Section 3

Describing Distributions

Investigation 3.1
How Many Pockets?

Overview

This investigation focuses on students collecting, analyzing, and interpreting numerical data. After listening to the story *A Pocket for Corduroy*, by Don Freeman, students generate questions about how many pockets they have in their clothes. Students count the number of pockets in their clothing. Data by the class are collected on sticky notes and organized, displayed, and compared in a **dotplot**.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level A activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Generate questions about the number of pockets in their clothing
- Organize and display data in a dotplot
- Analyze data and record observations
- Compare two groups of data using dotplots

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

K.CC.7 Compare two numbers between 1 and 10 presented as written numerals.

Note: The Common Core State Standards do not include the dotplot as a statistical graph until the sixth grade. However, the writers of this publication feel a dotplot is an excellent visual representation of organizing numbers between 1 and 10 and have included its use in this investigation at this beginning grade level.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Pre-K–2 All students should pose questions and gather data about themselves and their surroundings; sort and classify objects according to their attributes and organize data about the objects; and represent data using concrete objects, pictures, and graphs.

Materials

- Grid chart paper
- Sticky notes (2 colors – green and yellow)

Estimated Time

One day

Instructional Plan

✏ Formulate a Statistical Question

1. Read *A Pocket for Corduroy*, by Don Freeman.

   In this story, Lisa and her mother go to the Laundromat. Lisa brings her stuffed bear, Corduroy. Corduroy sees Lisa searching through her pockets before putting clothes in the washer. Corduroy notices he doesn't have a pocket and goes in search of one. He gets lost and spends the night in the Laundromat. The next day, Lisa returns, takes him home, and makes a pocket for him.

2. After reading the story, ask your students if they share Corduroy's desire for pockets in their clothes. Ask your students if they have pockets in the clothes they are wearing today. How many pockets do they have? Do boys and girls have the same number of pockets in their clothes? Discuss if you want your students to count their indoor clothes and outdoor clothes, or just their indoor clothes. Discuss what counts as a pocket.

The remainder of this investigation focuses on the question, "Do boys and girls have the same number of pockets in their clothes?"

Before collecting data, you may want to ask your students to raise their hands if they think girls have more pockets. Then, ask your students to raise their hands if they think boys have more pockets. Let students know they were making a prediction when they raised their hand and that they will use their number of pockets as data to see if their prediction about whether boys or girls have more pockets was correct.

Collect Appropriate Data

1. Have each student count the number of pockets in their clothing and record the number on a sticky note. Boys should write their number on one color of sticky note and girls on a different color of sticky note.

2. Draw a number line on the grid paper. The scale should start at 0 and go to the largest number of pockets observed by your students. The numbers should be equally spaced on the horizontal axis. Label the horizontal axis *Number of Pockets*.

3. Have students place their sticky note (number of pockets) above the number on the number line. To begin with, put all notes—whether from boys or from girls—vertically above the appropriate number. Figure 3.1.1 shows an example of a dotplot of the number of pockets from a group of first-graders.

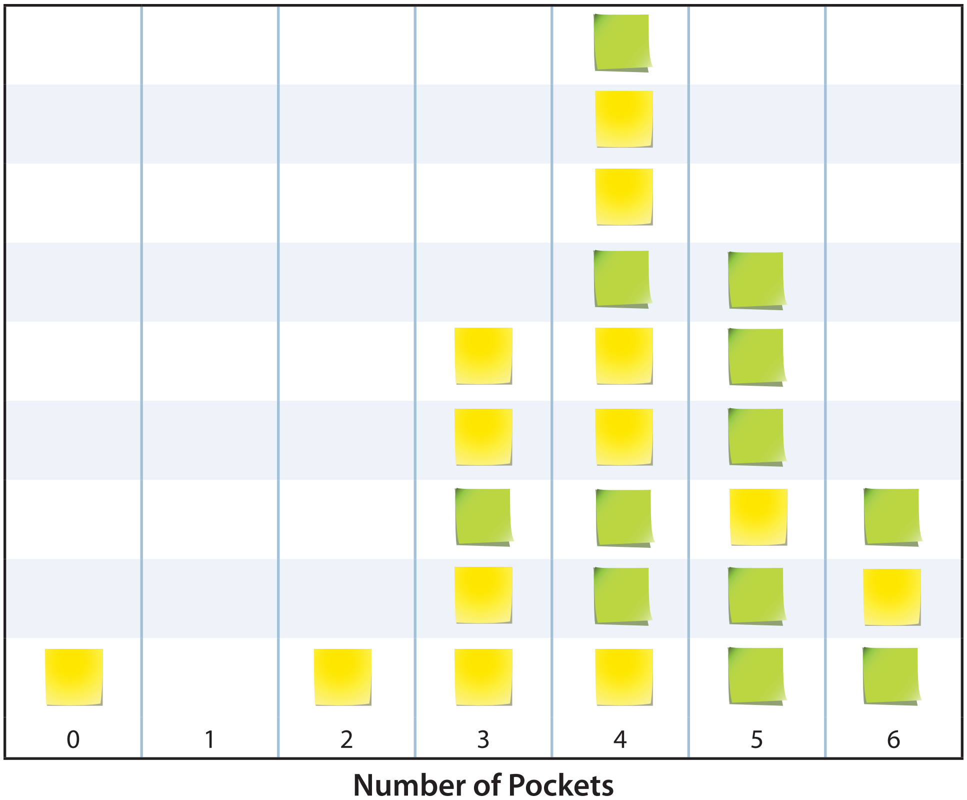


Figure 3.1.1 Dotplot of number of pockets

Analyze the Data

1. Ask the following questions:

   a. What is the lowest number of pockets? This is called the minimum.

   b. What is the highest number of pockets? This is called the maximum.

   c. What is the number of pockets that occurs most often? This number is called the mode. (The mode is four pockets in this example. It occurs nine times.)

2. Ask your students if boys and girls have about the same number of pockets. Because the data are mixed, suggest they make two graphs—one for the boys and the other for the girls. For comparison purposes, the two scales should be the same.

3. Draw a number line on another sheet of grid chart paper. Place this sheet next to the graph already constructed. Take the boys' sticky notes from the first graph and transfer to the new graph. The girls' sticky notes may need to be adjusted on the first graph once the boys' sticky notes have been removed. Note that the use of grid chart paper helps keep the

rows lined up. Otherwise, it is easy to misrepresent comparing column heights. See Figures 3.1.2 and 3.1.3.

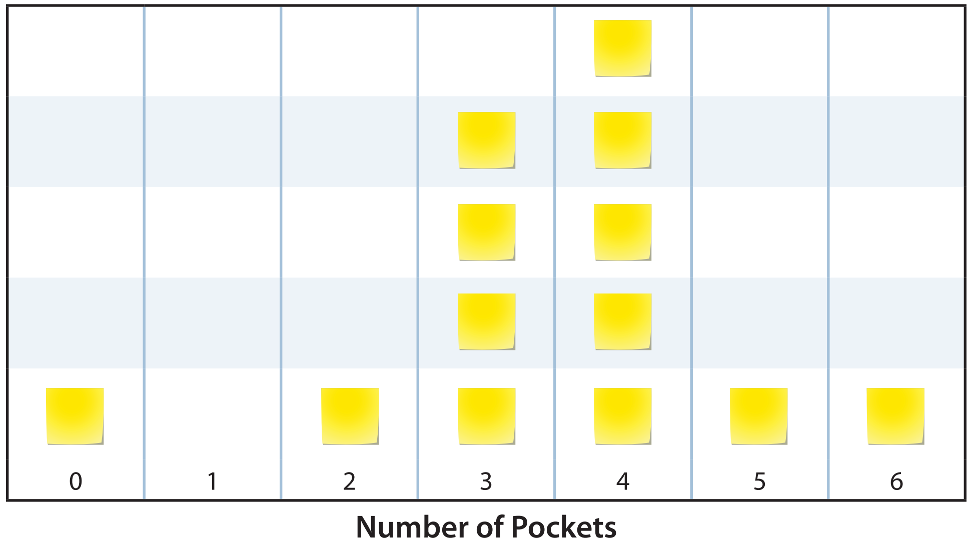


Figure 3.1.2 Dotplot of number of pockets for girls

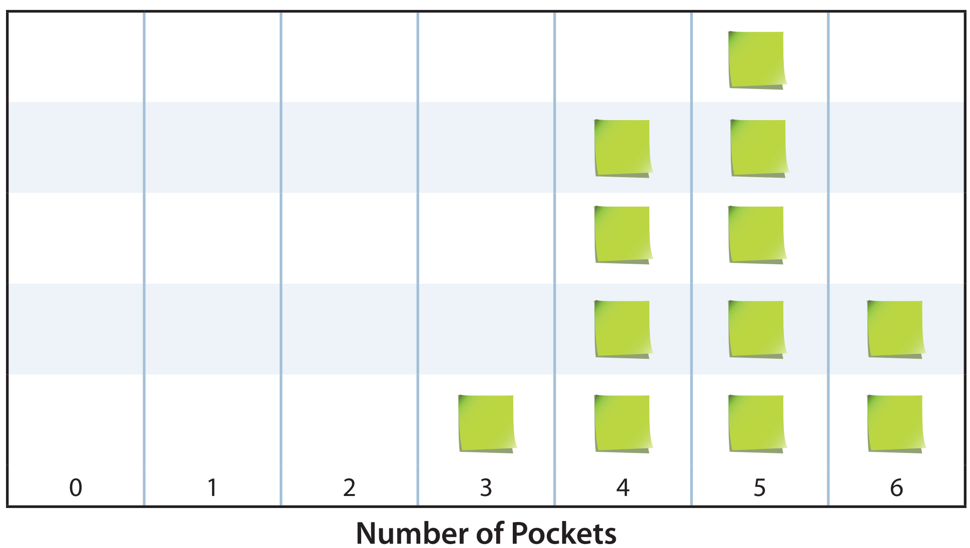


Figure 3.1.3 Dotplot of number of pockets for boys

4. Ask students the following questions:
   a. What is the minimum number of pockets for the girls? For the boys?
   b. What is the maximum number of pockets for the girls? For the boys?
   c. What is the mode number of pockets for the girls? For the boys?

d. How many boys have four or more pockets? What fraction of the boys have four or more pockets?

e. How many girls have four or more pockets? What fraction of the girls have four or more pockets?

Interpret the Results in the Context of the Original Question

1. Have your students recall that the original question was, "Do boys and girls have the same number of pockets in their clothes?" Have your students write a summary that begins with answering the question and then justifying their answer based on the analysis of the data conducted in class. They should include the data and graphs. They also should discuss how their prediction about whether boys or girls have more pockets compares to the actual data they collected.

2. Ask your students if they think their parents or grandparents have more or fewer pockets than they do? Why or why not?

3. Ask students how they might go about finding the answer to Question 2.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

After we listened to the story *A Pocket for Corduroy*, we became curious about the number of pockets we have on our clothes and whether girls or boys have more. So, on a green sticky note, each boy in our class wrote the number of pockets he had and the girls did the same thing, except their sticky notes were yellow. Then, we put them all in a dotplot. When we tried to answer our statistical question about whether girls or boys had more pockets, it was a little hard to do with all the green and yellow sticky notes together. So, to make it easier to answer our question, we drew two dotplots with the same scale, one for the boys and one for the girls.

By comparing the two dotplots, we concluded that, overall, boys have more pockets. We had to be careful because the number of boys and the number of girls was not the same. There were 12 boys and 13 girls. Still, 11 out of the 12 boys had four or more pockets; that's way over half of the boys. But 7 of the 13 girls, about half of them, had four or more pockets, so we concluded from our analysis that boys have more pockets than girls. We're going to continue this activity by asking our parents and grandparents how many pockets they usually have on their everyday clothes.

Assessment with Answers

A group of students counted the number of pockets in their clothes and drew the two graphs shown below.

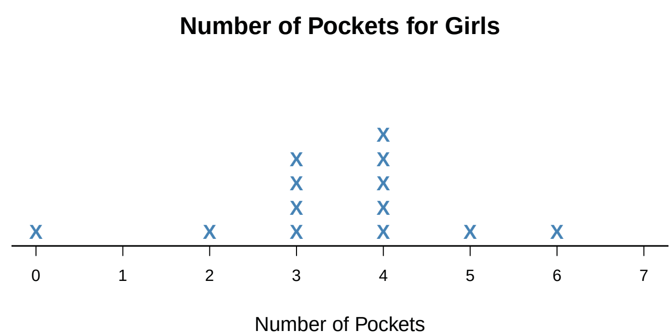


Figure 3.1.4 Dotplot of number of pockets for girls

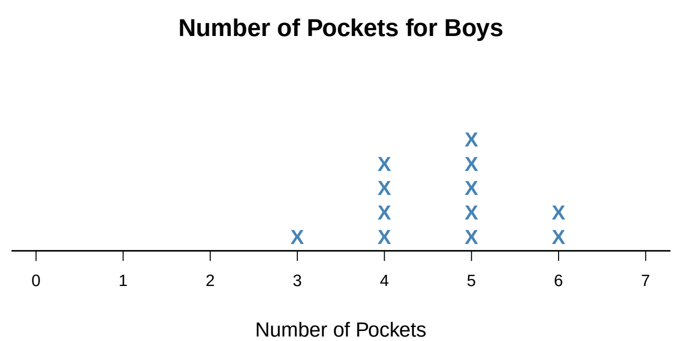


Figure 3.1.5 Dotplot of number of pockets for boys

1. What is the minimum number of pockets for the girls? The minimum number of pockets for the girls is zero.

2. What is the maximum number of pockets for the boys? The maximum number of pockets for the boys is six.

3. What is the mode number of pockets for the girls? The mode number of pockets for the girls is four.

4. How many more girls had four pockets than boys who had four pockets? There were five girls with four pockets each and four boys with four pockets each, so there was one more girl who had four pockets.

5. Who has more pockets, boys or girls? Use words, numbers, and graphs to explain your answer. Boys have more pockets than do girls. There are seven girls who have four, five, or six pockets, but there are 11 boys who have four, five, or six pockets. Every boy has three or more pockets, but there are two girls who have fewer than three pockets.

Extensions

1. Have your students collect data on the number of eyelets on their shoes. Students can compare the number of eyelets with the type of shoes. For example, compare the number of eyelets in gym shoes versus dress shoes.

2. Have your students ask their parents or an adult how many pockets they have. Have the students make a bar graph of the number of pockets from the adults and then compare those results with the class results.

3. Science connection: Have your students investigate what animals have pockets or pouches.

References

Chapin, S., A. Koziol, J. MacPherson, and C. Rezba. 2003. *Navigating through data analysis and probability in grades 3-5*. Reston, VA: National Council of Teachers of Mathematics.

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

Freeman, D. 1978. *A pocket for corduroy.* New York, NY: Penguin Group Publishers.

Greenes, C. E. (ed.) 2002. *Navigating through data analysis and probability in prekindergarten – grade 2*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics. www.corestandards.org.

Investigation 3.2
Who Has the Longest First Name?

Overview

This investigation is based on one found in the Appendix for Level A in *Guidelines for Assessment and Instruction in Statistics Education (GAISE): A Pre-K–12 Curriculum Framework*. During the first week of school, a third-grade teacher is trying to help her students learn one another's names by playing various games. During one of the games, a student named MacKenzie noticed that she and her classmate Zacharius each have nine letters in their names. MacKenzie conjectured that their names were longer than everyone else's names, which gave the teacher an opportunity to introduce a statistics lesson.

In this investigation, students analyze the length (number of letters) of their first names. The data will be organized and displayed in **dotplots** to develop the **median** as a measure of center and the **range** as a measure of variability of first name lengths.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level A activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Collect data and organize the results in a dotplot
- Find measures of center (median and mode) for the data
- Consider what measures of center are appropriate for categorical versus quantitative data (addressed in the extensions)
- Find a measure of spread (range) for the data

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

6.SP.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

6.SP.2 Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape.

6.SP.3 Recognize that a measure of center for a numerical data set summarizes all its values with a single number, while a measure of variation describes how its values vary with a single number.

6.SP.4 Display numerical data in plots on a number line, including dotplots, histograms, and box plots.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 3-5 All students should design investigations to address a question and consider how data collection methods affect the nature of the data set; represent data using tables and graphs such as line plots, bar graphs, and line graphs; use measure of center, focusing on the median, and understand what each does and does not indicate about the data set.

Materials

- Sticky notes
- Masking tape

Estimated Time

Two days

Instructional Plan

Formulate a Statistical Question

1. Have your students discuss what they would like to know about their first names. Make a list of their responses. The following are some possible responses:

   Do any of us have the same name?

   What is our most common first name?

   Who has the longest name?

   Who has the shortest name?

   What is the most common first letter?

   What is the most common length for our first names?

2. This investigation focuses on the statistical question, "How do the lengths of first names vary in our class?"

Collect Appropriate Data

1. Hand out a sticky note to each student. Have your students write their first name on a sticky note, as well as the number of letters in their first name. For example:

2. Have your students place their sticky notes on the board in no particular order. They are placed randomly so that the class will be able to observe

how a dotplot can be used to organize the notes. Figure 3.2.1 is an example of the data for one class.

Figure 3.2.1 Example of sticky notes with names and lengths

✏ Analyze the Data

Have your students discuss how they may be able to organize all these sticky notes. They may suggest grouping the notes by the number of letters, as shown in Figure 3.2.2.

Figure 3.2.2 Example of sticky notes organized by number of letters

Reorganize the sticky notes into a preliminary dotplot. Note that, as much as possible, there should be no gaps or overlaps between each note in a column. See Figure 3.2.3.

Figure 3.2.3 Dotplot of length of first names

Ask the following questions:

a. What is the shortest name? How many letters does it have?

b. What is the longest name? How many letters does it have?

c. Is it possible for someone to have a name shorter than our shortest name?

d. Is it possible for someone to have a name longer than our longest name?

e. What is the range of the data? The range of the data is the largest number minus the smallest number. (9 – 3 = 6 for this example)

f. What is the most common number of letters in the first names for our class? This value is called the mode. (6 for this example)

To introduce the concept of the median, students need to be lined up side by side according to the number of letters in their name. Note that it might be helpful to have your students form a human dotplot first. That will get all the students together who have the same number of letters.

a. Put a piece of masking tape on the floor labeled with numbers from the smallest number of letters to the largest number of letters in their names.

Figure 3.2.4 Labeled masking tape

b. Have your students line up corresponding to the length of their name (e.g., Josh and Ella would both line up above the 4) in either order. Once they have created a human dotplot, have them form a single-file line side-by-side, keeping them in order based on the length of their name.

c. Ask your students who they think is exactly in the middle of this display.

d. Have one student each at either end of the line sit down until one student is left standing in the middle. If there is an even number of students in the class, you may want to include yourself in the display to ensure students will be finding the median of an odd number of data points. For example, have the first student above 3 (Zak) and the last student above 9 (Octavious) sit down at the same time. Have the second student above 3 (Sam) and the second-to-last student above 9 (Christian) sit down at the same time. Continue this process until you are left with one person standing. The number of letters in that person's name is the median of the data set. Note that if the number of students is even, there will be two students left standing. The median is the value half-way between their values.

e. Ask your students how they could use the dotplot on the board with the sticky notes to model what they did in the human dotplot in order to find the median length of their names. Have your students make suggestions. They should realize that they could either remove sticky notes one at a time from both ends until one was left or put an X through the notes until they get to the middle one, in the case of an odd number of sticky notes.

f. A new student arrives after the above has been completed. Her name is Seraphinia. How would adding her to the data set affect the center measures of mode and median and the spread measure of range? Note that Seraphinia has 10 letters, so the mode remains at 6. Also note that, with Seraphinia, there is now an even number of names. An even number of data points creates two "middles." The median is taken to be the value half-way between the two middles. In this example, both middle values are 6, so the median remains at 6. The range is now 10 – 3 = 7, an increase of one letter.

Interpret the Results in the Context of the Original Question

1. Have your students recall the original question: How do the lengths of first names vary in our class? Ask them to write a report that answers the question, along with providing a justification of it using their analysis.

2. Ask your students the following questions:

    a. Based on our data, what is the most common or typical length of our first names?

    b. Do you think this would be the same in the classroom across the hall? Why or why not?

    c. Do you think this would be the same in a middle- or high-school classroom? Why or why not?

    d. Do you think this would be the same in Mexico? In China?

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

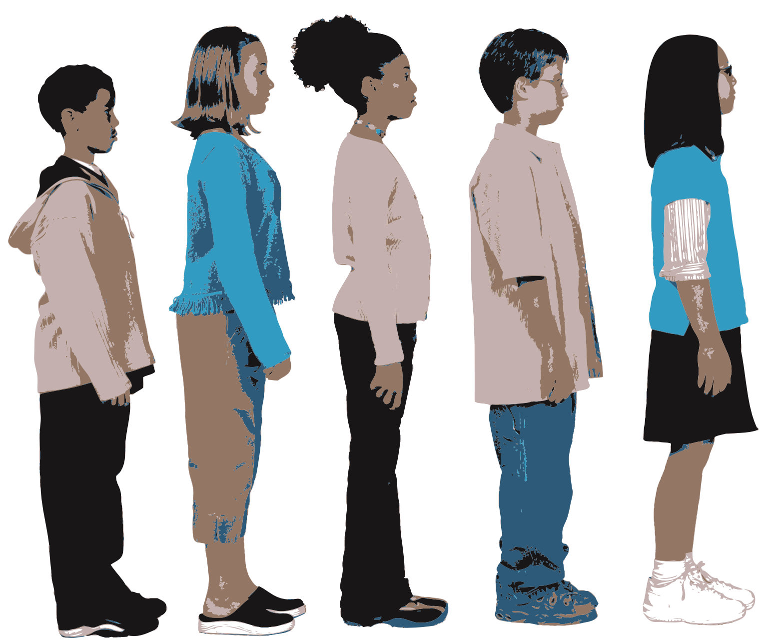
On the first day of school, our teacher had us play games to learn each other's names. After that, she was showing us some statistics by having us study the lengths of our first names. The question was, "How do the lengths of first names vary in our class?"

We used sticky notes with our names and the number of letters in our names written on them. After putting all of the sticky notes on the board all messed up, we organized them into a dotplot with the number of letters on the horizontal axis. But, we didn't do the graph the right way the first time, because we didn't keep the columns nice and straight and in line with the other columns. We had to remember to keep the rows in line, also. When we corrected that, we saw that there were more of us whose first names had six letters than any other number. It was the highest in the dotplot. That's called the mode number of letters.

We also calculated the middle number of letters by lining up from fewest number of letters to most and then having low and high sit down until we got to one person left. That number is called the median. It's the middle number of letters, 6 (Alicia), with 12 of us below Alicia and 12 of us above Alicia.

The day after we did that analysis, we got a new student in class, Seraphinia. When we added her, she had the longest name. The range of letters was 9 – 3 = 6 before Seraphinia, but 10 – 3 = 7 letters with her. The mode stayed at 6 because it was still the highest. To find the median, we sat down like before, but now there were two middles, Alicia and Connor. They both have 6 letters in their names, so the median is still 6.

We want to continue doing this study by looking at names from different countries to see if their number of letters differs from ours. We think that maybe Chinese names are shorter.



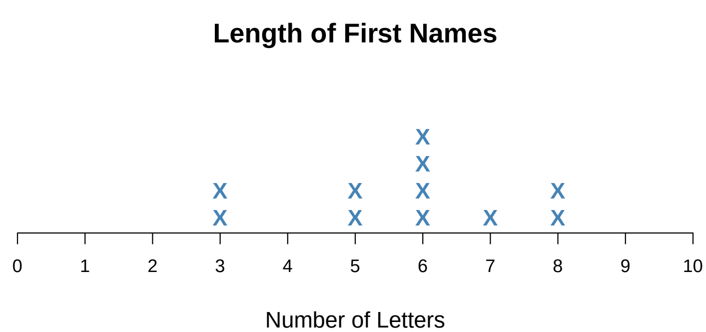
Assessment with Answers

With the help of his family and friends, Jose collected data regarding the lengths of first names of his family and friends. Table 3.2.1 shows the data Jose collected.

Table 3.2.1 Length of First Name

	Family and Friends First Names	Number of Letters in First Name
1	Hector	6
2	Amada	5
3	Che	3
4	Ricardo	6
5	Camila	6
6	Roberto	7
7	Carlos	6
8	Raymundo	8
9	Gabriela	8
10	Diego	5
11	Tia	3

1. Make a dotplot of the length of the first names of Jose's family and friends.



2. Find the value of each of the following:

   Maximum value: 8

   Minimum value: 3

   Mode: 6

   Median: 6

   Range: 5

3. Write a summary of what you observed about the length of the first names of Jose's family and friends. Your summary should include reference to the dotplot and the measures of center and spread that you found. The dotplot shows that all of the name lengths are from 3–8 letters long. The most common length was 6 letters (four people had this length) and the median length was also 6 letters, meaning there were five names above the median (6, 6, 7, 8, 8) and five below (3, 3, 5, 5, 6).

Extensions

1. Have your students do an analysis of the length of their last names.

2. Your students may have had the experience of having to write their full name (first, last, and space between) on a form that has a set of boxes. Have your students combine the length of their first and last names, including the space between their names, and investigate how many boxes the form should have so most of the students in class can fill in their entire name.

   Note: Discuss with your students what they feel "most of the students" means in this context. Some students will think just over half, while others will want to say all except for one or two students. Some may even interpret it as "most often" and want to use the mode. Allow students to answer the question based on their definition of "most."

3. Students may be tempted to find the median of a categorical set of data. The following exercise will demonstrate why the median is a measure of center for numerical data only.

   a. Consider the statistical question, "What is your favorite type of pet?" Put a piece of masking tape on the floor labeled as follows:

Figure 3.2.5 Favorite pet

Have your students line up according to their favorite type of pet, thus forming a human bar graph. They have to choose only one pet category.

b. Ask your students which type of pet is the most popular? Who thinks their favorite type of pet is in the middle?

   Note: The first question is asking for the mode, while the second question is designed to begin a discussion about why categorical data do not have a median. The following parts, c through e, should help with the discussion.

c. Recalling the process of creating a human dotplot to determine the median length of first names, have your students at either end sit down until one is left standing in the middle (two if the number of students is even). Ask your students what the median is. Students might respond with the name of the student or the type of pet.

d. Change the order in which the pets are listed on the floor and have your students make a new bar graph.

Figure 3.2.6 Favorite pet

e. Have your students find the middle again by sitting down starting at each end. Ask what the median or middle is. It is likely that this answer will be different from the first bar graph. Discuss with your students that changing the order of the categories has changed the median response, but the data are exactly the same in both cases. Have them do the process a third time. Yet another median appears for the same data. Discuss with your students that a median cannot have different values for the same data set. So, finding a median for categorical data does not make sense. To find a median, data must be able to be ordered from smallest to largest. That can be done for numerical data, but it cannot be done for categorical data.

4. Ask the class to work together to determine if the following data sets can have a median: number of pets, eye color, number of siblings, ways to get to school. Then, have your students work in pairs to create an example of a data set that could have a median and one that could not have a median. They should write a sentence or two with their examples, explaining why their data set can or cannot have a median.

References

Bereska, C., L. C. Bolster, C. A. Bolster, and R. Scheaffer. 1998. *Exploring statistics in the elementary grades: Book one, grades k–6*. White Plains, NY: Dale Seymour.

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics. www.corestandards.org.

Investigation 3.3
How Expensive Is Your Name?

Overview

This investigation focuses on developing the **mean** as a measure of center. In addition, students will compare two sets of numerical data and build on the objectives in Investigation 3.2, "Who Has the Longest First Name?" Students will collect data regarding the amount of money it would cost to monogram their first name on a shirt. The mean, median, and mode will be compared for the number of letters in their first name and for the total worth of their first name. Students will be asked to focus on the shapes of the distributions and the idea that although the scale of the distributions shifted, the shape of the distributions remained the same.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Collect data and organize the results in a dotplot
- Find measures of center (mean, median, and mode) for the data
- Compare two distributions by using measures of center and the shapes of the distributions

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

6.EE.2 Write, read, and evaluate expressions in which letters stand for numbers.

6.SP.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

6.SP.2 Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape.

6.SP.3 Recognize that a measure of center for a numerical data set summarizes all its values with a single number, while a measure of variation describes how its values vary with a single number.

6.SP.4 Display numerical data in plots on a number line, including dotplots, histograms, and boxplots.

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 3-5 All students should describe the shape and important features of a set of data and compare related data sets, with an emphasis on how the data are distributed.

Grades 6-8 All students should formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population; find, use, and interpret measures of center and spread, including mean and interquartile range.

Materials

- Sticky notes
- Interlocking cubes (e.g., Unifix cubes)
- Chart paper

Estimated Time

Two days

Instructional Plan

Formulate a Question

1. Before beginning this investigation, display on the board the dotplot of the number of letters in your students' first names. This plot was made in Investigation 3.2, "Who Has the Longest First Name?" If you do not have this graph, you should collect the length of each student's first name and make a dotplot of the data. See Figure 3.3.1.

Figure 3.3.1 Dotplot of length of first names

2. Begin this investigation by asking your students where they have seen a person's name printed on a shirt or jersey. Share with your students that some people have their initials monogrammed or sewn on their dress shirts or a piece of luggage. Explain to your students that, in many of those cases, there is a charge to sew the letters on the shirt. Suggest to your students that you wonder how much it would cost to have their names sewn onto a school T-shirt. Tell your students to assume it costs 5¢ per letter to sew their name on a T-shirt.

3. Explain to your students that the statistical question they will be focusing on is "How expensive is it to monogram first names onto T-shirts?" and how this question is related to the question from Investigation 3.2, "How do the lengths of first names vary in our class?"

4. Refer to the dotplot (Figure 3.3.1) of the length of your students' first names and explain to your students that they are going to make a dotplot of the cost of sewing their first name on a T-shirt. Ask your students how this new plot will be similar and different from the plot on the board.

Collect Appropriate Data

1. Have your students write their names on a sticky note, along with the cost for sewing on their first name. (See Figure 3.3.2.)

2. Have your students put their sticky notes on the board. Do not attempt to organize the sticky notes at this time.

Figure 3.3.2 Sticky note of first name and cost of sewing on first name

Analyze the Data

1. Have your students discuss how the sticky notes could be organized. Encourage them to group the notes by cost.

2. Arrange the sticky notes to create a dotplot. Place this plot alongside the "length of name" plot. Figure 3.3.3 is an example of a dotplot

based on the names shown in Investigation 3.2, “Who Has the Longest First Name?”

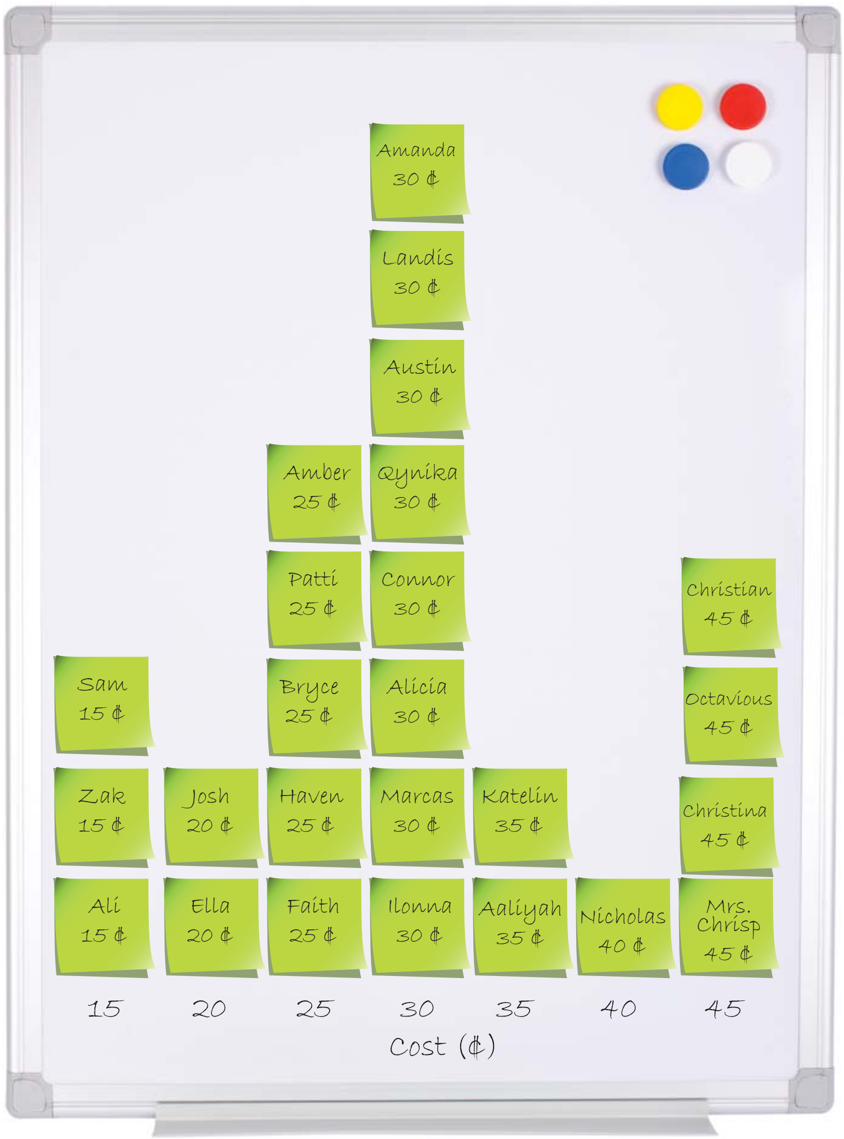


Figure 3.3.3 Dot plot of the cost of sewing on letters

3. Refer to the dotplot of the cost of sewing on the names and ask the following questions:

   a. Who has the least expensive name to sew on? How much does this name cost to sew on? Are these the same students who had the shortest names?

b. What is the most expensive name to sew on? How much does this name cost to sew on? Are these the same students who had the longest names?

c. What is the most common cost to sew on a name (i.e., the **mode**)? Note that the mode is 30 cents. Some students will incorrectly want to answer 8, the frequency of the mode.

d. How does the mode for the cost of sewing on the letters compare to the mode for the number of letters in your first names? Note that the mode of the cost of sewing on the letters data is five times the mode of the length of name data.

4. What is the **median**, or middle number, for the length of name plot? **Note:** If you have not discussed how to find the median of a set of data, see Investigation 3.2, "Who Has the Longest First Name?"

5. What is the median (middle) cost for sewing names?

6. Ask your students how the median cost for sewing on a name compares to the median number of letters in first names. **Note:** The median of the sewing cost data will be five times the median of the length of name data.

7. Introduce the concept of **mean** by developing the "fair share" sense of measuring the center of the data set of number of letters in first names.

   a. The data set in this example has 25 students. Divide these students into five groups of four students, with the sixth group having five students.

   b. Instruct your students that they will need as many interlocking cubes as there are letters in their names. For example, Josh will need four interlocking cubes because his name has four letters.

   c. You want your students to discover that another measure of center for a set of data (in addition to the mode and median) is based on distributing the data evenly among the subjects, a "fair share" concept. In mathematics, the operation of division creates a fair share distribution of n data points among k subjects. For example, if four students have a total of 28 letters in their names, then each of them would have a fair share number of 28/4 = 7 letters. In statistics, this number is called the average, or **mean**, number of letters (i.e., the number of letters each student would have were each to have the same number of letters). To this end, here is a suggestion to follow:

i. Have the groups discuss what the mode and median centers are for their names. Note that if there are no duplications, then there is no mode. Ask them to describe how the mode measures the center of a data set (most often). Ask them to describe how the median measures the center of a data set (middle of the ordered values). Ask them to determine another description of center; lead them to a fair share, even distribution, leveling off concept (mean).

ii. Have them demonstrate the fair share concept with their cubes in two ways. Suppose that one group of four students has a total of 28 letters in their first names (e.g., Sara, Landis, Octavious, Christian).

    1. For this group, combine all 28 cubes and then distribute them one by one until they are all distributed. Each of the four students will have 7 cubes, the mean for that group.

    2. An alternative procedure is to instruct each student to have a stack of cubes equal to the number of letters in his/her name. Then, move cubes from one stack to another, trying to equalize their stacks so all the stacks have the same number of cubes. Each will have 7 cubes, the mean. Note that Sara and Landis will gain cubes, while Octavious and Christian need to share some of theirs.

iii. In most cases, there will be cubes left over. Have your students discuss what to do with the leftovers so that they can be distributed evenly. For example, suppose a group of four students were Ali, Amanda, Marcas, and Aaliyah. They have a total of 22 letters. So, when distributed evenly, each will have five cubes with two left over. The two need to be split in half, with each student getting one of the halves. So, the mean would be five and a half. Another way to handle the leftovers is to see how far the two leftovers can be distributed around the group of four students. Clearly, the two would cover half of the four students (i.e., halfway around).

Note: Each group of students now knows how to find the mean for their set of four or five students. Ask them how they would find the mean number of letters in the first names for all the students in the class. Some may suggest that what they did in their individual group would have to be extended and done using the whole class. (There is a total of 148 letters, which, when divided by 25, yields a mean of 5 and 23/25 letters.)

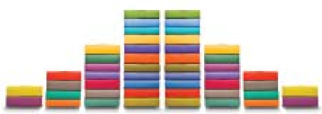
Mode, Median, and Mean Are Centers

The **mode** is whatever there is the most of. The **median** is the middle value of ordered numerical data, and the **mean** is the "fair share" value.

Others might suggest the mean of the class would be the fair share mean of the group means. This would be correct if all the groups were the same size. See the extension at the end of this investigation to determine what to do when not all of the groups are the same size.

8. Have your class determine the class mean number of letters in their first names and then ask them what the mean cost of sewing on the names for the entire class is. They should answer that the mean cost is five times the mean number of letters. They also could find the mean cost of sewing the names for the class by using the information in the dotplot, adding the cost of all the names and dividing by the number of students in the class (i.e., fair share).

Shape of a Distribution



Symmetric

Skewed

9. Examine the **shape of a distribution**.

   a. Ask your students to look at the distribution (dotplot) of the lengths of names and the distribution (dotplot) of the costs of sewing on the names.

   b. Ask your students not to focus on the numbers in the display, but rather on the way the distribution is shaped. (Trace the top of the distributions with your finger as you are showing them what to focus on.)

   c. Ask your students how the shape of the distribution for the cost of sewing on the names compares to the shape of the distribution for the number of letters in their first names. They should realize that the shapes are similar, even though the cost dotplot was shifted to the right. Ask them to discuss why the dotplots should have a similar shape. They should realize that since each letter cost 5 cents, the distribution shifted by a factor of 5. This occurred because the original data were used to determine how much a letter would cost to sew on. As a result, each data point was multiplied by 5 to determine the cost of sewing on the name. Note that this also explains why the mode, median, and mean changed by a factor of 5 cents.

Note: The students also should notice that if the dotplots are drawn on the same scale, the cost dotplot has more variability than does the names dotplot. Measures of variability are discussed in Investigation 3.4, "How Long Are Our Shoes?"

✏ Interpret the Results in the Context of the Original Question

Recall the original question, "How expensive is it to sew first names onto a T-shirt?" Also recall that it was coupled with "How do the lengths of first names vary in our class?" The investigation focused on comparing the centers and shapes of the two distributions. Ask your students to write a report that discusses how the two distributions were similar and how and why they differed. They are to provide an answer by including graphs and measures they formulated in their analysis.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

In a previous activity, we analyzed the number of letters in first names by drawing a dotplot that looked like this for our class.

Our student council voted to have T-shirts made with our first names on them and asked our class to do a statistical study on the cost. So, we made a statistical question of "How expensive is it to sew first names onto T-shirts?" We asked our domestic arts teacher what it would cost to do the sewing. She said most embroidery businesses would charge around 5 cents per letter. With that estimate, we first determined the data set for the cost of sewing first names based on our data set for the number of letters in our first names. For example, Janice has six letters in her name, so her cost would be five times six, or 30 cents. We did that for all 25 students in our class and produced the following dotplot:

We decided to compare our two dotplots. The first thing is that the scales are different. The number of letters scale is 3, 4, 5, 6, 7, 8, 9. The cost of sewing scale is 15, 20, 25, 30, 35, 40, 45. It is five times the letters scale. This makes sense since each letter costs 5 cents.

From our last study, we found the mode (most often) number of letters was 6 and the median (middle) number of letters was also 6. So it makes sense that the mode and median for the cost data set should be five times as much (i.e., 30 cents). Our teacher had us learn another measure of center called the mean. It is a fair share, or equal number, for everyone.

What we did was work in groups of four, and working with cubes that represented the number of letters in our names, put them all together in a pile and then handed them out to each other. If we had done that for our whole class according to our dotplot, all the letters in our first names would have totaled 148. Handing them out to all 25 of us gave each of us 5 cubes with 23 cubes left over. To hand out the 23 evenly, each of us would get 23/25 of a cube, so the mean (fair share) center is 5 23/25 letters for the whole data set.

After we found that the mean number of letters was 5 23/25, we found the mean cost of the letters by multiplying 5 23/25 by 5 (the cost of each letter).

The answer was a mean cost of 29 3/5 cents, which was the cost each one of us would have if we all had the same cost.

The last thing we did was to compare the shapes of the two dotplots. We saw that they were the same except the cost one was more spread out. Just like we find measures of center (mode, median, mean), our teacher said we also will learn how to measure how spread out a data set is. We can't wait to find out.

Assessment with Answers

1. Jose collected first names from his family and friends. Table 3.3.1 shows the first names and their lengths. Determine the amount of money it would cost for each family member to have their first name sewn on a T-shirt with a cost of 4 cents per letter and complete the last column in Table 3.3.1.

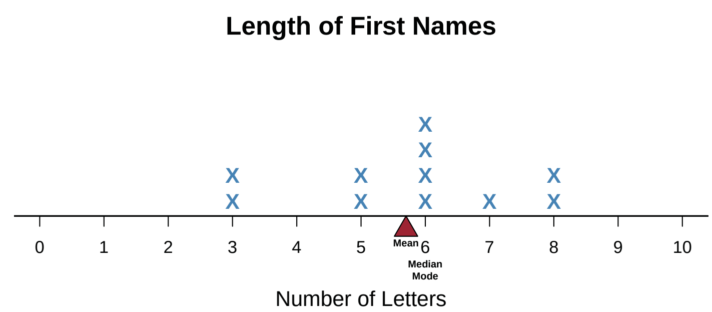
Table 3.3.1 Cost to Sew on First Name

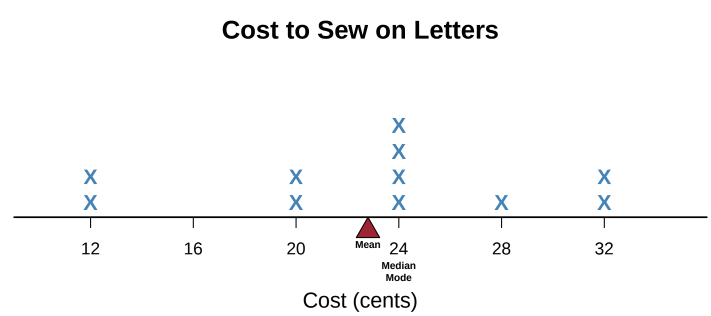
	Family and Friends First Name	Number of Letters in First Name	Cost to Sew on First Name
1	Hector	6	24 ¢
2	Amada	5	20 ¢
3	Che	3	12 ¢
4	Ricardo	6	24 ¢
5	Camila	6	24 ¢
6	Roberto	7	28 ¢
7	Carlos	6	24 ¢
8	Raymundo	8	32 ¢
9	Gabriela	8	32 ¢
10	Diego	5	20 ¢
11	Tia	3	12 ¢

2. Find the mean, median, and mode of the number of letters.

	Number of Letters	Cost of Letters
Mean	5.73	22.92 ¢
Median	6	24 ¢
Mode	6	24 ¢

3. Use words, numbers, and graphs to explain how the mean, median, and mode of the number of letters compare to the mean, median, and mode of the cost of the letters. A dotplot of each data set should be included with the location of the mean, median, and mode labeled on the graph.
The mean, median, and mode for the cost of the letters are four times the number of letters.





Extension

Introduce the concept of a weighted mean.

1. Place students in different sized groups of 3, 4, and 5.

2. Have each group find the mean length of their first names. Record the size of each group and the group mean.

3. Ask your students whether the mean of the entire class can be found by adding the group means and dividing by the number of groups. Discuss with them that this cannot be done because not all the groups are the same size. Ask them to verify that it could be done if the groups sizes were the same.

4. Discuss with your students the following method for using group means based on different group sizes to find the class mean.

   a. To find the class mean, all lengths have to be collected and distributed equally to the class members. To find the total value of the lengths, multiply the group means by the number of members in the group and add the results. The sum is the total length of first names. Finally, divide the total length by the number of students in the class.

An example:
Assume there were three groups. Group 1 had four students with a mean of 4.25 letters. Group 2 had four students with a mean of 4.5 letters. Group 3 had three students with a mean of 5 letters. To find the mean number of letters for the class, calculate the following:

$$\frac{(4 \times 4.25) + (4 \times 4.5) + (3 \times 5)}{11} = \frac{17 + 18 + 15}{11} = \frac{50}{11} = 4\frac{6}{11}$$

   a. Ask your students what 4 x 4.25, 4 x 4.5, and 3 x 5 represent.

   b. Ask your students what the 50 in the calculation represents.

   c. Ask your students what the unit is for 4 6/11.

   d. Ask your students to explain what 4 6/11 means in words. Note that they could say that each would have 4 5/11 letters if all 11 students had first names that were the same length.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics. www.corestandards.org.

Investigation 3.4
How Long Are Our Shoes?

Overview

In Investigation 1.1, "What Color Are Our Shoes," the focus was on students collecting, analyzing, and interpreting **categorical** data related to the color of their shoes. This investigation does the same for **numerical** data related to students' shoes. Students generate questions about what they would like to know about their shoes (e.g., length, number of eyelets). Decisions about how to conduct appropriate measurements and units are made. Graphs such as dotplots and boxplots are drawn to illustrate the data. Measures of center (**mode, median, mean**) and spread (**range, IQR, mean absolute deviation**) are computed. The detection of **outliers**, unusual values, will be explored. Conclusions are drawn based on the analysis in the context of the question(s) asked. This activity is based on an activity in *Exploring Statistics in the Elementary Grades: Book One, Grades K–6*, by Carolyn Bereska, L. Carey Bolster, Cyrilla A. Bolster, and Richard Scheaffer.

Note: If measuring the length of shoes or feet is not appropriate for your students, you may substitute hand span and call the investigation "How Wide Are Our Hands?"

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Make measurements in centimeters
- Calculate measures of center (mode, median, mean) and spread (range, interquartile range, mean absolute deviation) for a data set
- Display data in a dotplot and boxplot and observe clusters, gaps, and outliers

Common Core State Standards for Mathematical Practices

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.

Common Core State Standards Grade Level Content

6.SP.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

6.SP.2 Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape.

6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

6.SP.4 Display numerical data in plots on a number line, including dotplots, histograms, and boxplots.

6.SP.5 Summarize numerical data sets in relation to their context, such as by:

a. Reporting the number of observations

b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement

c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered

d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered

NCTM Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 6–8 All students should formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them; select and use appropriate statistical methods to analyze data; find, use, and interpret measures of center and spread, including mean and interquartile range; discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stemplots, boxplots, and scatterplots.

Materials

- Centimeter rulers

Estimated Time

2–3 days

Instructional Plan

Formulate a Statistical Question

Begin this investigation by holding up one of your student's shoes. Ask your students to write questions they would be interested in investigating about middle-school students' shoes. Possible questions might be the following:

a. How many eyelets are in shoes?

b. How long are shoelaces?

c. How long is a middle-school student's shoe?

d. Is the ratio of the length of a shoe to the width of a shoe a constant?

This investigation is based on the question, "How long is a middle-school student's shoe?"

Collect Appropriate Data

Before students begin measuring, discuss questions that need to be answered, including the following:

1. Should feet or shoes be measured?
2. Should the right or left shoe/foot be measured? One's right and left feet are not necessarily the same length.

3. If a shoe is chosen, should it be measured on or off the foot?

4. What unit of measurement should be used, inches or centimeters?

5. How can we be sure the measure is taken accurately? All measurements need to be taken in the same manner.

This investigation is based on measuring the shoe, in centimeters, that corresponds to the dominant foot (identified as the one mostly used when kicking something like a football). The shoe is kept on the foot, and the measurement is taken to the nearest whole number.

To minimize measurement error, tape a centimeter ruler to the floor with 0 at the wall. Students work in pairs. One student will press his or her heel against the wall as he/she steps on the ruler. The other will measure the length of the shoe by pressing another ruler vertically against the toe of the shoe, marking the distance the shoe is from the wall. The students will then switch roles. Table 3.4.1 shows a sample set of shoe length data collected from a class of sixth-graders.

Table 3.4.1 Sample Set of Shoe Length Data (cm)

20	17	17	19	20	17	19	14	17	20	15	19	
18	15	21	17	14	16	23	16	18	19	16	18	18

Analyze the Data

1. In the analysis of the data, the students should focus on the shape, center, and spread. The first step is to look at the shape of the distribution. Begin by organizing the data in a frequency table. Table 3.4.2 shows the frequency table for the sample sixth-grade class.

Table 3.4.2 Frequency Table of Sample Sixth-Grade Class Data

Length	14	15	16	17	18	19	20	21	22	23
Tally	II	II	III	~~IIII~~	IIII	IIII	III	I		I
Frequency	2	2	3	5	4	4	3	1	0	1

Note: Even though no student's shoe length is 22 cm, it must be listed in the frequency table so the list of lengths is consecutive.

2. To further analyze shape, have the students make a dotplot.

   a. After making a dotplot of the class data, discuss with the students

the shape of the distribution with any evidence of clusters, gaps, and outliers. Figure 3.4.1 is a dotplot of the sample sixth-grade class data.

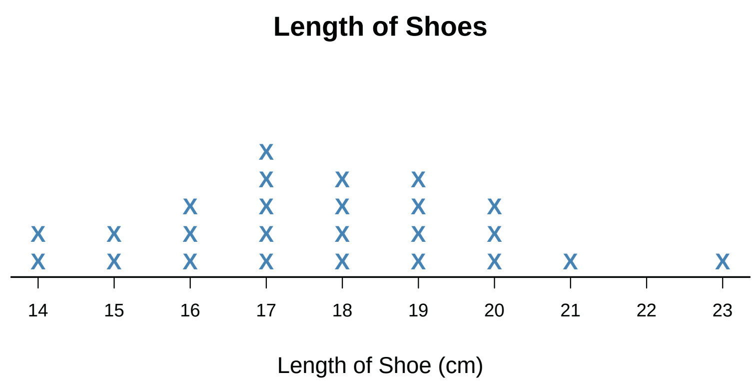


Figure 3.4.1 Dotplot of sample sixth-grade class data

Note: In this example, the shape of the distribution is fairly symmetrical and somewhat mound-shaped with a short gap between 21 and 23 cm. A possible outlier might be 23 cm.

b. Ask students to make a boxplot. Table 3.4.3 shows the five-number summary for the sample data shown in the frequency table. Figure 3.4.2 displays the boxplot of the sample sixth-grade class data. It is drawn above the dotplot to help in the analysis of the data.

Table 3.4.3 Five-Number Summary of Sample Sixth-Grade Class Data

Min	Q1	Median	Q3	Max
14	16	18	19	23

Note that the first quartile is the median of the data points strictly below the overall median of the data set, and the third quartile is the median of the data points strictly above the overall median of the data set.

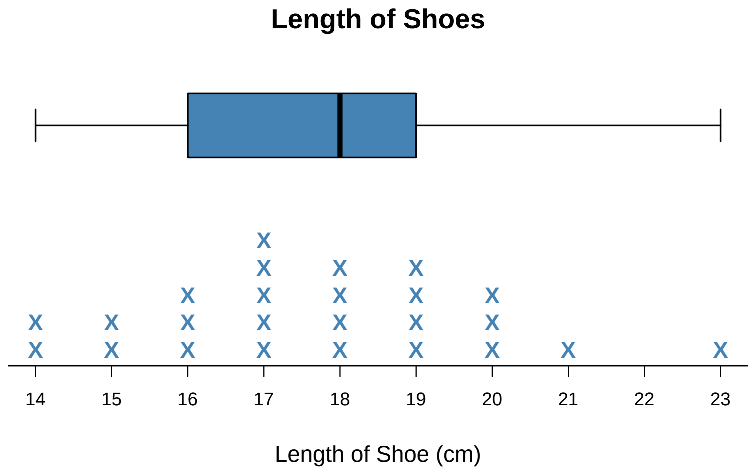


Figure 3.4.2 Boxplot and dotplot of the sample sixth-grade class data

c. Discuss with students how to interpret the boxplot. Students should understand that there are about the same **number of shoes** between the minimum and Q1, Q1 to Q2, Q2 to Q3, and Q3 to the maximum, or approximately 25% of the data in each of these four intervals. Note that Q2 is the median.

d. Demonstrate for your students how to determine if there are any outliers using the following steps:

Find the interquartile range (IQR): IQR = Q3 - Q1 = 19 - 16 = 3.

Multiply 1.5 x IQR = 1.5 x 3 = 4.5.

Add this value to Q3: 19 + 4.5 = 23.5 (called the upper fence) and subtract it from Q1: 16 - 4.5 = 11.5 (called the lower fence).

All data points that fall outside these fences are taken to be outliers.

Note that even though 23 is the highest value and follows a gap in the dotplot (hence a possible outlier), it is not an outlier according to the boxplot IQR rule.

3. The next step in the analysis of the data is to focus on the measures of center**.**

a. The question your students are investigating is how long a middle-school student's shoe is. For the given data set, one possible answer is

17 cm—the mode—because more students had a shoe length of 17 than any other shoe length.

b. Another answer would be 18 cm—the median—because 18 lies in the middle of the ordered shoe lengths. There are 12 data points to the left of one of the 18s and 12 to its right.

c. Another possibility is the shoe length determined if all the students had the same shoe length. Such a measure is called the arithmetic average, or the mean, and is interpreted as the "fair share" shoe length (i.e., if all 25 students had the same length shoe, that length would be 17.72 cm). If the 25 shoes were lined up toe to heel, the sum of the lengths would be 443 cm. If all 25 shoes were of the same length, they would each have length 443/25 = 17.72 cm. (See Investigation 3.3, "How Expensive Is Your Name," for a development of the fair share sense of mean.)

Characterizing a Spread

Range is the length of an interval.

The **interquartile range (IQR)** is the distance between the first and third quartiles.

The **mean absolute deviation (MAD)** is the arithmetic average of the absolute deviations.

4. The next step in the analysis of data is to characterize a **spread** of the shoe length distribution.

   a. The data cover the interval from the minimum of 14 to the maximum of 23. The length of the interval is the maximum - minimum, or 9 cm. This length is called the **range** of the distribution. (Note that in statistics, the range is a single number. It is tempting to say that the data *range* from 14 to 23 cm, but this would not be a correct interpretation of the statistical term range.)

   b. The boxplot suggests another measure of spread, namely the distance between the first and third quartiles, 19 - 16 = 3 cm. This measure, called the **interquartile range,** denoted by **IQR,** provides a measure of spread of the middle 50% of the shoe lengths. (Recall that the first quartile is the median of the data points strictly below the median of the data set and the third quartile is the median of the data points strictly above the median of the data set.)

   c. When a distribution is relatively symmetric, the mean is a good measure of center. Another measure of spread is constructed incorporating how far the data are from its mean, on average. The fair share absolute deviation, called the **mean absolute deviation** and denoted by **MAD**, is the arithmetic average of the absolute deviations.

   d. The first step to find the mean absolute deviation is to find how far the data are from the mean by subtracting the mean from each shoe length. Table 3.4.4 shows the 25 shoe lengths from the sample sixth-grade class and the deviations the data are from the mean.

e. Ask your students to find the sum of the deviations the data are from the mean. They should find the sum equals 0. Some of the deviations are positive if the data are greater than the mean and negative if the data are less than the mean. This indicates that the mean is the balance point of a set of data placed on a line that balances the total of the positive deviations from it with the total of the negative deviations from it.

f. The next step to finding the mean absolute deviation is to find the absolute value of the deviations. They are in the third column of Table 3.4.4.

g. The final step is to find the mean of the absolute values. **Note:** Squaring the deviations instead of taking absolute values is the basis of another measure of spread, the standard deviation, a high-school topic.

Table 3.4.4 shows the calculations for the deviations from the mean and the absolute deviations from the mean for the sixth-graders' class data set of shoe lengths. For this example, the MAD = 43.28/25 = 1.73 cm (rounded to hundredths). On average (fair share), the shoe lengths are 1.73 cm away from the mean shoe length of 17.72 cm.

Table 3.4.4 Table of Calculations to Find the Mean Absolute Deviation

Shoe Length	Length - Mean	|Length - Mean|	Shoe Length	Length - Mean	|Length - Mean|
20	20 - 17.72 = 2.28	2.28	18	18 - 17.72 = 0.28	0.28
17	17 - 17.72 = -0.72	0.72	15	15 - 17.72 = -2.72	2.72
17	17 - 17.72 = -0.72	0.72	21	21 - 17.72 = 3.28	3.28
19	19 - 17.72 = 1.28	1.28	17	17 - 17.72 = -0.72	0.72
18	18 - 17.72 = 0.28	0.28	14	14 - 17.72 = -3.72	3.72
20	20 - 17.72 = 2.28	2.28	16	16 - 17.72 = -1.72	1.72
17	17 - 17.72 = -0.72	0.72	23	23 - 17.72 = 5.28	5.28
19	19 - 17.72 = 1.28	1.28	16	16 - 17.72 = -1.72	1.72
14	14 - 17.72 = -3.72	3.72	18	18 - 17.72 = 0.28	0.28
17	17 - 17.72 = -0.72	0.72	19	19 - 17.72 = 1.28	1.28
20	20 - 17.72 = 2.28	2.28	16	16 - 17.72 = -1.72	1.72
15	15 - 17.72 = -2.72	2.72	18	18 - 17.72 = 0.28	0.28
19	19 - 17.72 = 1.28	1.28	Sum	0	43.28

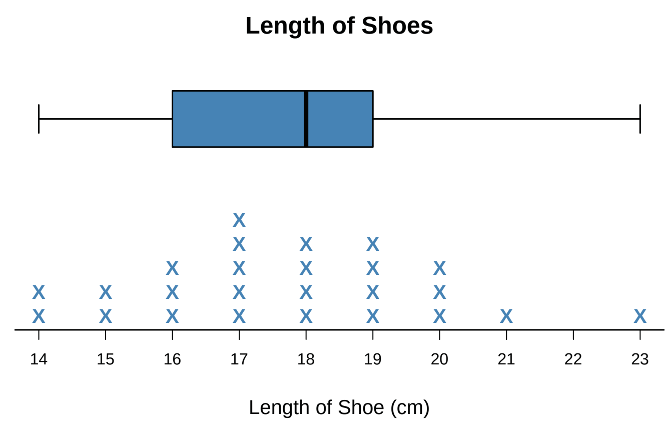
Interpret the Results in the Context of the Original Question

Ask your students to write a report that begins with an answer to the original question, "How long is a middle-school student's shoe?" They need to support their answer by focusing on what they found out about the shape, center, and spread of their data. Graphs (dotplot and boxplot) should be included.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

We wanted to know about the lengths of our shoes. From collecting our shoe lengths in a frequency table, we drew the following dotplot and boxplot.



It was neat to put the two graphs together to see what one of them showed that the other didn't. For example, the dotplot showed a possible outlier at 23 because of a gap between 21 and 23, but when we did the IQR calculation to detect outliers in a boxplot, 23 was not an outlier.

The dotplot shows the shape of the data to be fairly symmetrical. It's a little hard to see that in the boxplot because the distance from the median to Q3 is short and the distance from Q3 to the max is long. That means 25% of us, or about 6 of us, were close together between the median of 18 and Q3 = 19, but 25% of us, or about 6 of us, were spread out between Q3 = 19 and the max of 23. That's not too symmetric.

Regarding how spread out our shoe lengths are, we calculated three measures of spread. The first is the range, which is the overall distance from the minimum to the maximum, 23 - 14 = 9 cm. The second is the range of the middle 50% of the data. It is called the interquartile range. Its value is the distance from the first quartile to the third quartile, which is 19 - 16 = 3 cm. So, half our shoe lengths occupy an interval of length 3 cm. That's pretty closely packed. The third measure of spread is based on the distance each point is from the mean. This distance in statistics is called a deviation. We discovered that if you calculate all the deviations above the mean, they will be positive and the shoe lengths below the mean will be negative. It was kind of neat to see that when we added them all together, the answer was 0. We now have two meanings for the mean: It's the value everyone would have if everyone were to have the same value and a balance point of the data set put on a line. To find another measure of spread, we took the mean of the absolute values of the deviations and called it MAD—that stands for the mean absolute deviation. It was 1.73 cm for our shoe lengths, which means, on average, all our shoe lengths are 1.73 cm away from the mean shoe length of 17.72 cm.

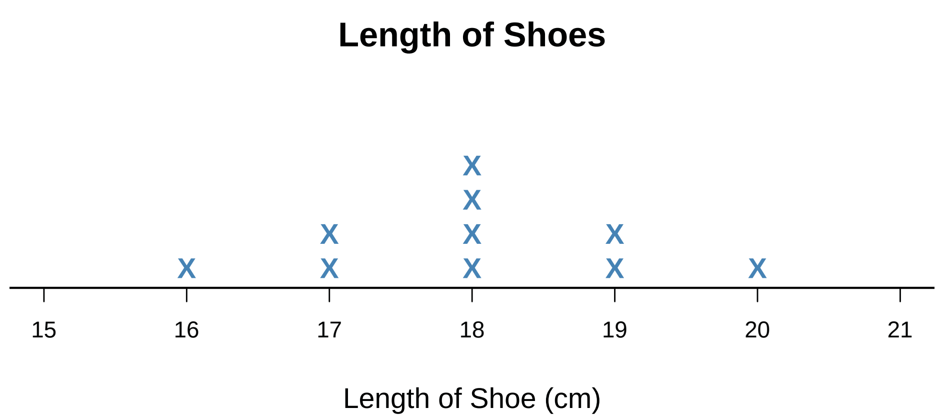
Our teacher told us that when we get to high school, we will learn another really cool measure of spread that is important and we will be able to understand it because of our working with MAD.

Assessment with Answers

Chris, a seventh-grader, collected the shoe length of a group of 10 students from the eighth grade to investigate the question regarding the shoe length of eighth graders. The data are shown in the following frequency table:

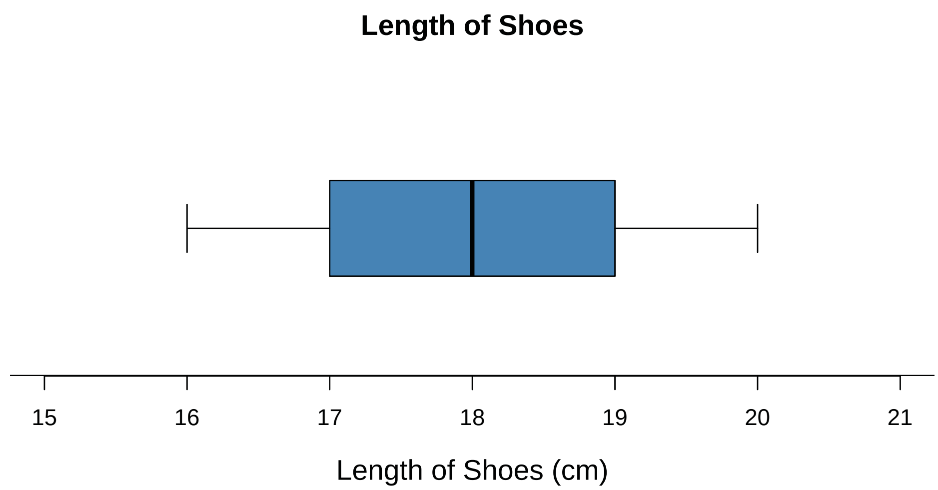
Shoe Length (cm)	16	17	18	19	20
Frequency	1	2	4	2	1

1. Draw a dotplot of the data and describe the shape of the distribution.



Distribution is mound-shaped and symmetric.

2. Draw a boxplot of the data and describe the spread of the distribution.



Five-number summary: Min = 16, Q1 = 17, Med = 18, Q3 = 19, Max = 20

3. Find the mean of the data. 18

4. Is the mean a useful measure of center? Yes, since the distribution is mound-shaped and symmetric.

5. Explain how to interpret the mean in this context. If all the students had the same length shoe, the length would be 18 cm.

6. Find the mean absolute deviation of the shoe lengths and interpret it in the context of this problem.

Shoe Length	Deviations from the Mean	Absolute Deviations
16	16 - 18 = -2	2
17	17 - 18 = -1	1
17	17 - 18 = -1	1
18	18 - 18 = 0	0
18	18 - 18 = 0	0
18	18 - 18 = 0	0
18	18 - 18 = 0	0
19	19 - 18 = 1	1
19	19 - 18 = 1	1
20	20 - 18 = 2	2

Sum of the absolute deviations = 8

Mean absolute deviation = 8/10 = .8

On average, the shoe lengths are 0.8 cm from the mean length of 18 cm.

Extensions

1. An alternative approach to developing the mean as a balance point is to have your students look at the dotplot and make a guess as to where a fulcrum should be placed to balance the graph. Suppose their guess for the balance point is 18. The first five rows of a table that shows the deviations of the data from their guess of 18 would be the following:

Table for Calculating Deviations

Shoe Length	Shoe Length - 18
20	20 - 18 = 2
17	17 - 18 = -1
17	17 - 18 = -1
19	19 - 18 = 1
18	18 - 18 = 0

Using 18 as a guess for the mean, the sum of the deviations of all 25 shoe lengths is -7. A properly placed fulcrum would be one such that the total of the negative deviations balances (cancels out) the total of the positive deviations. Since the sum was -7, the guess for the balance point was too high.

Have the students try 17. The sum of the deviations using 17 as the guess for the mean is 18, which would mean the guess of 17 was too low.

Trying 17.7 produces a total of 0.5. This would mean that 17.7 would be very close to the actual balance point. We know the mean is 17.72, so that if we now tried 17.72, the sum of all 25 deviations would turn out to be 0.

2. Add famous athletes' shoe sizes to your class data. For example, Shaquille O'Neal has shoe size 22 (about 47 cm); Michael Phelps has 14 (about 30.5 cm); and Lebron James is 16 (about 31.8 cm). How are the original conclusions affected by the inclusion of these three data points?

3. Investigate any relationship between shoe length in centimeters and shoe size. From your findings, what would a student's shoe size be if the student has a shoe length of 17 cm?

References

Bereska, C., L. C. Bolster, C. A. Bolster, and R. Scheaffer. 1998. *Exploring statistics in the elementary grades: Book one, grades k–6*. White Plains, NY: Dale Seymour.

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics. www.corestandards.org.