

Parent Newsletter

Chapter 12: Data Analysis and Displays

Students will...

- Find the mean, median, and mode of a data set.
- Identify and remove outliers.
- Explain the effects of changing values in data sets.
- Find ranges of data sets.
- Find standard deviations of data sets.
- Make and interpret box-and-whisker plots.
- Find interquartile ranges of data sets.
- Compare box-and-whisker plots.
- Describe shapes of distributions.
- Choose appropriate measures of central tendency and dispersion to represent data sets.
- Interpret scatter plots.
- Identify relationships from scatter plots.
- Find lines of fit.
- Solve real-life problems.
- Use residuals to determine whether models are a good fit.
- Find lines of best fit using technology.
- Identify correlations and causations.
- Read two-way tables.
- Find marginal frequencies.
- Make two-way tables.
- Find relationships in two-way tables.
- Choose appropriate data displays
- Identify and analyze misleading data displays.



Shapes of Box-and-Whisker Plots

Skewed left:

- Left whisker longer than right whisker
- Most data on the right

Symmetric:

- Whiskers about the same length
- Median in the middle of the data

Skewed right:

- Right whisker longer than left whisker
- Most data on the left

Key Ideas

Mean

The mean of a data set is the sum of the data divided by the number of data values.

Median

Order the data. For a set with an odd number of values, the median is the middle value. For a set with an even number of values, the median is the mean of the two middle values.

Mode

The mode of a data set is the value or values that occur most often.

Standard Deviation

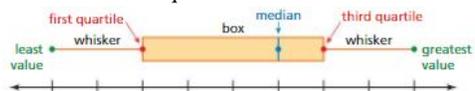
The standard deviation of a data set is a measure of how much a typical value in the data set differs from the mean. It is given by

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

where n is the number of values in the data set. The symbol \bar{x} represents the mean. It is read as “ x -bar.”

Box-and-Whisker Plot

A box-and-whisker plot displays a data set along a number line using medians. Quartiles divide the data set into four equal parts. The median (second quartile) divides the data set into two halves. The median of the lower half is the first quartile. The median of the upper half is the third quartile.



Symmetric and Skewed Distributions

Skewed left:

- The “tail” of the graph extends to the left.
- Most data are on the right.

Symmetric

- The data are evenly distributed on each side of the highest bar.

Skewed right:

- The “tail” of the graph extends to the right.
- Most data are on the left.

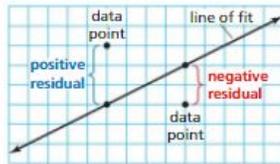
Key Ideas

Scatter Plot

A scatter plot is a graph that shows the relationship between two data sets. The two sets of data are graphed as ordered pairs in a coordinate plane.

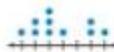
Residuals

A residual is the difference between the y -value of a data point and the corresponding y -value found using the line of fit. A residual can be positive, negative, or zero.



Data Displays

- Pictograph: shows data using pictures 
- Bar Graph: shows data in specific categories 
- Circle Graph: shows data as parts of a whole 
- Line Graph: shows how data change over time 
- Histogram: shows frequencies of data values in intervals of the same size. 
- Stem-and-Leaf Plot: orders numerical data and shows how they are distributed

1	0	2	3	6
2	1	1	5	
3	9			
4	0	6		
- Box-and-Whisker Plot: shows the variability of a data set using quartiles 
- Dot Plot: shows the number of times each value occurs in a data set 
- Scatter Plot: shows the relationship between two data sets using ordered pairs in a coordinate plane 

Key Terms

A **measure of central tendency** is a measure that represents the center of a data set.

A **measure of dispersion** is a measure that describes the spread of a data set.

The **range** of a data set is the difference between the greatest value and the least value.

The **standard deviation** of a data set is a measure of how much a typical value in the data set differs from the mean.

A **box-and-whisker plot** displays a data set along a number line using medians.

Quartiles divide a data set into four equal parts.

The five numbers that make up a box-and-whisker plot are called the **five-number summary** of the data set.

The **interquartile range** is the difference of the third quartile and the first quartile.

A **scatter plot** is a graph that shows the relationship between two data sets.

A **line of fit** is a line drawn on a scatter plot close to most of the data points.

A **residual** is the difference between the y -value of a data point and the corresponding y -value found using the line of fit.

The process used to find the line of best fit for a set of data is **linear regression**.

A precise line of fit that best models a set of data is the **line of best fit**.

A value that tells whether the correlation is positive or negative and how closely an equation models the data is a **correlation coefficient**.

When a change in one variable x results in a change in another variable y , it is called **causation**.

A **two-way table** displays two categories of data collected from the same source.

Each entry in the table is called a **joint frequency**.

The sums of the rows and columns in a two-way table are called **marginal frequencies**.

Games

- M and M and M

This is available online in the Game Closet at www.bigideasmath.com.

What's the Point?

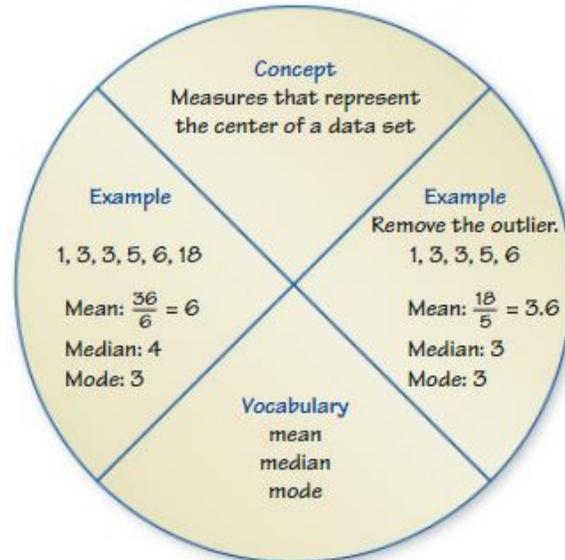
The STEM Videos available online show ways to use mathematics in real-life situations. The Chapter 12: Fuel Economy Part 2 STEM Video is available online at www.bigideasmath.com.



Reference Tools

A **Concept Circle** can be used to organize information about a concept. Students write the concept above the circle. Then students write associated information in the sectors of the circle. Concept circles can have more or fewer than four sectors. Students can place their concept circles on note cards to use as a quick study reference.

Measures of Central Tendency



Standards

California Common Core:

8.SP.1, 8.SP.4, S.ID.1, S.ID.2, S.ID.3, S.ID.5, S.ID.6a, S.ID.6b, S.ID.6c, S.ID.7, S.ID.8, S.ID.9

Quick Review

- Data can have one mode, more than one mode, or no mode. When each data value occurs only once, there is no mode.
- An outlier is a data value that is much greater than or much less than the other values.
- A box-and-whisker plot shows the variability of a data set.
- If you can draw a line through the median of a box-and-whisker plot, and each side is a mirror image of the other, then the distribution is symmetric.
- If all of the bars of a histogram are the same height, then the distribution is a flat, or uniform, distribution. A uniform distribution is also symmetric.
- Outliers can affect the mean of a data set more than they affect the median.
- Scatter plots can show unusual features of a data set, such as outliers, or gaps and clusters in the data.
- A line of fit does not need to pass through any of the data points.
- A causal relationship exists when one variable causes a change in another variable.

Essential Questions

- How can you use measures of central tendency to distribute an amount evenly among a group of people?
- How can you measure the dispersion of a data set?
- How can you use a box-and-whisker plot to describe a data set?
- How can you use a histogram to characterize the basic shape of a distribution?
- How can you use data to predict an event?
- How can you find a line that best models a data set?
- How can you read and make two-way tables?
- How can you display data in a way that helps you make decisions?