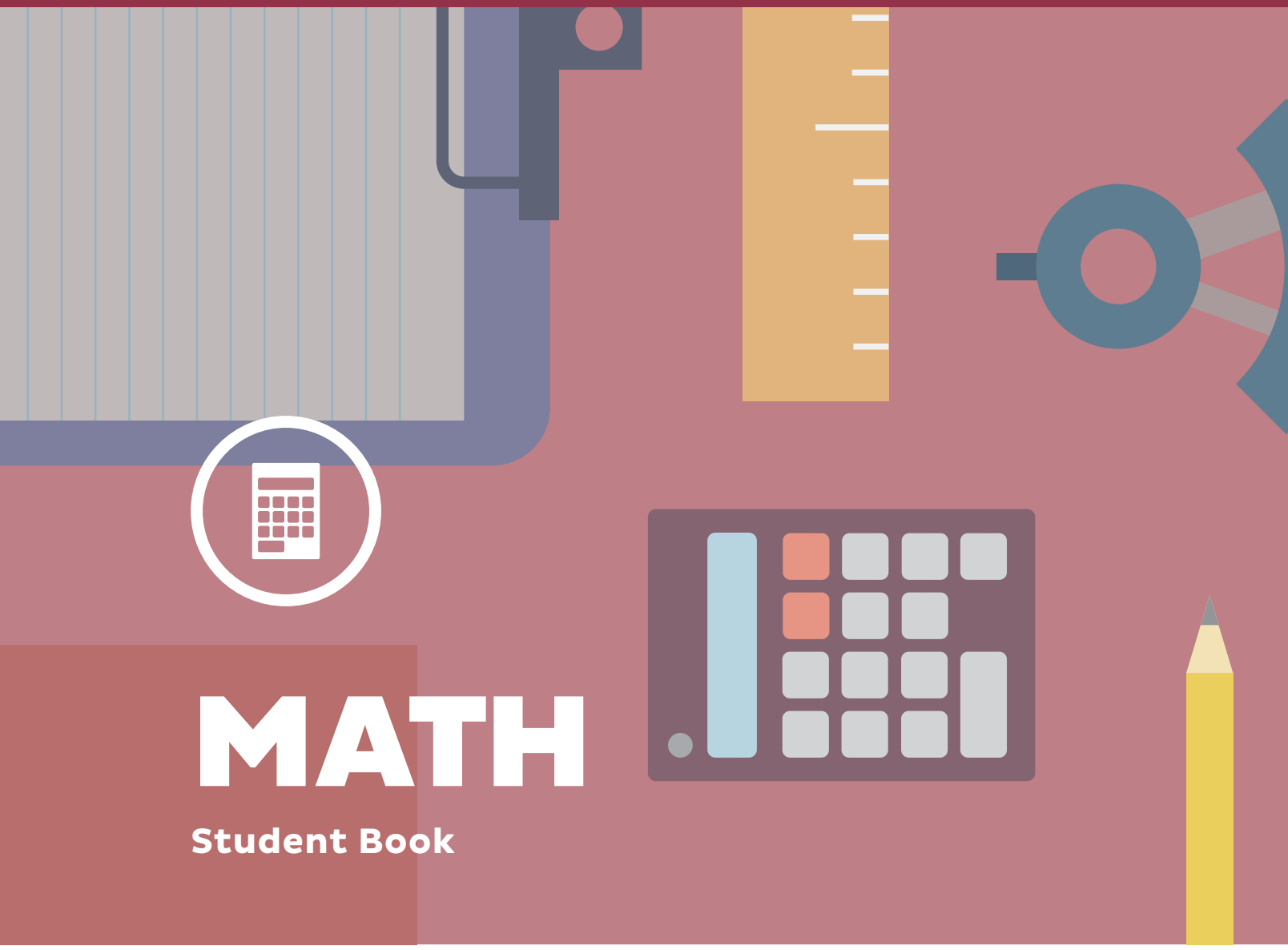
LIFEPAC®



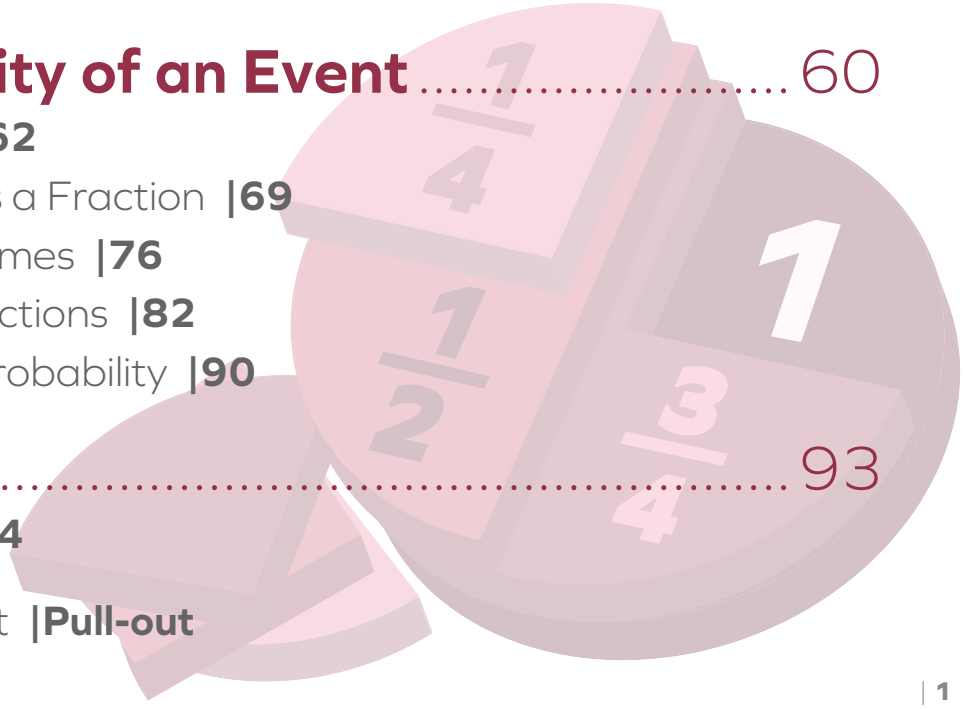
MATH

Student Book

5th Grade | Unit 8

Alpha Omega
PUBLICATIONS

MATH 508
DATA ANALYSIS AND PROBABILITY



Author:
Glynlyon Staff

Editor:
Alan Christopherson, M.S.

Media Credits:
Page 3: © Mirexon, iStock, Thinkstock; **4:** © prapann, iStock, Thinkstock; **6:** © Vectoraart, iStock, Thinkstock; **35:** © pixel_dreams, iStock, Thinkstock; **56, 58:** © block37, iStock, Thinkstock; **60:** © Baksiabat, iStock, Thinkstock; **68:** © marlanu, iStock, Thinkstock; **93:** © pattilabelle, iStock, Thinkstock.

Alpha Omega
PUBLICATIONS

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

DATA ANALYSIS AND PROBABILITY

In this unit, you will learn about ways to analyze data and you will be introduced to probability. You will learn that the way data is collected can affect the results. You will become familiar with several ways to display data. You will learn about measures of central tendency as tools to describe data sets. All the types of graphs and measures give you ways to analyze data.

You will learn about likelihood and probability, and will find that likelihood can be expressed as a fraction. You will learn that probability is a fraction from 0 (impossible) to 1 (certain), and that a probability of $\frac{1}{2}$ means the outcome is as likely as unlikely. You will learn about using a tree diagram to find the sample space for two independent events. Finally, you will learn that theoretical probability and experimental probability are used to predict results for any number of trials.

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:

- Analyze data using the mean, median, mode, and range.
- Choose the best way to display data, including a: frequency table, line plot, stem-and-leaf plot, bar graph, line graph, and pictograph.
- Use probability to determine the likelihood of events

1. COLLECTING AND ANALYZING DATA

Nutmeg and Pepper discuss how to keep track of their nuts for the winter. Pepper suggests they make piles and hope they don't run out. Nutmeg suggests a frequency table.

If you have a lot of information, either in different categories, or in different amounts, how can you make sense of it? One way to organize data is to use a frequency table, as Nutmeg suggests. In this lesson, you will learn about collecting data and organizing it in a frequency table.

Objectives

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

- Collect data.
- Organize data using a frequency table.
- Find the mean, median, mode, and range of a set of data.
- Organize data using a line plot.
- Construct a line plot.
- Organize data using a stem-and-leaf plot.
- Construct a stem-and-leaf plot.

Vocabulary

Study these new words. Learning the meanings of these words is a good study habit and will improve your understanding of this LIFEPAC.

biased question. A question that leads individuals toward a certain answer.
biased sample. A sample not representative of the entire population.
categorical data. Data grouped using non-numerical criteria.
central tendency. Movement in a particular direction.
data. Information (often numerical).
extreme values. The smallest and largest values in a data set.
frequency table. A display that shows data in groups or intervals.
interval. The distance between numbers on a graph or data display.
leaf. The last digit in a stem-and-leaf plot.
line plot. A graph showing frequency of data on a number line.
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 data set.
numerical data. Data represented by quantities.
outlier. A value that is far removed from the rest of the values in a data set.
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 the population chosen to represent the entire group.
statistics. The collection, organization, and analysis of numerical information.
stem. The leading digit(s) in a stem-and-leaf plot.
stem-and-leaf plot. A data display that shows data arranged in numerical order by intervals.
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 unsure of the meaning when you are reading, study the definitions given.*

Collecting Data

We collect **data** on a topic because we want to answer a question about that topic, or make a prediction about that topic. How tall is a typical fifth grader? What color do people like best? How many acorns have I collected for winter? The area of mathematics called **statistics** helps us draw conclusions about the data that we collect.

The first part of collecting data on a topic is deciding how much data we need, and from whom or where to collect the data. Often, we collect data by taking a **survey**. If we want to know what subject 5th graders like best, we can't ask every 5th grader, so we take a **sample** of the **population**. A sample is a small part of the population chosen to represent the entire group.

Connections

When we see surveys that say that 65 percent of Americans think the president is doing a good job, they do not ask every American for their opinion! A carefully selected sample, based on statistical analysis, is used to represent the population.

If the sample is too small, our conclusions are not valid. If we ask four 5th graders what their favorite subject is and two say math and two say history, would that mean that 5th graders like only math and history? Probably not. However, if we ask thirty 5th graders what their favorite subject is, some might say history, some might say reading or other subjects. But, if twenty out of thirty say math, we could conclude that math was the favorite subject, and it would not be necessary to survey one hundred more 5th graders.

The amount of data necessary to draw a valid conclusion will vary from topic to topic. In general, if there is not a clear conclusion, then more data is needed. For instance, if we want to find out how tall a typical 5th grader is, but we only measure four boys on the basketball team, we may get a clear conclusion, but it won't be valid. So, whom we survey also affects our results.

The second part of collecting data is asking the right question. If we asked 5th graders what their favorite subject is between math and history, we wouldn't necessarily find out the favorite subject because they might like a different subject. Or, if we asked, "Do you like fun and exciting math, or boring old history?" our data would be biased because the way the question is asked would lead people to choose math. When a sample leads people to a certain conclusion, either by who is surveyed (such as four boys on the basketball team), or by how the question is asked (called a **biased question**), then it is called a **biased sample**.

The goal in statistics is to collect a large enough **random sample** to draw valid conclusions. A random sample is a sample in which *every* member of the population has an equal chance of being selected. Instead of measuring only basketball players to find the height of students, which would be a biased sample, the first ten names from each 5th grade class could be measured.

Organizing Data Using a Frequency Table

Once we've collected the data, we need a way to organize it. This can be done using a frequency table. A frequency table lists each piece of data from the survey. A frequency table for the topic of favorite subject might look like this:

If we construct the table before doing the survey, the table can also be used to help collect the data. If we are asking what the favorite subject is, we don't know what the responses will be, but we can add a row for each subject as it is selected. Each time a subject is selected, a tally is added to that row. When the survey is complete, we can total the tallies.

SUBJECT	TALLIES	FREQUENCY
Math	𝍸 𝍸 ||	12
Science	||||	4
History	𝍸	5
Language	||	2
Art	𝍸 ||	7

Make sure that the total number of tallies is the same as the number of people in your survey. For instance, in the class above there are thirty students. Was everyone surveyed? Yes, because the tallies add to thirty:

$$12 + 4 + 5 + 2 + 7 = 30$$

What can we conclude from the data in the frequency table? We can see that almost half of the class likes math (12 out of 30; 15 out of 30 would be $\frac{1}{2}$), so it looks like it is the favorite subject.

Let's see how Nutmeg and Pepper constructed their frequency table for the nuts they have gathered for the winter.

Nutmeg and Pepper have collected acorns, walnuts, pecans, and almonds. So Nutmeg made a table with those categories. Pepper began going through the pile as Nutmeg tallied each nut. Then they totaled each category. What can you tell about the nuts that Nutmeg and Pepper collected? Of the 100 nuts they collected, half of the nuts (50) are acorns. There are half as many walnuts (25) as acorns.

NUT	TALLIES	TOTAL
Acorns	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	50
Walnuts	𝍸 𝍸 𝍸 𝍸 𝍸	25
Pecans	𝍸	5
Almonds	𝍸 𝍸 𝍸 𝍸	20

Data that falls into categories, such as different subjects and different kinds of nuts, is called **categorical data**. In a frequency table, there is one row for each category.

Data can also be given as quantities, called **numerical data**. For instance, if we wanted to know how tall a typical 5th grader is, we would need to measure the height of several 5th graders, and the quantities would be given as heights in inches.

When collecting numerical data, it is often easier to make the frequency table after the data is collected. We will see later that is helpful to have the data listed in order, but we usually don't know what numbers we'll get. So, if we make the table after the data has been collected, we can make a row for each value in order.

We can learn a lot by looking at the organized data.

Example:

Given the frequency table shown, what can you conclude about the height of a 5th grader?

HEIGHT (in.)	TALLIES	FREQUENCY
55	|	1
56	||	2
57	𝍸	5
58	𝍸 𝍸 |	11
59	𝍸 ||	7
60	|||	3
61		0
62	|	1

Solution:

We will look at the data to draw conclusions.

We can see that thirty students were measured:

$$1 + 2 + 5 + 11 + 7 + 3 + 0 + 1 = 30$$

The students are all between 55 and 62 inches tall.

Most of the students (23 out of 30) are 57 to 59 inches tall:

$$5 + 11 + 7 = 23$$

About $\frac{1}{3}$ of the students are 58 inches tall:

$\frac{10}{30} = \frac{1}{3}$, and 11 students are 58 inches tall.

So, a typical 5th grade student is around 58 inches tall.

Key point!

Even though there was no value of 61, a space is still shown to make it clear that there was no value of 61 in the data set and to show that there is a gap from 60 to 62.

Sometimes the difference between the smallest data point and the largest can be wide. In a case like this, it is not helpful to make the frequency table before we collect the data because there would be too many rows.

Here is a set of data for how long 5th grade students can hold their breath in seconds:

5, 30, 14, 23, 45, 17, 32, 50, 55, 24, 31, 28, 12, 32, 46, 22, 19, 47, 34, 9, 37, 22, 35, 21, 38

If we made a frequency table, there would be over 50 rows, from 5 to 55! However, we can organize the data in **intervals** of 10 seconds. This will make it easier to see where the data is clustered. First, we'll list the data in order by 10's so we can see how much data is in each interval:

5, 9,
12, 14, 17, 19,
21, 22, 22, 23, 24, 28,
30, 31, 32, 32, 34, 35, 37, 38,
45, 46, 47,
50, 55

Now we can tally the data.

SECONDS	TALLIES	FREQUENCY
0 – 9	||	2
10 – 19	||||	4
20 – 29	𝍸 |	6
30 – 39	𝍸 |||	8
40 – 49	|||	3
50 – 59	||	2

We can see that twenty-five students were tested:

$2 + 4 + 6 + 8 + 3 + 2 = 25$

14 (6 + 8), or a little more than $\frac{1}{2}$ ($\frac{14}{25} > \frac{1}{2}$) held their breath between 20 and 40 seconds. So, we could say a typical 5th grade student can hold his or her breath for about 20 to 40 seconds.

Let's Review!

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

- ✓ For data to be valid, it must be unbiased and from a random sample.
- ✓ Data can be categorical or numerical, and can be organized in a frequency table.

1.1 **Complete this activity.**

Match the terms with their definitions.

a. _____ biased question
b. _____ biased sample
c. _____ categorical data
d. _____ data
e. _____ frequency table
f. _____ interval
g. _____ numerical data
h. _____ population
i. _____ random sample
j. _____ sample
k. _____ statistics
l. _____ survey

1. a sample not representative of the entire population
2. data represented by quantities
3. the collection, organization, and analysis of numerical information
4. a question that leads individuals toward a certain answer
5. a sample in which every member of the population has an equal chance of being selected; unbiased sample
6. information (often numerical)
7. a small part of the population chosen to represent the entire group
8. the distance between numbers on a graph or data display
9. a sampling of a population used to make predictions
10. display that shows data in groups or intervals
11. data grouped using non-numerical criteria
12. all of the possible data in a given topic

1.2 **Circle each correct letter and answer.**

Which example would be likely to give a valid conclusion?

a. Six students are surveyed about their favorite color.
b. People are asked, "Is our mayor doing a good job?"
c. Thirty students are randomly sampled about their eye color.
d. Four students with blond hair are asked about their favorite color.

1.3 If we wanted to know how high a 5th grade student could jump, how many students would be reasonable to test?

a. 5 b. 12 c. 30 d. 200

1.4 Which data set would be numerical?

a. favorite color b. hair color c. place of birth d. height of trees

1.5 How many people were in the survey shown in this frequency table?

a. 4
b. 15
c. 27
d. 30

AGE	TALLIES	FREQUENCY
9	||	2
10	𝍸 𝍸 𝍸	15
11	𝍸 |||	8
12	||	2

1.6 Use the table from Exercise 1.5. What is the typical age of people shown in the frequency table?

a. 9 b. 10 c. 11 d. 12

1.7 Students were asked what day of the week they were born. Which statement is true?

a. Fifth grade students are not born on Saturday.
b. Most fifth grade students are born on Sunday.
c. There is not enough data to draw a valid conclusion.
d. Most fifth grade students are born on Tuesday or Wednesday.

DAY	TALLIES	FREQUENCY
Monday	|	1
Tuesday	|||	3
Wednesday	|||	4
Thursday	||	2
Friday	|||	4
Saturday		0
Sunday	𝍸	5

1.8 Use the table from Exercise 1.7. How many students were surveyed about the day of the week they were born?

a. 7 b. 19 c. 20 d. 25

1.9 If you were trying to find out how far students could jump and you thought that there would be a wide variety of distances, which of the following would you do? (There may be more than one correct answer.)

a. Make the frequency table first.
b. After the data is collected, arrange it in order.
c. Make a row for every data value.
d. Make a frequency table using intervals.

SELF TEST 1: COLLECTING AND ANALYZING DATA

Each numbered question = 6 points

Circle the correct letter and answer.

1.01 Which data set would be numerical?

a. Eye color b. Kinds of pets c. Age d. Favorite food

1.02 Which example would be likely to give a valid conclusion?

a. Five people are randomly sampled and asked about their eye color.

b. Sarah asks the girls on her soccer team what their favorite sport is.

c. In a data set of 20 data points, the mean and median are different and there is no mode.

d. Ten students from each class are randomly sampled to find the favorite subject at school.

1.03 How many people were in the survey shown in this frequency table?

a. 10

b. 27

c. 30

d. 42

AGE	TALLIES	FREQUENCY
9	||	2
10	𝍸 𝍸	10
11	𝍸 |||	8
12	𝍸 ||	7

1.04 What is wrong with this frequency table?

a. There is not enough data.

b. There should be a row for 58.

c. The tallies are not totaled correctly.

d. There is nothing wrong with the table.

HEIGHT (inches)	TALLIES	FREQUENCY
54	|||	3
55	𝍸 |||	7
56	||||	5
57	||	2

1.05 What is the mode for the following set of data? 2, 5, 7, 8, 11, 11, 12

a. 8 b. 10 c. 11 d. no mode

1.06 What is the median for the following set of data? 2, 5, 7, 8, 11, 11, 12

a. 8 b. 9.5 c. 10 d. 11

1.07 What is the mean for the following set of data? 2, 5, 7, 8, 11, 11, 12

a. 8 b. 9 c. 10 d. 11

1.08 What is the range for the following set of data? 2, 5, 7, 8, 11, 11, 12

a. 8 b. 9 c. 10 d. 11

1.09 What is the median of the Hours of TV Watched data set?

a. 4
b. 5
c. 5.5
d. 6

Hours of TV Watched on Saturdays

1.010 Use the data from Question 1.09. What is the mean of the data?

a. 4 b. 5.5 c. 7
d. 9

1.011 Use the data from Question 1.09. What is the outlier for the data set?

a. 4 b. 5.5 c. 9 d. there is none

1.012 What is the best way to display data if there is a wide range and we want to see individual data?

a. stem-and-leaf plot b. frequency table
c. line plot d. frequency table with intervals

1.013 What is the mode of the Heart Rate data set?

a. 60
b. 70
c. 0
d. none

Heart Rate

5	0
6	0 0 0
7	0 0
8	0 0
9	0

5 | 0 = 50 beats per minute

1.014 Use the data from Question 1.013. What is the mean of the data set?

a. 70 b. 68 c.
69 d. 60

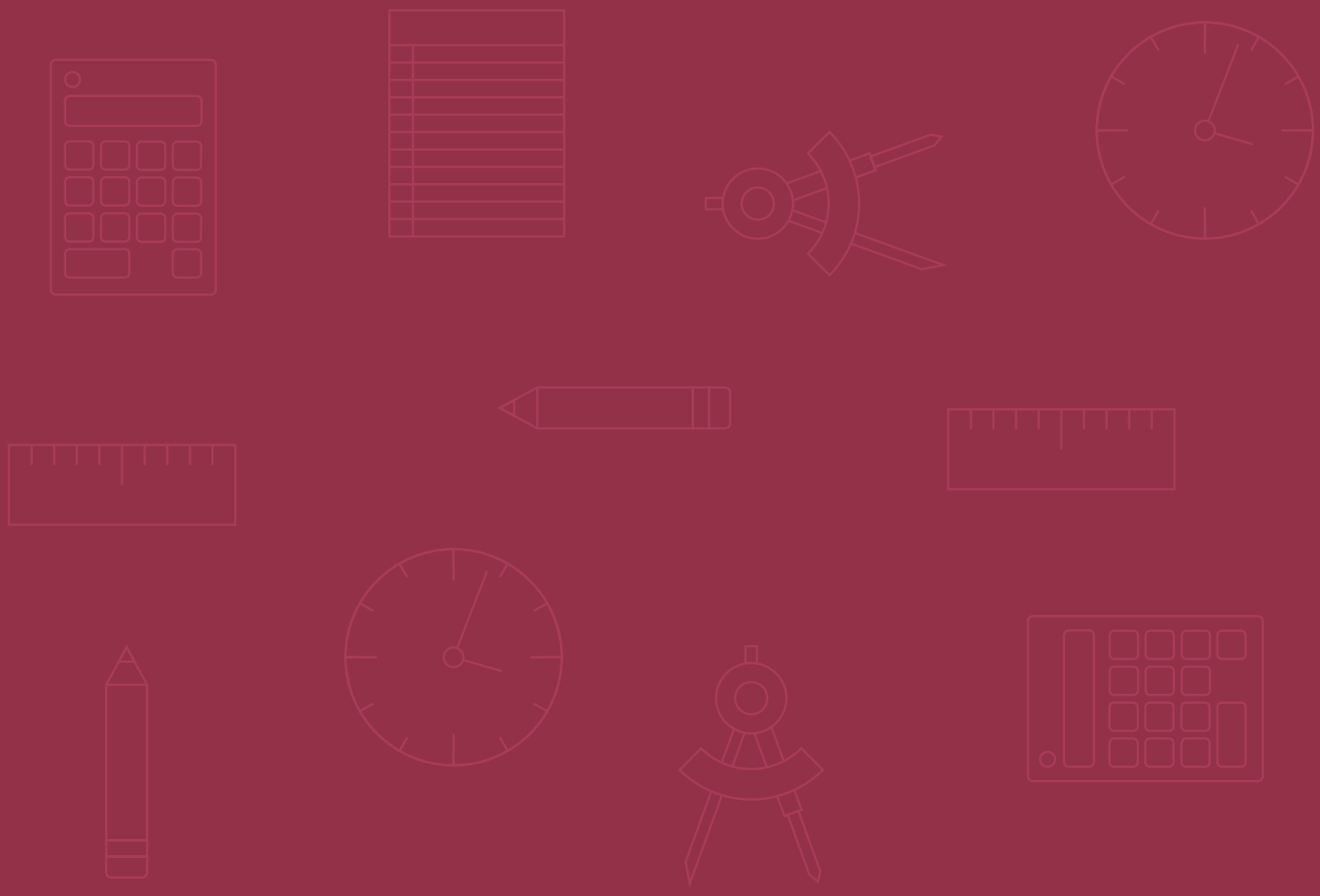
1.015 Use the data from Question 1.013. What is the median for the data set?

a. 40 b. 60 c. 65 d. 70

Teacher check: Initials _____

Score _____ Date _____

72/90



Alpha Omega
PUBLICATIONS

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

800-622-3070
www.aop.com

MAT0508 – Jan '16 Printing

ISBN 978-0-7403-3488-7