Algebra 1

**Univariate Datasets**

**Unit Summary:** In this unit, you will learn about univariate data and the statistical computation, representation, and analysis that can be done using this type of data.

**Lesson 2 – Describing Distributions**

**Key Words:**

* **box plot** – a visual representation that shows the five-number summary of a dataset
* **dot plot** – a visual representation of data that consists of individual data points plotted as dots on a graph
* **histogram** – a representation of a frequency distribution by means of rectangles whose widths represent class intervals and whose areas are proportional to the corresponding frequencies
* **interquartile range** – the range of values of the variable in a statistical distribution that lies between the upper and lower quartiles
* **lower quartile** – the middle value of the lower half of a dataset
* **measures of central tendency** – the extent to which statistical values fall around a middle value
* **measures of spread** – the extent to which statistical values vary
* **median** – the middle value in an ordered set of data, or the mean of the two middle numbers
* **range** – the difference between the least and greatest of a set of values
* **upper quartile** – the middle value of the upper half of a dataset

**Objective 1:** In this section, you willrepresent data on a dot plot.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* *Dot plots* can be used to find trends in the data because it is clearly noticeable.
* *Dot plots* are better used to represent **small sets of data**.
* You can interpret data from a *dot plot* by looking at the shape, center, and spread of the data.
  + Shape: By examining the shape of a dataset, trends can be found and used as information for making decisions or making predictions.
  + Center: Mean, median, and mode can be determined from a dot plot because they show each individual data point.
  + Spread: You can calculate the range, or difference between maximum and minimum values, to analyze the spread of the data.

|  |  |
| --- | --- |
| **Example:** The following dot plot represents the number of gallons of paint ordered per order at a hardware store. Use the dot plot to answer the following questions: | |
| 1. **Are smaller orders more common or larger orders?** | More than half of the data points are greater than or equal to 4 gallons, and so larger orders are more common. |
| 1. **How many gallons of paint is the most commonly ordered? Least commonly ordered?** | The most common order size is for 6 gallons of paint. The least commonly ordered sizes are for 1 gallon and 7 gallons. |
| 1. **How many orders for gallons of paint did the hardware store get? What were the largest and the smallest order sizes?** | There were a total of 22 orders with the largest order for 7 gallons of paint and the smallest order for 1 gallon. |

**Objective 2:** In this section, you will represent data in a histogram.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* When analyzing a **large dataset**, *histogram* models allow for data points to be grouped together.
* *Histograms* use intervals to represent data in common groups.
* Review the analysis process using the following example:

|  |  |  |
| --- | --- | --- |
| **Data Collection** | **Grouping** | **Histogram** |
| The number of tacos eaten per person is provided in the list for a group of 25 people who attended a party.   |  |  |  | | --- | --- | --- | | Tacos Per Person | | | | 21 | 15 | 21 | | 18 | 19 | 17 | | 19 | 20 | 18 | | 16 | 16 | 21 | | 22 | 19 | 16 | | 19 | 15 | 20 | | 16 | 14 | 23 | | 17 | 17 | 20 | | 23 |  |  | | Create a frequency table with groupings, or intervals, and the number of data points that fall in that interval.   |  |  | | --- | --- | | Tacos Eaten | Frequency | | 14-15 | 3 | | 16-17 | 7 | | 18-19 | 6 | | 20-21 | 6 | | 22-23 | 3 | | Use the data from the frequency table to create a histogram. |

* *Histograms* can also be used to describe aspects of the data.

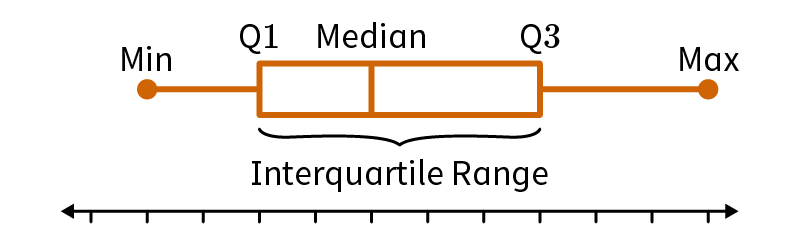
|  |  |
| --- | --- |
| **Example:** Use the provided histogram to answer the following questions. | |
| 1. **How many individuals were surveyed? In other words, how many data points are in the data set?** | By adding up the frequency, or amount of people in each grouping, you can determine the total number of people surveyed.  Fifty people were surveyed. |
| 1. **How many surveyed individuals send between five and six text messages a day?** | Locate the grouping for 5-6 text messages and determine the frequency.  Zero people send between five and six text messages per day. |

**Objective 3:** In this section, you will represent data in box plots.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* A *box plot* is a visual representation of a data set that provides a quick way to compare data.
* A *box plot* does not show a full dataset but rather provides a set of statistics called the *five-number summary*.
  + ***minimum***: smallest value in the dataset
  + ***lower quartile (Q1)***: middle value between the minimum and the median
  + ***median***: middle value in the dataset, or the mean of the two middle numbers and is often denoted as Q2.
  + ***upper quartile (Q3)***: middle value between the median and the maximum
  + ***maximum***: largest value in the dataset



|  |  |
| --- | --- |
| **Example:** The data values represent the weight in grams of 15 different walnuts that were collected from a park. Find the five-number summary and construct a box plot to represent the data. Interpret the meaning of Q1.  24, 25, 24, 27, 29, 29, 30, 32, 31, 34, 34, 35, 34, 36, 37 | |
| **Step 1:** Reorder the data values from smallest to largest. | 24, 24, 25, 27, 29, 29, 30, 31, 32, 34, 34, 34, 35, 36, 37 |
| **Step 2:** Identify each part of the five number summary | * The minimum is the smallest data value, which is 24. * The median is the middle most data value. Since there is an odd number of data values, the median is the middle most data value.   + 24, 24, 25, 27, 29, 29, 30, 31, 32, 34, 34, 34, 35, 36, 37 * The lower quartile is the middle value between the minimum and the median. Therefore, Q1=27   + **24, 24, 25, 27, 29, 29, 30**, 31, 32, 34, 34, 34, 35, 36, 37 * The upper quartile is the middle value between the median and the maximum. Therefore Q3=34.   + 24, 24, 25, 27, 29, 29, 30, 31, **32, 34, 34, 34, 35, 36, 37** * The maximum is the largest data value, which is 37.   The five number summary in order is 24, 27, 31, 34, 37. |
| **Step 3:** Construct a box plot. | * Begin by drawing a number line with units that will fit both the minimum and maximum values of the dataset. * Next, mark the values from the five-number summary on the number line and draw a box from the lower quartile (Q1) to the upper quartile (Q3) with a line through the median (Q2). * Then, draw a line from Q1 to the minimum and again from Q3 to the maximum. |
| **Step 4**: Interpret the meaning of Q1. | Since Q1=27, about 25% of the data is below 27 and about 75% of the data is above 27. This means that 25% of the walnuts are less than 27 grams, and about 75% of the walnuts are more than 27 grams. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the image to answer the question.  A dot plot labeled Hours Playing Video Games.  The dot plot represents the number of hours spent playing video games. How many hours is the most common amount spent playing video games?  Most people played video games for \_\_\_\_ hours. | 6 |
| P 2 | Use the image to answer the question.  An unlabeled dot plot.  The dot plot represents a sample of data. If each dot represents a person who participated, how many people were involved?  \_\_\_\_ people participated in the sample. | 15 |
| P 3 | Use the table to answer the question.    The frequency table shows the data that could be represented by a histogram. Based on the table, describe the shape of the data. Enter the option number that corresponds to the correct description of the dataset.  Option #1: Most of the data is on the left side of the graph.  Option #2: Most of the data is on the right side of the graph.  Option #3: Most of the data is in the center of the graph.  The shape of the data can be described as Option #\_\_\_\_. | 2 |
| P 4 | Use the table to answer the question.    The frequency table shows the data that could be represented by a histogram. Based on the table, which interval will have the highest bar?  The interval that will have the highest bar is \_\_\_\_\_. | 21-25 |
| P 5 | What is the five-number summary for this set of test scores?  90, 95, 89, 84, 74, 99, 97, 86, 93 | 1. 64 2. 85 3. 90 4. 96 5. 99 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the image to answer the question.  An unlabeled dot plot.  The dot plot shows the number of hours spent studying for an assessment. Based on the data distribution, which of the following statements is correct? | More students studied three hours or less than the number of students who studied four hours or more. |
| Q 2 | Use the image to answer the question.  An unlabeled dot plot.  What is the minimum value selected? | 1 |
| Q 3 | Use the image to answer the question.  A histogram's x-axis is labeled 'Text Messages Sent per Day' and ranges from 1 to 14 in intervals of 2. The y-axis is labeled 'Frequency' and ranges from 0 to 14 by 4-unit increments. 6 bars are plotted for 6 of the intervals on the x-axis.  What statement informally describes the shape of the histogram? | Most of the data is on the right side of the graph. |
| Q 4 | Use the table to answer the question.    The table of data is used to create a histogram. Which interval will have the highest bar on the histogram? | 15-19 |
| Q 5 | The dataset 3, 6, 2, 8, 6, 6, 9, 10, 3 represents the scores on the most recent quiz. Describe the change in the distribution of data if an additional number is added. If you add the score of 4, what part of the five-number summary will change? | Q3 |

**Lesson 3 – Centers of Distributions**

**Key Words:**

* **center** – a point or part that is equally distant from all sides, ends, or surfaces of something
* **distribution** – the position, arrangement, or frequency of occurrence over an area or throughout a space or unit of time
* **dot plot** – a visual representation of data that consists of individual data points plotted as dots on a graph
* **left-skewed** – a graph with a tail that extends to the left
* **mean** – a measure of the center of a dataset found by adding all items in a set and dividing by the total number of items
* **median** – the middle value in an ordered set of data, or the mean of the two middle numbers
* **outlier** – a statistical observation that is markedly different in value from the others of the sample
* **right-skewed** – a graph with a tail that extends to the right
* **skewed** – a graph or dataset that is not symmetrical, containing more data on one side than the other

**Objective 1:** In this section, you will compare the means and medians of symmetrical data by using dot plots.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* A symmetrical dataset occurs when the ***mean* and *median* are the same**.
* If a *dot plot* is symmetrical, one side of the graph will be a reflection of the other side.
* Recall that *mean* and *median* are two different measures of *center.* 
  + The *mean,* or average*,* is found by adding all items in a set and dividing by the total number of items.
  + The *median,* or middleis found by organizing the data points in numerical order and finding the middle number.

|  |  |
| --- | --- |
| **Example:** Use the following dot plot to compare median and mean. | |
| **Step 1:** Find the median. | List the data points in order from smallest to largest. The median will be the center of the data set.  5, 6, 6, 7, 7, 7, 8, **8**, 8, 9, 9, 9, 10, 10, 11  The median is 8. |
| **Step 2:** Find the mean. | Add the data points together and divide by the total number of data points. In all, there are 15 data points.  5 + 6 + 6 + 7 + 7 + 7 + 8 + 8 + 8 + 9 + 9 + 9 + 10 + 10 + 11 = 120  The mean is 8. |
| **Step 3:** State the answer. | The mean and median are the same, which is 8. Therefore the data is symmetric. |

**Objective 2:** In this section, you will compare the means and medians of skewed data represented in a dot plot.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* *Skewed* data is when there are more data points on one side of the graph than the other.
* In a *dot plot*, this will appear as more dots on one end of the graph and a tail of dots on the other end.
* In *skewed* data, the ***mean* and *median* will be different**.

|  |  |
| --- | --- |
| **Left-Skewed Data** | **Right-Skewed Data** |
| The data points are on the higher end of the range, on the right of the graph, and a tail extends to the **left**.  For example: | The data points are on the lower end of the range, on the left of the graph, and a tail extends to the **right**.  For example: |

|  |  |
| --- | --- |
| **Example:** Using the following dot plot, does the dataset appear skewed? Compare median and mean of the dataset. | |
| **Step 1:** Use the shape of the graph to determine if the data is skewed. | The data appears to be right-skewed because most of the data is in the lower range of the graph and there is a tail to the right. |
| **Step 2:** Calculate the mean and median. | The mean is 1.63.   * 0 + 0 + 0 + 0 + 0 + 1 + 1 + 1 + 2 + 2 + 2 + 2 + 3 + 3 + 4 + 5 = 26   The median is 1.5.   * 0, 0, 0, 0, 0, 1, 1, **1**, **2**, 2, 2, 2, 3, 3, 4, 5 * There is an even amount of numbers, therefore there are two values in the middle. Take the average of the two values to get the median. |
| **Step 3:** State the answer. | The mean is higher than the median. The data is skewed right because the few higher points, on the tail, are pulling the mean to the right |

**Objective 3:** In this section, you will explain how the shape of distribution determines the most appropriate measure of center.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* All data distributions, symmetrical and skewed, have *measures of center*. It’s important to use the most appropriate *measure of center* so that the data can be used accurately.
* You can determine whether the *mean* is appropriate based on the existence of *outliers*.
  + *Outliers* can significantly affect the value of the mean by raising or lowering it.
  + Recall that an *outlier* is a number that is significantly higher or lower than the majority of the dataset.
* When working with ***symmetrical* data**, the best measure of center is the ***mean***.
* When working with ***skewed* data**, the best measure of center is the ***median***.

|  |  |
| --- | --- |
| **Example** | **Measure of Center** |
|  | The data is right-skewed. The median is the most appropriate because the mean would be significantly affected by the presence of an outlier. |
|  | The data is symmetrical. The mean would be the best measure of center because it is not affected by an outlier. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the image to answer the question.  A dot plot has a horizontal axis ranging from negative 1 to 3 by 1-unit increments. The plot has 1 dot above negative 1, 2 dots above 0, 4 dots above 1, 3 dots above 2, and 1 dot above 3.  Compare the mean and the median. Is the data symmetrical?  Type 1 for yes.  Type 2 for no. | 2 |
| P 2 | Use the image to answer the question.  A dot plot has a horizontal axis ranging from 80 to 85 by 1-unit increments. The plot has 2 dots above 80, 2 dots above 81, 0 dots above 82, 1 dot above 83, 4 dots above 84, and 3 dots above 85.  Compare the mean and median of the following data. Is the data symmetrical?  Type 1 for yes.  Type 2 for no. | 2 |
| P 3 | Use the image to answer the question.  A number line with arrows at both the ends ranges from 0 to 5 in unit increments. It is labeled as Number of Followers per Day. Closed circles are plotted above the number line.  Sam has started a social media video channel. Sam is so excited about his new channel that he has been keeping track of his followers. He has created a dot plot for the first week of followers. Which statement about the dot plot is true?  Option #1: The data on the dot plot is skewed to the right.  Option #2: The data on the dot plot is skewed to the left.  Option #\_\_\_ is true. | 2 |
| P 4 | Use the image to answer the question.  A number line with arrows at both the ends ranges from 4 to 4 point 5 in increments of 0 point 25. It is labeled as Student’s height in Feet. Closed circles are plotted above the number line.  A group of students created a dot plot of their heights. Based on the dot plot, find the mean and median height of the students in this class to show that the two measures are different in this skewed data set. Round your answer to the nearest tenth.  The mean height of the students is \_\_\_\_ feet, and the median height of the students is \_\_\_ feet. | 4.6; 4.5 |
| P 5 | Use the image to answer the question.  Two dot plots are titled First Period and Second Period respectively. Both number lines with arrows at both ends range from 0 to 10 in unit increments and are labeled as Number of Books. Vertical columns of closed dots are plotted above certain numbers.  Which data distribution is the most symmetrical and suggests using the mean as the most appropriate measure of center?  Option #1: First Period  Option #2: Second Period  Option #\_\_\_\_ is the correct response. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the image to answer the question.  A dot plot is labeled 'Number of Siblings.' The horizontal axis ranges from 0 to 5 by 1-unit increments. The plot has 2 dots above 0, 7 dots above 1, 4 dots above 2, 4 dots above 3, 2 dots above 4, and 1 dot above 5.  Compare the values and calculate the median of the given data. | 2 |
| Q 2 | Use the image to answer the question.  A dot plot is labeled 'Alarm Clock Set.' The horizontal axis ranges from 5 AM to 7 in increments of 30 minutes. The plot show 2 dots above 5 AM, 3 dots above 5:30, 5 dots above 6:00, 4 dots above 6:30, and 1 dot above 7.  Calculate the mean and median of the given data. Round your answers to the nearest half-hour increment as needed. | Mean: 6:00, Median: 6:00 |
| Q 3 | When comparing the means and medians of skewed data, which of the following is correct? | The mean and median are never the same. |
| Q 4 | Use the image to answer the question.  A distribution curve is plotted by joining the four points A, B, C, and D on a plane.  A team of students were comparing the mean and median of a dot plot and noticed that the dots made a graph. They labeled the mean, median, minimum, and maximum of the graph. Which point is most likely the mean of the skewed data? | C |
| Q 5 | Explain how the shape of a distribution determines the appropriate measure of center. | Symmetrical data means that the mean is the best measure of center. |

**Lesson 4 – Estimations of Centers**

**Key Words:**

* **balance point** – the number in which the distance from the highest and the lowest point are equivalent
* **descriptor** – something (such as a word or characteristic feature) that serves to describe or identify
* **estimate** – to judge tentatively or approximately the value, worth, or significance
* **histogram** – a representation of a frequency distribution by means of rectangles whose widths represent class intervals and whose areas are proportional to the corresponding frequencies
* **mean** – a measure of the center of a dataset found by adding all items in a set and dividing by the total number of items
* **median** – the middle value in an ordered set of data, or the mean of the two middle numbers
* **outlier** – a statistical observation that is markedly different in value from the others of the sample

**Objective 1:** In this section, you will estimate the mean and median of data distributions.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* If the distribution is left leaning, in other words *right-skewed*, the *mean* will fall to the right of the *median* because the outliers, or tail, are pulling the *mean* to that side of the graph.
* Similarly, if the distribution is right leaning, or *left-skewed*, the *mean* will fall to the left of the *median*.
* This information is helpful when estimating the mean of a data distribution.
  + When estimating the *mean*, you must first find the *median* of the dataset.

|  |  |
| --- | --- |
| **Example** | **Estimate the Mean** |
|  | * The data is right leaning, or left-skewed. Therefore, the mean will fall to the left of the median. * The median is 4. * A reasonable estimate for the mean is 3. |
|  | * The data is left leaning, or right skewed. Therefore, the mean will fall to the right of the median. * Since this is a histogram, we can only estimate the group that the median will fall in. * The median will be in the third grouping labeled 72-72.9. * A reasonable estimate for the mean is about 73. |
|  | * The data is symmetrical, therefore the mean and median will be the same. * The median is 18, therefore so is the mean. |

**Objective 2:** In this section, you will explain mean as a balance point when both mean and median are not good descriptors.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* The *mean* as a balance point means the data are **evenly distributed** around the *mean*.
* The distances from the *mean* to the high data points and low data points are the same.
* When the data has a U-shape, you will treat the *mean as a balancing point*.
  + This is different than using the *mean as the center,* which is used for symmetrical data that is not U-shaped.
* A *balancing point* refers to data that are the same distance from each edge of the set to the center (*mean*).

|  |  |
| --- | --- |
| **Example** | **Descriptor** |
|  | The graph is symmetrical but with a U-shape. All points in this set are balanced around 4.  The data is best represented with the **mean as the balancing point** of data. It is the best descriptor because the data has a U-shaped. |
|  | The data is symmetrical and does not have a U-shape.  The best descriptor is the **mean as the center** because the representation is a set of symmetrical data meaning it isn’t affected by any outliers. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the image to answer the question.  A bar graph titled Frequency versus Age of Students in Algebra 1 has 4 bars. The horizontal axis is labeled Age of Students in Algebra 1, and the vertical axis is labeled Frequency.  Which of the following options is true about the relationship between the mean and median of the data in the provided graph?  Option #1: The mean is greater than the median.  Option #2: The mean is less than the median.  Option #3: The mean is equal to the median.  Option #\_\_\_ is the true statement. | 1 |
| P 2 | Use the image to answer the question.  A dot plot is titled Number of Pets. The number line ranges from 0 to 6 in increments of 1.  Estimate the median of the data in the dot plot.  The median of the data set is \_\_\_\_. | 1 |
| P 3 | In what type of data would you use the mean as a balance point to describe the data? Select the correct option below.  Option #1: bell-shaped symmetrical data  Option #2: right-leaning data  Option #3: left-leaning data  Option #4: U-shaped data  Option #\_\_\_\_ | 4 |
| P 4 | Use the image to answer the question.  A dot plot is graphed on a horizontal axis labeled 0 to 8. The data are 3 dots above 0, 2 dots above 1, 1 dot above 2 and 3, 0 dots above 4, 1 dot above 5 and 6, 2 dots above 7, and 3 dots above 8. The axis is labeled yearly vacations.  Use the mean to find the balancing point of the given dataset in the graph.  The balancing point of the dataset is \_\_\_\_. | 4 |
| P 5 | Use the table to answer the question.    Create a dot plot for the given data, and then use the mean to find the balancing point of the data.  The balancing point of the data is \_\_\_. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the image to answer the question.  A bar graph titled Number of Classes has 5 bars. The horizontal axis is labeled Number of Classes, and the vertical axis is labeled Frequency.  Estimate the median of the data in the histogram. | The median is in the bin 10-14.9 |
| Q 2 | Use the image to answer the question.  A dot plot is titled Number of Hours Calculus Students Studied For Exams. The number line ranges from 0 to 7 in increments of 1.  Estimate the mean and median of the data in the dot plot, and then determine which of the following statements is true about the data. | The mean of the data set is greater than the median. |
| Q 3 | Use the image to answer the question.  A dot plot is titled Number of States Visited. The number line ranges from 2 to 9 in increments of 1.  Estimate the mean of the data in the dot plot. | The mean number of states visited is 5. |
| Q 4 | Explain a situation when neither the mean nor the median is a good descriptor for a dataset. | when the data is U-shaped |
| Q 5 | Use the table to answer the question.    Create a dot plot for the data in the table. Use the mean to find the balancing point of the dataset. | The balancing point is 5. |

**Lesson 5 – Deviation from the Mean**

**Key Words:**

* **deviation** – the difference between a value in a frequency distribution and a fixed number (such as the mean)
* **mean** – a measure of the center of a dataset found by adding all items in a set and dividing by the total number of items

**Objective 1:** In this section, you will calculate the deviations from the mean for symmetric datasets that have the same mean.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

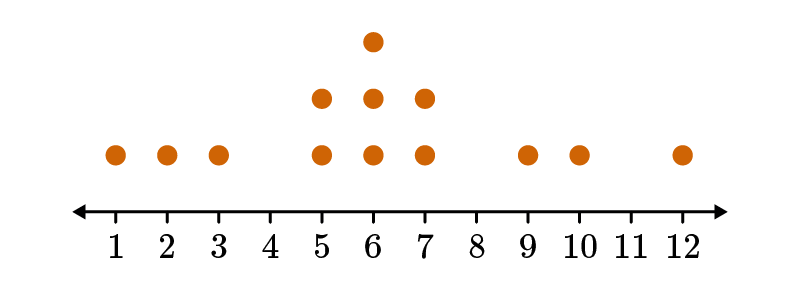
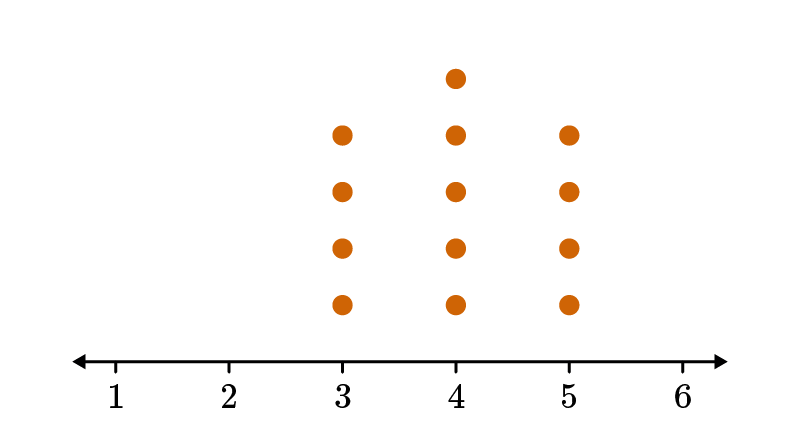
* The *average deviation* tells you the average distance a data point is from the *mean*. In other words, the *average deviation* can tell you how spread out the data is.
  + If the spread is large, then the *average deviation* will be higher.
  + If the spread is small, or clumped, the *average deviation* will be lower.

|  |  |
| --- | --- |
| **Example:** What is the average deviation from the mean of the following dot plot? | |
| **Step 1:** Find the mean. | Add all of the data points and divide by the total number of data points.   * 2 + 3 + 3 + 4 + 4 + 4 + 5 + 5 + 5 + 5 + 6 + 6 + 6 + 7 + 7 + 8 = 80   The mean is 5. |
| **Step 2:** Find the **distance difference** between each data point and the mean. Use a table to help organize your work. | The mean is 5. How far is each data point from the number 5?   |  |  | | --- | --- | | **Data Point** | **Deviation** | | 2 | 3 | | 3 | 2 | | 3 | 2 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 5 | 0 | | 5 | 0 | | 5 | 0 | | 5 | 0 | | 6 | 1 | | 6 | 1 | | 6 | 1 | | 7 | 2 | | 7 | 2 | | 8 | 3 | |
| **Step 3:** Find the mean of the deviations. | Now, take the deviations and perform the same process from Step 1 to find the mean.   * 3 + 2 + 2 + 1 + 1 + 1 + 0 + 0 + 0 + 0 + 1 + 1 + 1 + 2 + 2 + 3 = 20   The average deviation is 1.25. |

**Objective 2:** In this section, you will compare deviations, interpreting larger deviations as greater spread or variability and smaller deviations as smaller spread or variability.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* Recall that the *average deviation* measures the spread, or variability, of the data.
  + Large *deviation* means that the data is more spread out.
    - For example, this dot plot would have a larger deviation because the data is spread to the edges of the graph.
    - 
  + Small *deviation* means that the data is less spread out, more clustered around the mean.
    - For example, this dot plot would have a smaller deviation because the data is more clustered around the mean and less spread out over the graph.
    - 

|  |  |
| --- | --- |
| **Example:** Consider the descriptions of the following two graphs and compare them. What can you tell about the graphs?  Graph 1: mean = 70; deviation = 1  Graph 2: mean = 70; deviation = 4 | |
| **Step 1:** Compare the means. | Both descriptions express that the means of the data are equivalent. Since the means are equivalent, the graphs will have the same balancing point. |
| **Step 2:** Compare the deviations. | When you compare the deviations, the second graph has a deviation that is 3 points higher than that of the first graph. Since the second set has a higher deviation, we know that the points will be spread further out from the mean.  Graph 1 data is more clustered together than Graph 2 data. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Complete the table to find the deviation of each data point. | 1. 2.25 2. 0.25 3. 0.25 4. 2.75 |
| P 2 | Dataset A, {7,9,9,12}, and dataset B, {6,8,10,13}, both have a mean of 9.25. Calculate the average deviation from the mean of each dataset.  The average deviation from the mean of dataset A is \_\_\_. The average deviation from the mean of dataset B is \_\_\_\_. | 1. 1.375 2. 2.25 |
| P 3 | Use the image to answer the question.  A horizontal number line with arrows at both ends is titled at the top as Soccer Team A. Below the line is the title, Number of Goals per Game. The line is numbered 0 to 5 in increments of 1 and has a total of 10 closed points plotted.  A soccer team tracks the number of goals they score in a game on the dot plot. Complete the table to record the deviation of each data point. | 1. 2 2. 1 3. 0 4. 1 5. 2 |
| P 4 | Use the tables to answer the question.    Compare the deviations in the datasets, and determine which has the larger average deviation. Enter the option number of your response.  Option #1: Table 1 has the larger average deviation.  Option #2: Table 2 has the larger average deviation. | 1 |
| P 5 | Use the images to answer the question.      Compare the deviations and determine which graph has the larger average deviation: Graph #1 or Graph #2.  Graph #\_\_\_ has the larger average deviation. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Five friends recorded the number of action figures each friend has. The following is the dataset they recorded: {7,9,10,11,11}. Which of the following correctly calculates the average deviation from the mean number of action figures? | The average deviation from the mean is 1.28. |
| Q 2 | Use the image to answer the question.  A horizontal number line with arrows at both ends is titled at the top as Soccer Team B. Below the line is the title, Number of Goals per Game. The line is numbered 0 to 5 in increments of 1 and has a total of 10 closed points plotted.  Which of the following correctly calculates the average deviation from the mean number of goals per game scored by soccer team B? | The average deviation from the mean is 0.4. |
| Q 3 | Given the following descriptions, which of the following correctly compares the average deviations and describes how the center of graph 1 relates to that of graph 2?  graph 1: mean=7; deviation=5  graph 2: mean=10; deviation=5 | The center of graph 1 is three places away from graph 2 and points are on average five away from the center in both graphs. |
| Q 4 | Use the images to answer the question.      Which of the two graphs has a larger average deviation? | The first graph has a larger average deviation. |
| Q 5 | Use the tables to answer the question.    Compare the datasets in the tables. Which of the following correctly describes the relationship of their average deviations? | The second table has a larger average deviation. |

**Lesson 6 – Standard Deviation**

**Key Words:**

* **deviation** – the difference between a value in a frequency distribution and a fixed number (such as the mean)
* **standard deviation** – a measure of variation whose calculation is based on the distance between each data value and the mean
* **variance** - the average squared distance between the data values and the mean of a data set
* **variation** – a measure of the change in data, a variable, or a function

**Formulas:**

* Standard Deviation:

**Objective 1:** In this section, you will calculate the standard deviation of symmetric datasets.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* Recall that the *average deviation* represents how spread out a set of data points are from the *mean*.
* The *standard deviation* requires more calculation and helps interpret how significant a data point is compared to the graph.

|  |
| --- |
| **Standard Deviation** |
|  |

* Here is what each piece of the formula means and the sequence of steps to take in order to calculate the *standard deviation* of a set of data:
  + 1. means to take each data value, , subtract the mean, , then square each result.
  + 2. means to add all the values from step 1.
  + 3. Divide the result of step 2 by one fewer than the number of values in the dataset, . This is called the *variance*.
  + 4. Finally, take the square root of the result from step 3.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Example:** Find the standard deviation of the given data set.   |  |  |  |  | | --- | --- | --- | --- | | Value | 6 | 7 | 8 | | Total Observations (frequency) | 1 | 3 | 1 | | |
| **Step 1:** Find the mean of the dataset. | 6 + 7 + 7 + 7 + 8 = 35  The mean is 7. |
| **Step 2:** Find the difference between the mean and each data point. | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Data Point,** | 6 | 7 | 7 | 7 | 8 | | **Difference from Mean,** | -1 | 0 | 0 | 0 | 1 | |
| **Step 3:** Square the differences found in Step 2. | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Difference from Mean,** | -1 | 0 | 0 | 0 | 1 | | **Square of the Difference from Mean,** | 1 | 0 | 0 | 0 | 1 | |
| **Step 4:** Divide the sum of the squared differences by *n-1*. | The variance is 0.5 |
| **Step 5:** Find the square root of the variance. | The standard deviation is approximately 0.71. |

**Objective 2:** In this section, you will interpret the standard deviation as a measure of the spread of variability for a symmetric data set that represents the typical distance from a data point.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* Both *standard deviation* and *average deviation* measure the spread of data.
* The *standard deviation* is the average distance the data points are away from the mean.
* [Recall](#Bookmark1) how to calculate *standard deviation*.

|  |  |
| --- | --- |
| **Example:** Given the following two datasets, find and interpret the standard deviation of each.  data set A: 2, 2, 5, 6, 9, 10, 11, 14, 15, 18, 18  data set B: 2, 5, 5, 6, 6, 10, 14, 14, 15, 15, 18 | |
| **Step 1:** Calculate the [standard deviation](#Bookmark1) for each data set. | * The standard deviation of dataset A is 5.83. * The standard deviation of dataset B is 5.40. |
| **Step 2:** Interpret the standard deviation. | * Dataset A has a larger standard deviation, showing that it is more spread out than dataset B. * Dataset B is more consistent than dataset A because it is not as spread out. |

**Objective 3:** In this section, you will use standard deviation to compare the relative variability of distributions.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* *Standard deviation* is useful when comparing the *variability* of distributions.
* The purpose of comparing *standard deviations* is to observe the changes which can show trends in data.
* If two data sets have the same *standard deviation*, that means that they have the same *variability* (spread) of data points.

|  |  |
| --- | --- |
| **Example:** A toy company produces two different types of toys: Toy A and Toy B. The company needs to compare the variability in the production quality of the two toy types to determine **which one has a more consistent output**. They collect data on the number of defective toys produced each day for both toy types over a span of 10 business days.  The data is summarized as follows: | |
| **Step 1:** Calculate the mean and [standard deviation](#Bookmark1) of Toy A. | * Calculate the mean.   + The mean, , is 1.9. * Calculate the [standard deviation](#Bookmark1).   + The standard deviation is 1.2. |
| **Step 2:** Calculate the mean and [standard deviation](#Bookmark1) of Toy B. | * Calculate the mean.   + The mean, , is 2.7. * Calculate the [standard deviation](#Bookmark1).   + The standard deviation is 2.11. |
| **Step 3:** State the answer. | * Toy A has less variability or inconsistency in production quality. * The toy company can conclude that Toy A shows a more consistent production quality compared to Toy B. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Leah asked five of her friends how many sports they play. She received the following responses: 1, 1, 2, 3, and 3. Calculate the standard deviation of the data.  The standard deviation is \_\_\_\_\_. | 1 |
| P 2 | Use the table to answer the question.    Calculate the standard deviation of the dataset. Round the answer to the nearest hundredth.  The standard deviation is \_\_\_\_\_. | 1.22 |
| P 3 | Use the table to answer the question.    Which dataset has the **highest** variability?  Dataset \_\_\_\_\_ has the highest variability. | 2 |
| P 4 | Use the table to answer the question.  Find the standard deviation of the data set, rounded to the nearest hundredth.    The standard deviation of the dataset is \_\_\_\_\_. | 1. 0 2. -4 3. 16 4. 2.74 |
| P 5 | Find the standard deviation to the nearest hundredths and then compare the variability of the datasets. Enter the number of the correct option.  Dataset A: 10, 8, 10, 8, 8, 5  Dataset B: 9, 10, 8, 10, 11, 6  Option #1: The standard deviation of Dataset B is 1.79. The standard deviation of Dataset A is 1.83. The standard deviation of Dataset B is larger than that of Dataset A, so Dataset B has more variability.  Option #2: The standard deviation of Dataset A is 1.67. The standard deviation of Dataset B is 1.63. The standard deviation of Dataset A is larger than that of Dataset B, so Dataset A has more variability.  Option #3: The standard deviation of Dataset A is 1.83. The standard deviation of Dataset B is 1.79. The standard deviation of Dataset A is larger than that of Dataset B, so Dataset A has more variability. | 3 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    Calculate the standard deviation of the data set. Round the answer to the nearest hundredth. | 0.71 |
| Q 2 | Use the table to answer the question.    Charles records the number of miles he walks every day. The chart includes how many days he walked the given number of miles for one week. Calculate the standard deviation for the number of miles he walked. Round the answer to the nearest hundredth. | 0.82 |
| Q 3 | Compare and interpret the measures to select the data with the **least** variation. | Dataset D has a standard deviation of 1.14 and a mean of 48. |
| Q 4 | If the mean of a dataset is 59, what is the deviation of the data point at 43? | -16 |
| Q 5 | Use the standard deviation to compare the variability of the datasets. Round to the nearest hundredths.  Dataset A: 11, 7, 12, 8, 9, 5  Dataset B: 10, 8, 10, 8, 8, 4 | The standard deviation of Dataset A is 2.58. The standard deviation of Dataset B is 2.19. The standard deviation of Dataset A is larger than that of Dataset B, so Dataset A has more variability. |

**Lesson 7 – Interquartile Range**

**Key Words:**

* **box plot** – a graphical way to display the median, quartiles, and extremes of a dataset
* **first quartile** – the value that is one quarter of the way through the dataset, calculated by finding the median of the lower half of the dataset
* **five-number summary** – a list of five calculations including the minimum, maximum, median, first quartile, and third quartile
* **interquartile range (IQR)** – the measure of variability of a dataset that is the difference between the first quartile and third quartile
* **mean** – a measure of the center of a dataset found by adding all items in a set and dividing by the total number of items
* **median** – the middle value in an ordered set of data, or the mean of the two middle numbers
* **outlier** – a statistical observation that is markedly different in value from the others of the sample
* **range** – the difference between the least and greatest values of an attribute or of the variable of a frequency distribution
* **skewed** – not symmetrical
* **third quartile** – the value that is three quarters of the way through the dataset, calculated by finding the median of the upper half of the dataset

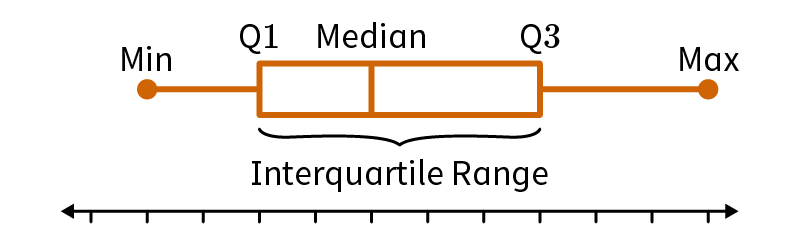
**Formulas:**

* Interquartile Range:

**Objective 1:** In this section, you will construct box plots by calculating five-number summaries and interquartile ranges for skewed datasets and identify any outliers.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* [Recall](#Bookmark2) how *box plots* are created using the *five-number summary*.
  + *Five-number summary*: minimum, first quartile (Q1), median, third quartile (Q3), maximum
  + 
* Data represented in a box plot can be symmetrical or skewed. The data is symmetrical when both of these occur:
  + The ends of the box are equidistant from the median line inside the box
  + The whiskers are the same lengths
* The *interquartile range (IQR)* of a dataset measures the spread of the middle 50% of the data. You can calculate the *interquartile range* using the formula:
* The interquartile range can be used to identify outliers in a dataset using the following calculations.
  + Lower fence =
    - Any number below this value is considered an outlier.
  + Upper fence =
    - Any number above this value is considered an outlier.

|  |  |
| --- | --- |
| **Example:** Use the interquartile range to help identify any outliers in Eithan’s dataset.  Dataset: 0, 15, 23, 29, 30, 38, 45, 52, 53, 59, 68, 68, 72, 78, 80, 90, 101, 105, 231, 670 | |
| **Step 1:** Calculate the [five-number summary](#Bookmark2). | * Minimum: 0 * First/Lower Quartile (Q1): 34 * Median: 63.5 * Third/Upper Quartile (Q3): 85 * Maximum: 670 |
| **Step 2:** Calculate the IQR. | * The Interquartile Range is 51. |
| **Step 3:** Calculate the lower fence of the data set. | * Any value below –42.5 would be considered an outlier. Eithan does not have any outliers in the lower end of his dataset. |
| **Step 3:** Calculate the upper fence of the data set. | * Any value above 161.5 would be considered an outlier. Both 231 and 670 are outliers in the upper end of Eithan’s dataset. |
| **Step 4:** State the answer. | Eithan’s dataset contains two outliers: 231 and 670. |

**Objective 2:** In this section, you will interpret the interquartile range as a way to describe the variability for skewed datasets.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* The *interquartile range* can be used to describe the **middle 50%** of the data to describe the *variability* of datasets, especially when they are skewed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Example:** Mrs. Stiltner is comparing the test scores of her classes. Find and compare the interquartile ranges (IQRs) of the two classes. What do they mean?   |  |  |  | | --- | --- | --- | |  | **Period 1** | **Period 2** | | **Minimum** | 32 | 60 | | **Q1** | 70 | 65 | | **Median** | 75 | 75 | | **Q3** | 78 | 84 | | **Maximum** | 95 | 100 | | |
| **Step 1:** Calculate the IQR for Period 1. | * The IQR for Period 1 is 8. |
| **Step 2:** Calculate the IQR for Period 2. | * The IQR for Period 2 is 19. |
| **Step 3:** Compare and interpret the IQRs. | This means that there is higher variability in the scores of Period 2. The middle 50% of students in Period 1 had scores that were closer together. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Calculate the interquartile range of the skewed dataset 0, 2, 6, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 10.  The interquartile range=\_\_\_\_\_ | 1.5 |
| P 2 | Calculate the values for the 5-number summary of the skewed dataset 11, 12, 12, 13, 13, 13, 14, 14, 15, 19, 25. | 1. 11 2. 12 3. 13 4. 15 5. 25 |
| P 3 | Determine if there are any outliers in the skewed dataset 11, 12, 12, 13, 13, 13, 14, 14, 15, 19, 25. Enter the number of the correct option.  Option #1: There are no outliers in the dataset.  Option #2: The value 25 is the only outlier in the dataset.  Option #3: The values 25 and 19 are the only two outliers in the dataset.  Option #4: The values 25, 19, and 15 are the only three outliers in the dataset.  The accurate statement is Option #\_\_\_\_\_. | 2 |
| P 4 | A meteorologist recorded the following temperatures: {71,74,81,78,67,69,64}. Which best describes the variability of the temperature? | The IQR is 11, indicating a fairly high variability for the temperatures for half of the recorded days. |
| P 5 | A meteorologist recorded the following temperatures: {71,74,81,78,67,69,64}. Today's temperature is 67º. If today’s temperature is included in the dataset, how does it change the interquartile range?  The interquartile range of the new dataset is \_\_\_. It changed by \_\_\_\_. | 1. 9 2. 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Calculate the 5-number summary of the skewed dataset 15, 16, 16, 17, 17, 18, 18, 18, 19, 19, 29. When constructing the box plot on a number line, which explains a reason why the dataset is skewed? | The right whisker is much longer than the left whisker. |
| Q 2 | Calculate the 5-number summary of the skewed dataset 5, 6, 6, 7, 7, 8, 8, 8, 9, 9, 14, 16. Which of the following choices is quartile 3? | 9 |
| Q 3 | Which option is true about the outliers of the skewed dataset 5, 6, 6, 7, 7, 8, 8, 8, 9, 9, 14, 16? | The outliers of the dataset are 14 and 16. |
| Q 4 | A stock analyst is comparing the interquartile range of the day’s stock prices of Stock A and Stock B. Stock A has an interquartile range of 3, and Stock B has an interquartile range of 11. What interpretation can be made from the interquartile range of each stock? | Stock B has more variability than Stock A. |
| Q 5 | A video game player had the following scores: 287, 389, 287, 329, 295, 290, 397, 387, 365. What does the IQR indicate about this dataset? | An IQR of 99.5 indicates that the video game player had a relatively small variability for their middle scores. |

**Lesson 8 – Comparing Distributions**

**Key Words:**

* **measure of central tendency** – a value intended to indicate the center of values in a collection of data; typically the mean, median, or mode
* **skewed distribution** – a distribution in which the chart's tail is longer on one side than the other
* **symmetric distribution** – a distribution in which the left and right sides mirror each other

**Objective 1:** In this section, you will compare two or more related symmetric distributions.

*Mathematical Practice Standard: Use appropriate tools strategically.*

**Big Ideas**:

* Recall that *mean*, *median*, and *mode* are each a *measure of central tendency* used to describe the middle of a dataset.
* Recall that a*symmetrical distribution*creates a graph in which a line can be drawn straight through the middle to create to mirror, or symmetric, images.
* In a *symmetric distribution*, the *mean* and *median* are located at the same point, they are the same.
  + This is true for symmetric distributions on a histogram, dot plot, or box plot.
  + See the three symmetric examples below.

|  |  |
| --- | --- |
| **Example** | **Measure of Center** |
|  | The mean and median are located in the center bin or precisely between the two center bins.  In this example, the mean and median are located in the bin between 24 and 30.  An estimate of the mean and median is . |
|  | Despite its unusual shape, this dot plot is still considered to be symmetric.  The mean and median are the same value, 5. |
|  | The box plot contains the median at exactly the center of the box, at the value of 70.  Therefore, the box plot is symmetric, and the mean is located in the same place.  The mean and median is 70. |

**Objective 2:** In this section, you will compare two or more related skewed distributions.

*Mathematical Practice Standard: Use appropriate tools strategically.*

**Big Ideas:**

* In *skewed distributions*, the best measure of central tendency is the *median*.
* Use the following graphs to compare how the *mean* moves when the graph is skewed.

|  |  |  |
| --- | --- | --- |
| **Right-Skewed** | **Symmetric** | **Left-Skewed** |
| The data is clustered on the left side of the chart and tails off on the right. The mean is greater than the median, and the median is greater than the mode. | The data is symmetrical. The mean and median are the same, at the center of the graph. | The data is clustered on the right side of the chart and tails off to the left. The mean is less than the median, and the median is less than the mode. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the images to answer the question.  A bar graph is titled 'Quiz Scores, Class A.'  A histogram is titled 'Quiz Scores, Class B.' The x-axis ranges from 60 to 81 and the y-axis ranges from 0 to 4, both by 1-unit increments. Six bars are drawn.  A histogram is titled 'Quiz Scores, Class C.' The x-axis ranges from 62 to 94 and the y-axis ranges from 0 to 3, both by 1-unit increments. Eight bars are drawn.  Compare the graphs that summarize the quiz scores for Class A, Class B, and Class C. All three classes had symmetrical datasets. Another student from each class will take the quiz. Which answer best describes which student’s score will be easiest to predict? | The score for a new student in Class A will be easiest to predict. It is likely that the score will be within 6 points of 70 because the IQR and range are relatively low. |
| P 2 | Use the images to answer the question.  A histogram is titled 'Quiz Scores, Class A.' The x-axis ranges from 65 to 76 and the y-axis ranges from 0 to 5, both by 1-unit increments. 5 bars are drawn.  A histogram is titled 'Quiz Scores, Class B.' The x-axis ranges from 60 to 81 and the y-axis ranges from 0 to 4, both by 1-unit increments. Six bars are drawn.  A histogram is titled 'Quiz Scores, Class C.' The x-axis ranges from 62 to 94 and the y-axis ranges from 0 to 3, both by 1-unit increments. Eight bars are drawn.  Three different classes took the same math quiz. The scores are summarized in the graphs for Class A, Class B, and Class C. Which statement best describes the class that had the greatest variability in quiz scores? | Class C had the greatest variability because it had the greatest range and IQR of any class dataset. |
| P 3 | Use the images to answer the question.  A histogram is titled 'Quiz Scores, Class A.' The x-axis ranges from 65 to 76 and the y-axis ranges from 0 to 5, both by 1-unit increments. 5 bars are drawn.  A histogram is titled 'Quiz Scores, Class B.' The x-axis ranges from 60 to 81 and the y-axis ranges from 0 to 4, both by 1-unit increments. Six bars are drawn.  Class A and Class B took the same math quiz. The scores are summarized in the histograms for Class A and Class B. Which statement best compares the scores of the two classes? | The range of quiz scores is greater in Class B than in Class A. |
| P 4 | Use the images to answer the question.  A histogram is titled 'Points Scored per Player from Eastern High School in the Championship Game.' The x-axis ranges from 0 to 11 and the y-axis ranges from 0 to 3, both by 1-unit increments.  A histogram is titled 'Points Scored per Player from Western High School in the Championship Game.' The x-axis ranges from 0 to 17 and the y-axis ranges from 0 to 3, both by 1-unit increments.  Compare the skewed distributions that summarize the number of points scored per player in the championship basketball game between Eastern High School and Western High School. Based on the distributions, which statement best compares the two teams? | The Eastern High School data is skewed left, while the Western High School data is skewed right. |
| P 5 | Use the images to answer the question.  A histogram is titled 'Points Scored per Player from Eastern High School in the Championship Game.' The x-axis ranges from 0 to 11 and the y-axis ranges from 0 to 3, both by 1-unit increments.  A histogram is titled 'Points Scored per Player from Western High School in the Championship Game.' The x-axis ranges from 0 to 17 and the y-axis ranges from 0 to 3, both by 1-unit increments.  Compare the skewed distributions that summarize the number of points scored per player in the championship basketball game between Eastern High School and Western High School. Which statement correctly compares the medians of the two distributions? | The median for Eastern High School is several points higher than that of Western High School because the Eastern’s dataset is skewed left. |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the image to answer the question.  Two box plots, 1 below the other, are over a number line ranging from 20 to 100 in increments of 10. The plots are labeled Test Grades. The top box plot represents No Homework and the bottom box plot represents Homework.  Compare the box plots summarizing the test scores of students who completed their homework versus those who did not complete their homework. Which statement is correct? | On average, students who completed their homework scored higher than students who did not. |
| Q 2 | Use the image to answer the question.  Two box plots, 1 below the other, are over a number line ranging from 20 to 100 in increments of 10. The plots are labeled Test Grades. The top box plot represents No Homework and the bottom box plot represents Homework.  Compare the box plots summarizing the test scores of students who completed their homework versus those who did not complete their homework. What was the lowest score among the students who did complete their homework? | 60 |
| Q 3 | Use the images to answer the question.  A histogram is titled 'Exam Scores per Student in Mr. Hanley's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 8 by 2-unit increments. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  A histogram is titled 'Exam Scores per Student in Ms. Blach's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 5 by increments of 1. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  Compare the skewed distributions. Mr. Hanley and Ms. Balch gave the same Algebra final exam to their classes. The histograms show the distribution of the scores for each class. In which range would you expect to find the median exam score for Ms. Balch’s class? | between 60 and 70 |
| Q 4 | Use the images to answer the question.  A histogram is titled 'Exam Scores per Student in Mr. Hanley's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 8 by 2-unit increments. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  A histogram is titled 'Exam Scores per Student in Ms. Blach's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 5 by increments of 1. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  Mr. Hanley and Ms. Balch gave the same Algebra ﬁnal exam to their classes. The histograms show the distribution of the scores for each class. Which best compares the overall scores of the two classes? | Mr. Hanley’s class scored higher overall, as shown by the left-skewed shape of the graph. |
| Q 5 | Use the images to answer the question.  A histogram is titled 'Exam Scores per Student in Mr. Hanley's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 8 by 2-unit increments. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  A histogram is titled 'Exam Scores per Student in Ms. Blach's Class.' The x-axis ranges from 40 to 100 by 10-unit increments and the y-axis ranges from 0 to 5 by increments of 1. The x-axis is labeled 'Exam Scores' and the y-axis is unlabeled.  Mr. Hanley and Ms. Balch gave the same Algebra ﬁnal exam to their classes. The histograms show the distribution of the scores for each class. One of the classes will get extra tutoring following the exam to help them with the topics they missed on the exam. Which class likely needs more tutoring help, and why? | Ms. Balch’s class needs the tutoring because the class data is skewed right, and her class median is lower. |