Section 9: One-Variable Statistics

Visit AlgebraNation.com or search "Algebra Nation" in your phone or tablet's app store to watch
the videos that go along with this workbook!

Dot Plots

Statistics is the science of collecting, organizing, and analyzing data.

Two major classifications of data:

- **Categorical:**
  - based on "qualities" such as color, taste, or texture, rather than measurements

- **Quantitative:**
  - based on measurements
The following Mathematics Florida Standards will be covered in this section:

<table>
<thead>
<tr>
<th>S-ID.1.1</th>
<th>Represent data with plots on the real number line (dot plots, histograms, and box plots).</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-ID.1.2</td>
<td>Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.</td>
</tr>
<tr>
<td>S-ID.1.3</td>
<td>Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).</td>
</tr>
</tbody>
</table>
Section 9: One-Variable Statistics

Section 9 – Topic 1

Dot Plots

Statistics is the science of collecting, organizing, and analyzing data.

There are two major classifications of data.

- **Categorical**
  - Based on "qualities" such as color, taste, or texture, rather than measurements

- **Quantitative**
  - Based on measurements

There are two types of quantitative data.

- **Discrete**
  - There is a finite number of possible data values.

- **Continuous**
  - There are too many possible data values so data needs to be measured over intervals.

Classify the following variables.

| Height                             | o Categorical  
|------------------------------------|----------------|
|                                    | o Discrete quantitative  
|                                    | o Continuous quantitative  

| Favorite subject                  | o Categorical  
|-----------------------------------|----------------|
|                                    | o Discrete quantitative  
|                                    | o Continuous quantitative  

| Number of televisions in a household | o Categorical  
|---------------------------------------|----------------|
|                                       | o Discrete quantitative  
|                                       | o Continuous quantitative  

| Area code                           | o Categorical  
|-------------------------------------|----------------|
|                                     | o Discrete quantitative  
|                                     | o Continuous quantitative  

| Distance a football is thrown       | o Categorical  
|-------------------------------------|----------------|
|                                     | o Discrete quantitative  
|                                     | o Continuous quantitative  

| Number of siblings                  | o Categorical  
|-------------------------------------|----------------|
|                                     | o Discrete quantitative  
|                                     | o Continuous quantitative  

A group of college students were surveyed about the number of books they read each month. The data set is listed below.

1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7

Let’s display the above data in a dot plot.

Each data value is represented with a _____ above the number line.
The dot plot shows the _________________ of data values.
Always include the title and an appropriate scale on the number line for the dot plot.
Dot plots are often used for:
  o smaller sets of data
  o discrete data

What is frequency?
Try It!

2. Mrs. Ferrante surveyed her class and asked each student, “How many siblings do you have?” The results are displayed below.

   0, 4, 2, 2, 3, 4, 8, 1, 0, 1, 2, 3, 0, 3, 1, 1, 2

   a. Construct a dot plot of the data.

   ![Dot plot](image)

   b. What observations can you make about the shape of the distribution?

   c. Are there any values that don’t seem to fit? Justify your answer.

BEAT THE TEST!

1. The cafeteria at Just Dance Academy offers items at seven different prices. The manager recorded the price each time an item was sold in a two-hour period and created a dot plot to display the data.

   ![Dot plot](image)

   Describe the data from the dot plot.

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!
College students were asked how well they did on their first statistics exam. Their scores are shown below.

100, 98, 77, 76, 85, 62, 73, 88, 85, 92, 93, 72, 66, 70, 90, 100

We can use a histogram to represent the data.

- **A histogram** is a bar-style data display showing frequency of data measured over ______________, rather than displaying each individual data value.
- Each interval width must be the __________.
- Always ________ the graph and ________ both axes.
- Choose the appropriate scale on the y-axis and the appropriate intervals on the x-axis.
- Histograms are often used for:
  - larger sets of data
  - continuous data

Describe an interval.

Represent the following students' scores on a histogram.

100, 98, 77, 76, 85, 62, 73, 88, 85, 92, 93, 72, 66, 70, 90, 100
Let's Practice!

1. Those same students from our first example were also asked how long in minutes it took them to complete the exam. The data is shown below.

   40.3, 42.4, 43.2, 44.1, 45.0, 55.7, 64.3, 70.3, 72.1, 32.3, 44.4, 54.5, 71.3, 66.1, 35.8, 67.2

   Construct a histogram to represent the data.

Try It!

2. Determine the sets of data where it would be better to use a histogram instead of a dot plot. Select all that apply.

   - Average daily temperatures for Albany, NY over a year
   - Daily temperatures for Albany, NY over a month
   - The results of rolling two dice over and over
   - Height of high school football players statewide
   - Finishing times of 125 randomly selected athletes for a 100-meter race
1. Last year, the local men’s basketball team had a great season. The total points scored by the team for each of the 20 games are listed below:

45, 46, 46, 52, 53, 53, 55, 56, 57, 58, 62, 62, 64, 64, 65, 67, 67, 76, 76, 89

Create a frequency table, and construct a histogram of the data.
Consider the following data set with an even number of data values.

6, 2, 1, 4, 7, 3, 8, 5

The minimum value of the data set is _____.
The maximum value of the data set is _____.
Some observations from our box plot:

- The lowest 50% of data values are from ____ to ____.
- The highest 50% of data values are from ____ to ____.
- The middle 50% (the box area) represents the values from ____ to ____.
  - The middle 50% is also known as the IQR (interquartile range).
- The first quartile represents the lower 25% of the data (____ percentile).
- The third quartile represents the first 75% of the data (____ percentile).
- 75% of the values are above ________________.
- 25% of the values are above ________________.
- The median of the lower half of the data is ______.
- The median of the upper half of the data is ______.

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set #1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Set #2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider the following data sets.

Data set #1: 1, 3, 5, 7, 9, 11, 13, 23
Data set #2: 1, 3, 5, 7, 9, 11, 13, 15

Complete the following table.

Construct the box plots for both data sets, one above the other.
Compare and contrast both box plots.

Explain which box plot is not symmetrical. Justify your answer.

**Let's Practice!**

1. Consider the following data set with an odd number of data values.

   3, 7, 10, 11, 15, 18, 21

   a. The minimum value of the data set is _____.
   b. The maximum value of the data set is _____.
   c. The median of the data set is _____.
   d. The first quartile of the data set is _____.
   e. The third quartile of the data set is _____.
   f. Use the five-number summary to construct a box plot.

---

**Try It!**

2. The time, rounded to the nearest hour, that 26 tourists spent on excursions in Cat Island, Mississippi on a given day was recorded as follows. (Cat Island is not actually an island for cats.)

   0, 3, 4, 4, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 10, 10, 10, 11, 11, 12

   a. Construct a box plot to represent the data. Label the minimum, maximum, first quartile, third quartile, and median.
   b. The bottom 25% of tourists spent, at most, ____ hours on excursions.
1. Mrs. Bridgewater recorded the number of Snapchats 10 different students sent in one day and constructed the box plot below for the data.

**Part A:** Use the following vocabulary to label the box plot.  
Hint: You will not use all of the words on the list.

<table>
<thead>
<tr>
<th>A. Average</th>
<th>E. Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. First Quartile</td>
<td>F. Minimum</td>
</tr>
<tr>
<td>C. Maximum</td>
<td>G. Third Quartile</td>
</tr>
<tr>
<td>D. Mean</td>
<td></td>
</tr>
</tbody>
</table>

Snapchats Sent In One Day

![Box Plot Diagram](image)

**Part B:** The 50<sup>th</sup> percentile of the data set is _____.

**Part C:** Half of the data values are between 8 and 12.

**Part D:** 75% of students send 12 or fewer Snapchats per day.

**Part E:** Add dots to the number line below to complete the dot plot so that it could also represent the data.

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!
Section 9 – Topic 5
Measures of Center and Shapes of Distributions

Data displays can be used to describe the following elements of a data set’s distribution:

- Center
- Shape
- Spread

There are three common measures of center.

- **Mean:** The ___________ of the data values.
- **Median:** The ___________ value of the ordered data set.
- **Mode:** The ________ _______________ occurring value(s).

Mr. Gray gave a test on a regular school day with no special activities. The scores are listed below.

60, 60, 70, 70, 70, 80, 80, 80, 80, 90, 90, 90, 100, 100

The dot plot for the data is as follows:

Looking at the dot plot, what do you think is the value of the median?

What is the value of the mean?

Why is it important to know where the center is?

The shape of a dot plot also gives important information about a data set’s distribution. The data in the previous dot plot is symmetrical and follows a **normal distribution**. What do you notice about the shape of a normal distribution?
Let’s Practice!

1. Mr. Gray then gave a test the day after a basketball game against the school’s rival. The scores were as follows.

   65, 65, 65, 65, 65, 70, 70, 70, 70, 75, 75, 75, 75, 75, 80, 80, 80, 80, 85, 90, 90, 95, 100

   a. What are the mean and the median of this data set?

   b. Which measure is a more appropriate measure of center, the mean or the median?

   c. Does this data set have a normal distribution? Why or why not?

   d. The shape of this distribution is ___________ _________.

Try It!

2. Mr. Gray then gave a test the day after a mid-week early release day. The scores were as follows.

   50, 60, 70, 70, 80, 80, 80, 90, 90, 90, 90, 90, 100, 100, 100

   a. Which value do you think will be smaller: the mean or the median?

   b. Consider the dot plot for the data.

   Which measure is a more appropriate measure of center, the mean or the median?

   c. The shape of this distribution is ___________ _________.

   d. For a normal-shaped data set the best measure of center is the ____________, whereas for a skewed-shaped data set, the ____________ is better.
1. Mr. Logan surveyed his junior and senior students about the time they spent studying math in one day. He then tabulated the results and created a dot plot displaying the data for both groups.

**Part A:** The value of the larger median for the two groups is _____.

**Part B:** The value of the larger mean for the two groups is _____.

**Part C:** Using one to two sentences, describe the difference between the number of minutes the juniors and seniors studied by comparing the center and shapes for the groups.

---

A meteorologist recorded the average weekly temperatures over a 13-week period and displayed the data below.

A meteorologist in a different state also recorded the average weekly temperatures over a 13-week period and displayed the data below.

Measures of spread tell us how much a data sample is spread out or scattered.

What are the differences between the spreads of the two data sets?

---

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!
There are two primary ways to measure the spread of data.

- **Interquartile Range (IQR)** represents the middle 50% of the data and is typically used to describe the spread of **skewed** data.

Consider the following data set.

5, 5, 6, 7, 8, 8, 8, 9, 10, 12, 12

What are the first and third quartiles of the data?

Calculate the interquartile range (IQR) of the data.

Why do you think IQR is used to measure spread in skewed data?

**Standard deviation** is the typical distance of the data values from the mean. The larger the standard deviation, the **farther** the individual values are from the mean. It is typically used for **normal** data.

Consider the dot plots below.

A.

B.

Which has a larger standard deviation? Explain your answer.

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!
Let's Practice!

1. The Bozeman Bucks and Tate Aggies cross-country teams ran an obstacle course. The times for each team are summarized below.

   **Bozeman Bucks’ Obstacle Course Times**
   
   4:25  4:43  4:49  5:02  5:12
   5:21  5:31  5:32  5:37  5:52
   5:54  6:08  6:20  6:26  6:33
   6:48  6:53  7:16  7:23  8:05

   **Tate Aggies’ Obstacle Course Times**

   Which statements are true about the data for the Bozeman Bucks and the Tate Aggies? Select all that apply.

   - The median time of the Bozeman Bucks is less than the median time of the Tate Aggies.
   - The fastest 25% of athletes on both teams complete the obstacle course in about the same amount of time.
   - The interquartile range of the Bozeman Bucks is less than the interquartile range of the Tate Aggies.
   - Approximately 50% of Tate Aggies have times between 5 and 6 minutes.
   - The data for the Bozeman Bucks is skewed to the left.

Try It!

2. The following box plots represent the starting salaries (in thousands of dollars) of 12 recent business graduates, 12 recent engineering graduates, and 12 recent psychology graduates.

   **Salaries of Recent Graduates**

   **Business**
   **Engineering**
   **Psychology**

   a. Describe the shape of each major’s data distribution.

   - Business:
   - Engineering:
   - Psychology:

   b. Which major has the largest median salary? The largest IQR?
1. Data on the time that Mrs. Lannister’s students spend studying math and science on a given night are summarized below.

<table>
<thead>
<tr>
<th>Math</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean: 75 minutes</td>
<td>Mean: 25 minutes</td>
</tr>
<tr>
<td>Minimum: 0 minutes</td>
<td>Minimum: 0 minutes</td>
</tr>
<tr>
<td>First Quartile: 65 minutes</td>
<td>First Quartile: 15 minutes</td>
</tr>
<tr>
<td>Median: 78 minutes</td>
<td>Median: 30 minutes</td>
</tr>
<tr>
<td>Third Quartile: 100 minutes</td>
<td>Third Quartile: 35 minutes</td>
</tr>
<tr>
<td>Maximum: 145 minutes</td>
<td>Maximum: 50 minutes</td>
</tr>
<tr>
<td>Standard deviation: 8 minutes</td>
<td>Standard deviation: 12 minutes</td>
</tr>
</tbody>
</table>

Tyrion spent 10 minutes studying math and 50 minutes studying science. If Tyrion spent all 60 minutes studying math, which of the following would be affected?

<table>
<thead>
<tr>
<th>Increases</th>
<th>Decreases</th>
<th>Stays the Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

**Interquartile Range of Math Time**

2. The data from a survey of the ages of people in a CrossFit class were skewed to the right.

**Part A:** The appropriate measure of center to describe the data distribution is the

- mean.
- median.

The

- interquartile range.
- standard deviation.

is the appropriate measure to describe the spread.

**Part B:** The box plot below represents the data. Calculate the appropriate measure of spread.

**Age Distribution in a CrossFit Class**

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!
Assume that we have a data set so large that we are not given a list of all the values. We are told the data follows a normal distribution with a mean of 16 and standard deviation of 4.

Label the distribution below with the values using the mean and standard deviation.

Suppose one of the data values is 20. An observation of 20 is _____ standard deviation(s) _________ the mean.

Suppose one of the data values is 8. An observation of 8 is _____ standard deviation(s) ______________ the mean.

Suppose an observation is 1.5 standard deviations above the mean. The value of that observation is ____________.

We can use the empirical rule to understand the data distribution.

**Empirical Rule**

- Approximately 68% of values are within one standard deviation of the mean.
- Approximately 95% of values are within two standard deviations of the mean.
- Approximately 99.7% of values are within three standard deviations of the mean.

Label the percentages on the previous distribution.
**Let's Practice!**

1. Suppose the amounts of water a machine dispenses into plastic bottles has a normal distribution with a mean of 16.2 ounces and a standard deviation of 0.1 ounces.

   a. Label the distribution below with the values using the mean and standard deviation.

   ![Normal Distribution Graph](image)

   b. The middle 95% of bottles contain between _____ and _____ ounces of water.

   c. Approximately 68% of bottles have between _____ and _____ ounces of water.

   d. What percentage of bottles contain more than 16.4 ounces of water?

   e. What is the probability that a randomly selected bottle contains less than 16.3 ounces of water?

   f. What percentage of bottles contain between 16.1 and 16.4 ounces of water?

---

**Try It!**

2. Choose the correct numbers to build a normal distribution graph based on a mean of 45.5 and standard deviation of 3.92 (All numbers will not be used, and some may be used more than once).

<table>
<thead>
<tr>
<th>13.5%</th>
<th>13.5%</th>
<th>37.66</th>
<th>34%</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.26</td>
<td>33.74</td>
<td>53.34</td>
<td>49.42</td>
</tr>
<tr>
<td>2.45%</td>
<td>2.45%</td>
<td>68%</td>
<td>45.5</td>
</tr>
<tr>
<td>34%</td>
<td>41.58</td>
<td>95%</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

   ![Normal Distribution Graph](image)
1. SAT mathematics scores for a particular year are approximately normally distributed with a mean of 510 and a standard deviation of 80.

   Part A: What is the probability that a randomly selected score is greater than 590?

   Part B: What is the probability that a randomly selected score is greater than 670?

   Part C: What percentage of students scored between 350 and 670?

   Part D: A student who scored a 750 is in the ___________ percentile.

Want some help? You can always ask questions on the Algebra Wall and receive help from other students, teachers, and Study Experts. You can also help others on the Algebra Wall and earn Karma Points for doing so. Go to AlgebraNation.com to learn more and get started!

Section 9 – Topic 9
Outliers in Data Sets

A survey about the average number of text messages sent per day was conducted at a retirement home.

5, 5, 5, 5, 5, 10, 10, 10, 10, 10, 15, 15, 15

The mean for this data set is 8.7 and the median is 10.

Grandma Gadget is up-to-date on the latest technology and loves to text her 25 grandchildren. She sends an average of 85 texts per day. Her data point is substituted for one of the original data points of 15.

The new data set is:

5, 5, 5, 5, 5, 10, 10, 10, 10, 10, 15, 15, 85

Which measure of center will be most affected by substituting Grandma Gadget – the mean or the median? Justify your answer.

Does Grandma Gadget’s data point have a greater effect on standard deviation or interquartile range? Justify your answer.
Grandma Gadget’s data point is called an **outlier**.

An **outlier** is an ______________ value in a data set that is very distant from the others.

**Let’s Practice!**

1. The table below lists the number of customers who visited a car dealership during 30 randomly selected days.

<table>
<thead>
<tr>
<th>26</th>
<th>29</th>
<th>27</th>
<th>33</th>
<th>29</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>36</td>
<td>26</td>
<td>31</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>28</td>
<td>11</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>33</td>
<td>37</td>
<td>31</td>
<td>26</td>
<td>37</td>
<td>33</td>
</tr>
<tr>
<td>29</td>
<td>35</td>
<td>37</td>
<td>29</td>
<td>27</td>
<td>33</td>
</tr>
</tbody>
</table>

Identify the outlier and describe how it affects the mean and the standard deviation.

The outlier is ______. The outlier in the data set causes the mean to ______________ and the standard deviation to ______________.

**Try It!**

2. The students in Mrs. Gomez’s class were surveyed about the number of text messages they send per day. The data set is as follows.

   0, 24, 26, 28, 30, 33, 35, 35, 36, 38, 39, 42, 42, 45, 50

   a. What value would you predict to be an outlier?

   b. How does the outlier affect the mean?

   c. How does the outlier affect the median?

   d. Which measure of center would best describe the data, the mean or the median?

   e. How does the outlier affect the standard deviation?

   f. How does the outlier affect the interquartile range?

   g. Which measure of spread would best describe the data, the standard deviation or the interquartile range?
1. The dot plot below compares the arrival times of 30 flights for two different airlines.

Arrival Times for Airlines P and Q

Airline P

Airline Q

A negative number represents the number of minutes the flight arrived before its scheduled time.

A positive number represents the number of minutes the flight arrived after its scheduled time.

A zero indicates that the flight arrived at its scheduled time.

Based on these data, from which airline would you choose to buy your ticket? Use your knowledge of shape, center, outliers, and spread to justify your choice.

2. After a long day at Disney World, a group of students were asked how many times they each rode Space Mountain. The values are as follows.

4, 3, 19, 1, 2, 4, 3, 5, 3, 4, 5, 4, 5

Part A: Are there any outliers in the data set above? Explain.

Part B: The outlier causes the to be greater than the

Part C: If the outlier were changed to 5, the interquartile range would and the standard deviation would

Test Yourself!

Practice Tool

Great job! You have reached the end of this section. Now it’s time to try the “Test Yourself! Practice Tool,” where you can practice all the skills and concepts you learned in this section. Log in to Algebra Nation and try out the “Test Yourself! Practice Tool” so you can see how well you know these topics!

Section 9: One-Variable Statistics