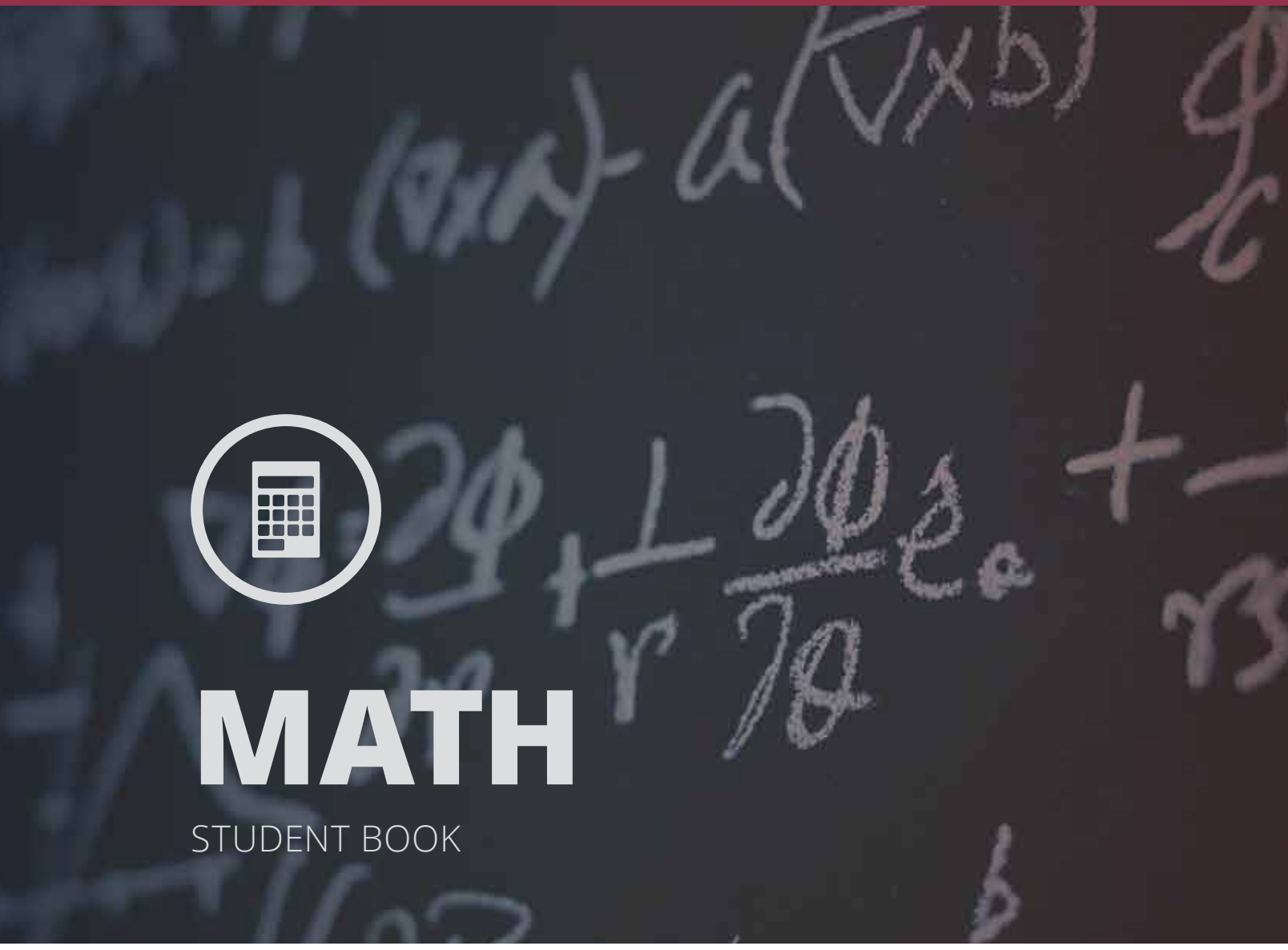
LIFEPAC®



MATH

STUDENT BOOK

6th Grade | Unit 2

Alpha Omega
PUBLICATIONS

MATH 602

Data Analysis

INTRODUCTION |3

1. COLLECTING AND DESCRIBING DATA 5

COLLECTING VALID DATA |6
MEDIAN, MODE, AND RANGE |12
DESCRIBING DATA |17
SELF TEST 1: COLLECTING AND DESCRIBING DATA |22

2. ORGANIZING DATA 24

FREQUENCY TABLES |25
HISTOGRAMS |30
LINE PLOTS |36
STEM-AND-LEAF PLOTS |43
SELF TEST 2: ORGANIZING DATA |50

3. DISPLAYING AND INTERPRETING DATA 53

BAR GRAPHS |54
LINE GRAPHS |59
VENN DIAGRAMS |65
VERTEX-EDGE GRAPHS |71
SELF TEST 3: DISPLAYING AND INTERPRETING DATA |77

4. REVIEW 80

GLOSSARY |89

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

Author:
Glynlyon Staff

Editor:
Alan Christopherson, M.S.

MEDIA CREDITS:
Page 5: © Ridofranz, iStock, Thinkstock; **12:** © Jupiterimages, Brand X Pictures, Thinkstock; **30:** © Ingram Publishing, Thinkstock; **59:** © bgsmith, iStock, Thinkstock; © lucky336, iStock, Thinkstock.

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

Data Analysis

Introduction

In this unit, you will be introduced to the topic of Data Analysis. You will learn about ways to collect and organize data, and many ways to display data. You'll learn what type of graph is best to display different data sets. You'll also discover different ways to describe data, called measures of central tendency. You will learn many new terms and have several concepts to build on in your future studies of statistics.

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.
- Find the measures of central tendency.
- Organize and display data in frequency tables, histograms, line plots, stem-and-leaf plots, bar graphs, and line graphs.
- Use Venn diagrams to solve problems, including counting problems
- Solve route problems with vertex-edge graphs.

1. COLLECTING AND DESCRIBING DATA

Mr. Smith asked his class at ABC School if they could determine the height of a typical sixth grader at their school. Four of the students arrived at the following conclusions:

- Cathy – A typical 6th grader is about 5 feet tall.
- Al – A typical 6th grader is about 6 feet tall.
- Debbie – A typical 6th grader is about $5\frac{1}{2}$ feet tall.
- Bob – A typical 6th grader is about $5\frac{1}{2}$ feet tall.

How could these students draw such different conclusions? To answer this question, we will begin to explore the mathematical subject of **statistics**. In statistics, we look at a set of data (such as the height of sixth graders) and see how it can be organized and analyzed to provide us with useful information.

In this lesson, we will look at how **data** is collected and how we can draw conclusions about a set of data using the **mean**.

Objectives

Review these objectives. When you have completed this section, you should be able to:

- Determine whether a sample is biased or random.
- Describe a set of data using the mean.
- Find the median, mode, and range for a set of data.
- Describe a set of data using the median, mode, and range of a set of numerical data.
- Determine when each measure of central tendency provides a good representation of the data.
- Determine how an outlier affects the measures of central tendency.

Vocabulary

biased question. A question that leads individuals toward a certain answer.

biased sample. A sample not representative of the entire population.

central tendency. Ways to describe or summarize data.

data. Information (often numerical).

extreme values. The smallest and largest values in a data set.

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.

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

population. All of the possible data in a given topic.

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

range. The difference between the largest and smallest data points.

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.

Note: *All vocabulary words in this LIFEPAC appear in* **boldface** *print the first time they are used. If you are not sure of the meaning when you are reading, study the definitions given.*

COLLECTING VALID DATA

If we are curious to know the answer to a question such as "What is the typical height of a sixth grader?" we can often find answers by collecting and analyzing data.

One way to collect data is to take a **survey** of the **population** we are interested in. If the population is small enough, the whole population can be surveyed. However, if the population is very large—let's say all sixth graders in your state—then we need a **sample** of sixth graders that represents the population.

Samples can be biased or random. A **biased sample** means that the sample does not represent the population. A biased sample might not represent the population because not enough data was collected or because the sample did not truly represent the population. The best kind of sample to get is a **random sample**. A random sample means that every member of the population has an equal chance of being selected. Let's look at the samples taken by Mr. Smith's students to determine if the samples are biased or random.

In the ABC School, each of the four students in Mr. Smith's class took a survey of sixth graders to collect data. Here is how each student collected the data:

- Al is on the sixth-grade basketball team. He measured himself and his teammates.
- Bob measured the first 20 sixth graders he saw at recess.
- Cathy measured herself and four of her friends.
- Debbie measured all 80 sixth graders.

What do you think about Al's sample? Do you think sampling basketball players would provide a realistic sample of typical sixth graders? Probably not. We would say Al's sample is biased because it does not represent the population. Cathy did not collect enough data, so her sample is biased also.

Bob's sample is random and unbiased, and he has enough data to represent the population. Bob and Debbie reached the same conclusion, yet Debbie measured all of the sixth graders, and Bob measured only 20 sixth graders. The more data we have, the more likely our conclusions are valid, but we need to be efficient also.

Did you know?

Part of statistics is carefully choosing a relatively small sample that will represent the whole population. For example, it would be impossible to survey every home in the country to see what TV show people watch on Thursday!

Example:

For each of the following topics, should we sample the population, or survey everybody?

1. the most popular movie star in America
2. the favorite food in Mr. Smith's class
3. the number of hours of TV that students at ABC School watch in a week

Solution:

For each case, we will look at the size of the population and determine the size of the sample.

1. The American population is too large to survey everybody, so we need a relatively small sample to determine the most popular movie star.
2. The number of students is small enough in Mr. Smith's class that the survey can include everybody.
3. The number of students in the whole school is too large to survey everyone's TV-watching habits. A small sample from each class could be taken.

It is important that the sample is random, so that the results will not be biased.

Example:

Which of the following samples would most likely be random?

1. People coming out of the library are asked about a library tax.
2. A questionnaire is included in the newspaper.
3. An alphabetical list of students is sampled by pulling assigned numbers out of a basket.

Solution:

For each sample, we need to look at whether the sample represents the population and whether it is random.

1. Not random: The people being surveyed are likely to have strong opinions about the library.
2. Not random: Only people with strong opinions are likely to respond.
3. Random: Each student is chosen randomly and will be likely to represent the population.

Another factor that can cause a sample to be biased is how a question on a survey is worded. A **biased question** seems to have a right or wrong answer. The question seems to be leading you to respond in a certain way.

Examples of biased questions:

Do you, like everybody your age, dislike classical music?
Do you agree with the unfair tax law?
Who is doing a better job, the efficient mayor, or the lazy chief of police?

People who give surveys need to be careful that questions are unbiased so that the results will be valid.

Examples of unbiased questions:

What is your height?
What food do you like best?
What color are your eyes?

FINDING THE MEAN

Once we have collected a random, unbiased sample, how do we analyze the data? There are several measures of central tendency that help us summarize the data. A measure of **central tendency** gives us one number that represents all the data. One of these is the **mean**, often called the average.

The mean tells us what each number would be if the data were evened out. The mean is found by adding up all the numbers in the data set, and dividing by the number of items in the set:

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

Example:

Find the mean for the high temperatures for a week in Honolulu, Hawaii:
89°, 87°, 85°, 84°, 86°, 83°, 88°

Solution:

To find the mean, we will add all the numbers, and divide by the number of temperatures. There are seven temperatures, so we will divide the total by seven.

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

$$\frac{89° + 87° + 85° + 84° + 86° + 83° + 88°}{7} = \frac{602°}{7}$$ Add the numbers.

$$\frac{602°}{7} = 86°$$ Divide by 7.

So, the mean high temperature for the week was about 86°.

So, we can say that the temperature in Honolulu was around 86° for the week.

Let's go back to our problem from the beginning of our lesson and take a look at the data Bob collected for the height of sixth graders at ABC School.

Example:

Bob measured 20 sixth graders in inches: 65, 67, 64, 68, 66, 66, 60, 72, 60, 69, 69, 65, 67, 64, 67, 67, 66, 68, 66, 66. Is his conclusion that sixth graders at ABC School are around $5\frac{1}{2}$ feet (66 inches) tall correct?

Solution:

To see if Bob is correct, we will use the mean to summarize the data by adding the numbers and dividing by the number of students (20).

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

$65 + 67 + 64 + 68 + 66 + 66 + 60 + 72$

$+ 60 + 69 + 69 + 65 + 67 + 64 + 67 + 67$ Add all the heights.

$+ 66 + 68 + 66 + 66 = 1{,}320$"

$\frac{1320}{20} = 66$" Divide by 20.

So, Bob's conclusion is correct: sixth graders at ABC School are around 66 inches ($5\frac{1}{2}$ feet) tall.

Think about it!

The mean will always be between the lowest and highest number in the data set. Can you see why?

Let's Review!

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

- ✓ Sampling is used to draw conclusions about a population that is too large to survey in its entirety.
- ✓ Biased samples and questions can affect the validity of the data.
- ✓ The mean is one measure of central tendency and helps to summarize the data.

Match the following items.

1.1

_____ a question that leads individuals toward a certain answer

_____ a sample not representative of the entire population

_____ ways to describe or summarize data

_____ information (often numerical)

_____ the sum of a set of data divided by the number of items in the set

_____ all of the possible data in a given topic

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

_____ a small part of a population chosen to represent the entire group

_____ the collection, organization, and analysis of numerical information

_____ a sampling of a population used to make predictions

a. central tendency
b. biased sample
c. population
d. random sample
e. statistics
f. biased question
g. sample
h. survey
i. data
j. mean

Circle each correct answer.

1.2 Which sample size of a population of 200 is most likely to give a reliable conclusion?

a. 10 b. 30 c. 50 d. 80

1.3 Which question is a biased question?

a. What is your favorite color out of red, blue, and green?
b. How old are you?
c. Do you prefer a beautiful sunny day or a depressing rainy day?
d. How many siblings do you have?

1.4 A sample of a population of 100 students is collected. Five students playing dodge ball at recess are asked if they prefer playing outdoors or indoors. Which of the following is *not true* of the sample?

a. The sample is not large enough.
b. The question is biased.
c. The sample is not random.
d. The students are not part of the population.

1.5 Students in Mr. Smith's class want to find out how far they can jump. Which of the following will help give a valid conclusion about the data?

a. Five students are chosen randomly to jump.
b. All 25 students are measured.
c. Students who want to jump are measured.
d. Students playing during recess are measured.

1.6 What is the mean for the following set of data?
12, 10, 11, 13, 14

a. 10 b. 12 c. 13 d. 14

1.7 The mean for a set of data is 10. Which of the following is the set of data with a mean of 10?

a. 9, 9, 9, 9, 9 b. 12, 11, 13, 14, 11 c. 8, 12, 10, 13, 7 d. 5, 8, 9, 6, 7

1.8 The high temperatures for each day last week increased by one degree each day: 65°, 66°, 67°, 68°, 69°, 70°, 71°. What was the mean temperature for the week?

a. 68° b. 65° c. 71° d. 70°

SELF TEST 1: COLLECTING AND DESCRIBING DATA

Circle each correct answer (each answer, 7 points).

1.01 What would be the *best* way to get an unbiased sample that represents the population for the following topic: What is the favorite sport in Mr. Smith's class?

a. Ask four people on your soccer team.
b. Ask, "Do you like silly soccer, or exciting football?"
c. Ask people reading at recess.
d. Randomly survey at least 20 people in the class.

1.02 How much data is needed to have a representative sample of the population?

a. 10 to 20 items.
b. At least 100 items.
c. All of the population.
d. It depends on the size of the population.

1.03 Which sample size out of a population of 1,000 is most likely to lead to a valid conclusion?

a. 50 b. 100 c. 150 d. 200

1.04 What is the mean of the set of data?
6, 7, 10, 12, 12, 13

a. 6 b. 7 c. 10 d. 11

1.05 What is the median for the set of data?
6, 7, 10, 12, 12, 13

a. 7 b. 10 c. 11 d. 12

1.06 What is the mode for the set of data?
6, 7, 10, 12, 12, 13

a. 10 b. 11 c. 12 d. 13

1.07 What is the range for the set of data?
6, 7, 10, 12, 12, 13

a. 6 b. 7 c. 11 d. 13

1.08 Which set of data has the same mode and median?

a. 3, 4, 6, 7, 8 b. 3, 3, 4, 5, 5 c. 6, 7, 7, 8, 9 d. 2, 2, 3, 4, 6

1.09 Which value in the data set is an outlier?
7, 8, 9, 10, 10, 22

a. 7 b. 9.5 c. 10 d. 22

1.010 Given the following information, which is the best description of the data?
Range — 5, from 12 to 17; Mode — 14; Median — 14.5; Mean — 15

a. The data is around 12.
b. The data is around 14.
c. The data is around 15.
d. The data is around 17.

1.011 What will happen to the mean if the outlier is removed?
4, 5, 5, 7, 9, 17

a. It will not change.
b. It will be the same as the median.
c. It will decrease.
d. It will increase.

1.012 A set of data is described as:
The data is around 7. If another measurement were taken it would probably be around 7.
Which group of measures would lead to this conclusion?

a. Range — 10
Mode — 7
Median — 9
Mean — 5

b. Range — 4
Mode — 6
Median — 7
Mean — 8

c. Range — 12
Mode — 9
Median — 9
Mean — 9

d. Range — 5
Mode — 9
Median — 9
Mean — 7

Place a check mark next to each correct answer (check all that apply; each answer, 8 points).

1.013 Given the following set of data, which measures have a value of 6?
3, 4, 6, 8, 9

☐ range
☐ mode
☐ median
☐ mean

1.014 Given the set of data below, which measure(s) will change if the outlier is removed?
4, 7, 9, 9, 10

☐ range
☐ mode
☐ median
☐ mean

80/100 **SCORE** _____ **TEACHER** _____ _____
initials date

Alpha Omega
PUBLICATIONS

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

800-622-3070
www.aop.com

MAT0602 – Apr '15 Printing

ISBN 978-0-7403-3466-5