



## Lesson 21: Summarizing a Data Distribution by Describing Center, Variability, and Shape

### Student Outcomes

- Given a data set, students are able to describe the data collected, including the number of responses, mean or median, and the MAD or the interquartile range (IQR).

### Lesson Notes

This lesson provides an opportunity for students to summarize a given data set. In the problem set of this module, students are expected to summarize the data collected in Lesson 17 by constructing a poster or an outline of a presentation. This lesson guides students through the four steps used to carry out a statistical study in order to prepare them for the presentation in Lesson 22.

### Classwork

Each of the lessons in this module is about data. What are data? What questions can be answered by data? How do you represent the data distribution so that you can understand and describe its shape? What does the shape tell us about how to summarize the data? What is a typical value of the data set? These questions, and many others, were part of your work in the exercises and investigations. There is still a lot to learn about what data tell us. You will continue to work with statistics and probability in grades seven and eight and throughout high school. You have already, however, started to learn how to uncover the stories behind data.

When you started this module, the four steps used to carry out a statistical study were introduced:

- Step 1:** Pose a question that can be answered by data.
- Step 2:** Collect appropriate data.
- Step 3:** Summarize the data with graphs and numerical summaries.
- Step 4:** Answer the question posed in step 1 using the numerical summaries and graphs.

In this lesson, you will carry out these steps using a given data set.

### Exploratory Challenge: Annual Rainfall in the State of New York (25 minutes)

This is an exploration lesson. Students should be given approximately 25 minutes to work independently in completing the template that provides a structure for summarizing the rainfall data.

Students are given the annual rainfall in inches for New York from 1983 to 2012. The data were obtained from the National Climate Data Center. (If any students need data for their presentation discussed in the problem set, this site provides climate data for regions, cities, and states and could be a source to help students struggling to obtain data.)

Before students organize their summary, discuss the context explained in the lesson. Make sure the students understand what the data represent by highlighting the words *annual* and *rainfall*. Ask students to guess a value that they think represents the typical rainfall for New York in a year. Record these guesses and refer to them after students have summarized this data set and calculated a measure of center. Also, ask students why a statistical study of rainfall is important. For example, when a reporter says that a certain year was unusually rainy, on what basis was that claim made?

Direct students to study the template that is included with this lesson. Review the four steps involved in a statistical study. Indicate that during the next 25 minutes they are expected to complete the template that organizes their statistical summary of the data.

#### Exploratory Challenge: Annual Rainfall in the State of New York

The National Climate Data Center collects data throughout the United States that can be used to summarize the climate of a region. You can obtain climate data for a state, a city, a county, or a region. If you were interested in researching the climate in your area, what data would you collect? Explain why you think this data would be important as a statistical study of the climate in your area.

For this lesson, you will use yearly rainfall data for the state of New York that were compiled by the National Climate Data Center. The following data are the number of inches of rain (averaged over various locations in the state) for the years from 1983 to 2012 (30 years).

45	42	39	44	39	35	42	49	37	42	41	42	37	50	39
41	38	46	34	44	48	50	47	49	44	49	43	44	54	40

Use the four steps to carry out a statistical study using this data.

**Step 1:** Pose a question that can be answered by data.

What is a statistical question that you think can be answered with these data? Write your question in the template provided for this lesson.

**Step 2:** Collect appropriate data.

The data have already been collected for this lesson. How do you think these data were collected? Recall that the data are the number of inches of rain (averaged over various locations in the state) for the years from 1983 to 2012 (30 years). Write a summary of how you think the data were collected in the template for this lesson.

**Step 3:** Summarize the data with graphs and numerical summaries.

A good first step might be to summarize the data with a dot plot. What other graph might you construct? Construct a dot plot or another appropriate graph in the template for this lesson.

What numerical summaries will you calculate? What measure of center will you use to describe a typical value for these data? What measure of variability will you calculate and use to summarize the spread of the data? Calculate the numerical summaries and write them in the template for this lesson.

**Step 4:** Answer your statistical question using the numerical summaries and graphs.

Write a summary that answers the question you posed in the template for this lesson.

The following directions should be considered as students develop a statistical summary of this data. Work with students individually or in small groups as they complete the template.

### Step 1: Pose a question that can be answered by data.

It is important that students are reminded of the two most important parts of the definition of a statistical question. A statistical question is (1) a question that is answered by data, and (2) a question that anticipates the data will vary. As students examine the data, point out to them that there is variability. Although students may vary the wording of their questions, it is anticipated most students will form a question that essentially asks, “What is the typical annual rainfall in New York?”

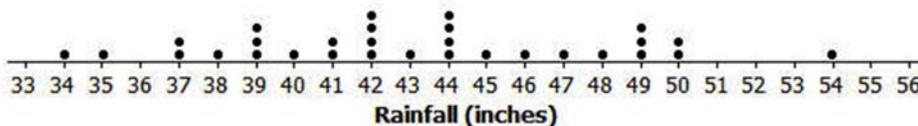
### Step 2: How do you think the data were collected?

The data are given to students in this lesson. This step will be more challenging as they carry out their own statistical study because they need to explain the plan they developed to collect their data. For this lesson, allow students to speculate on how the National Climate Data Center probably collected this data. As the data represent the annual rainfall for the state of New York, the Center had to collect rainfall totals from several reporting weather centers around the state. They calculated an average of those levels for each day of the year. At the end of the year, the National Climate Data Center averaged those daily results. Students might be asked how a rainfall level is measured at a weather center. A rain gauge might be a good visual to share with students.

### MP.4 Step 3: Construct graphs and calculate numerical summaries of the data.

This step represents most of the work students will be expected to do in this lesson. As a first step, encourage students who are not sure how to start summarizing the data to construct a dot plot. A blank grid is provided at the end of the Teacher Notes that can be duplicated for students who may need some structure in making a dot plot. This grid could also be used if any student decides to develop a box plot or a histogram of the data distribution.

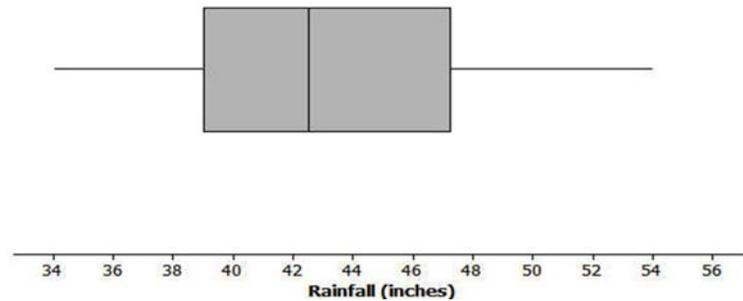
Dot plot of annual rainfall from 1983 to 2012



Students examine the dot plot and make decisions about the data distribution. For example, is the distribution approximately symmetric or is the distribution skewed? This dot plot shows a data distribution that is approximately symmetric.

Based on the decision that the distribution is approximately symmetric, students should proceed to calculate the mean as a measure of center and the MAD as a measure of variability. Some students might also choose to investigate this distribution with a box plot to answer the question about the symmetry.

Box plot of annual rainfall from 1983 to 2012



The box plot does not indicate a perfectly symmetrical distribution; however, it is approximately symmetric. The mean and the median of this data distribution are nearly equal to each other. Based on this decision, students should calculate the mean and the MAD.

The mean rainfall is 43 inches (to the nearest inch), and the mean absolute deviation (MAD) is 3.9 or 4 inches (to the nearest inch).

At the end of the Teacher Notes is a table that could be used for students who need structure in calculating the MAD. There are several steps in calculating the MAD that might need to be organized for some students.

#### Step 4: Answer your statistical question using the numerical summaries and graphs.

MP.3

This step asks students to write a short summary interpreting the graphs and numerical summaries. Students should connect this back to their statistical question. Students would indicate that the typical rainfall for New York is 43 inches per year. They would also indicate that a typical deviation from the mean is about 4 inches.

#### Template for Lesson 21

Step 1: What is your statistical question?

Step 2: How do you think the data were collected?

Step 3: Construct graphs and calculate numerical summaries of the data.

Construct at least one graph of the data distribution. Calculate appropriate numerical summaries of the data. Also indicate why you selected these summaries.

Step 4: Answer your statistical question using your graphs and numerical summaries.

**Closing (5 minutes)**

If time permits, look at the original guesses students made to this question. Did students have a pretty good idea of the annual rainfall in New York? Discuss this question with students.

**Lesson Summary**

Statistics is about using data to answer questions. The four steps used to carry out a statistical study include posing a question that can be answered by data, collecting appropriate data, summarizing the data with graphs and numerical summaries, and using the data, graphs, and summaries to answer the statistical question.

**Exit Ticket**

Consider a special type of Exit Ticket for this lesson. As students are expected to complete a summary of the four-step investigative study, use this opportunity to assess your students' understanding of this process as related to the question they formed and the data they collected. The Exit Ticket for this lesson is to complete the following direction (state this direction to the students):

*Based on your current preparation, summarize the four steps you are expected to complete as part of presenting a statistical study.*



Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 21: Summarizing a Data Distribution by Describing Center, Variability, and Shape

### Exit Ticket

Based on your current preparation, summarize the four steps you are expected to complete as part of presenting a statistical study.

## Exit Ticket Sample Solutions

Based on your current preparation, summarize the four steps you are expected to complete as part of presenting a statistical study.

**Step 1:** *State my statistical question. My question is based on collecting data that will vary.*

**Step 2:** *Devise a plan to collect data. I prepared a question to ask the students in my class. (Allow students to explain the question they asked, the responses they received, and the method they used for recording answers.)*

**Step 3:** *Summarize my data. I prepared a dot plot of the responses to the question. My dot plot indicated that the responses to my question were skewed to the left; therefore, I used the median of the data distribution to describe my center and the IQR to describe the variability. (Allow for a summary of the specific median or mean, and a specific summary of the variability as the MAD or IQR.)*

**Step 4:** *Based on my graphs and numerical summaries, I answered my question.*

## Problem Set Sample Solutions

The problem set for this lesson involves creating a poster or an outline for a presentation using the data collected in Lesson 17. The directions in the lesson indicate that students are expected to carry out the four steps either on their poster or outlined for a presentation. If students provided an adequate summary of the four-step process in the exit ticket, they could use their summary as a guide in completing the poster. Highlight the following with students:

For Step 1, students are expected to have a question clearly identified as their statistical question. The question should involve the data they collected. Students should have anticipated variability in the data.

For Step 2, students should indicate how they collected the data based on the plan proposed in Lesson 17. For example, for a question that investigates a typical height of students in the class, did every student state his or her height in inches or was there a way to measure everyone's height? For a question that investigates how many books students read, did students ask members of their class how many books they read each month?

For Step 3, students include graphs and numerical summaries of the data. (Again, if students need more structure in constructing their graphs, provide them with the coordinate grids for this lesson.) It is anticipated that students begin with a dot plot. From the dot plot, students might construct a box plot or a histogram. Based on the shape of the distribution, students select appropriate numerical summaries—either the mean and the mean absolute deviation (MAD) or the median and the interquartile range (IQR). Posters or outlines should indicate what summaries were used and why.

For Step 4, students should have a concluding statement that answers the statistical question. Students should provide a brief description of their numerical summaries and graphs.

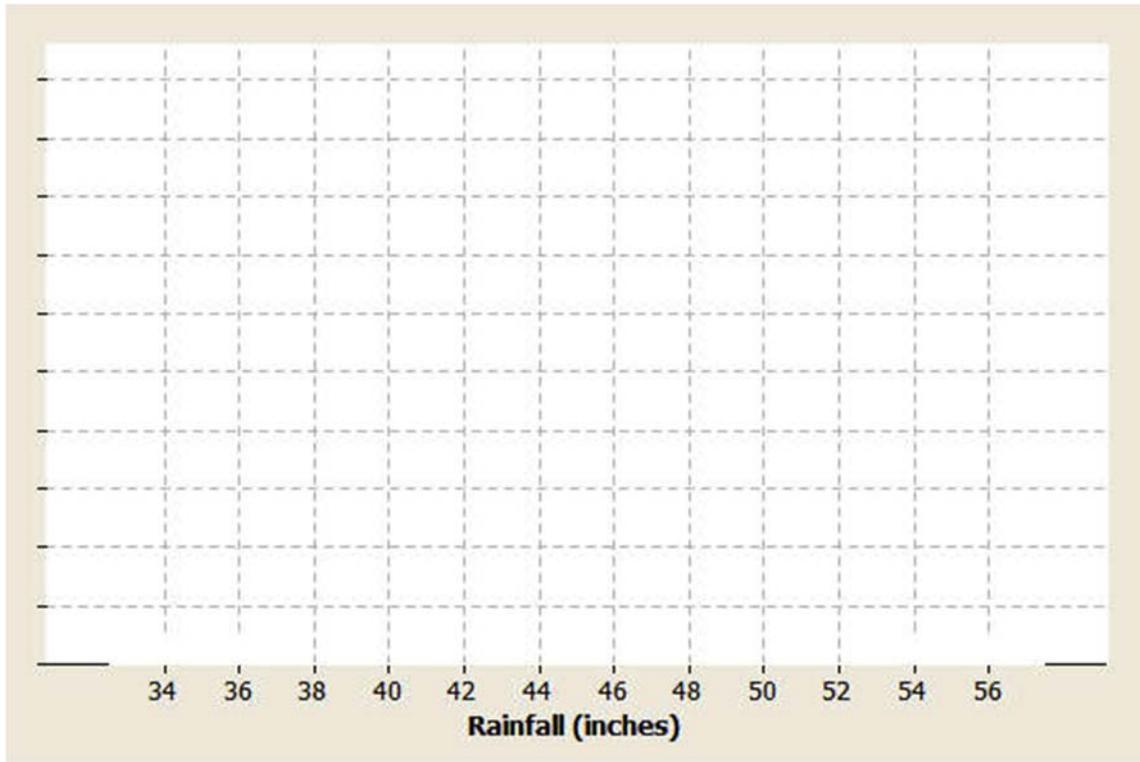
In Lesson 17, you posed a statistical question and a plan to collect data to answer your question. You also constructed graphs and calculated numerical summaries of your data. Review the data collected and your summaries.

Based on directions from your teacher, create a poster or an outline for a presentation using your own data. On your poster, indicate your statistical question. Also, indicate a brief summary of how you collected your data based on the plan you proposed in Lesson 17. Include a graph that shows the shape of your data distribution, along with summary measures of center and variability. Finally, answer your statistical question based on the graphs and the numerical summaries.

Share the poster you will present in Lesson 22 with your teacher. If you are instructed to prepare an outline of the presentation, share your outline with your teacher.

**Additional Resource Materials**

The following could be used to provide structure in constructing a dot plot, histogram, or box plot of the rainfall data. A similar type of grid (or graph paper) could be prepared for students as they complete the problem set. The grid provided for students should not include the units along the horizontal axis as that is part of what they are expected to do in preparing their summaries.



The following table could be used for students requiring some structure in calculating the mean absolute deviation, or MAD:

Data Value	Mean	Data Value – Mean	Absolute Value of Data Value – Mean
45			
42			
39			
44			
39			
35			
42			
49			
37			
42			
41			
42			
37			
50			
39			
41			
38			
46			
34			
44			
48			
50			
47			
49			
44			
49			
43			
44			
54			
40			