Algebra 1

**Bivariate Datasets**

**Unit Summary:** In this unit, you will learn how to display bivariate data using a two-way frequency table or a scatterplot. You will learn how to draw a scatterplot to represent the relationship between the data.

**Lesson 2 – Two-Way Frequency Tables**

**Key Words:**

* **bivariate data** – a type of data that derives from, relates to, or involves two variables
* **categorical data** – a type of data that can be broken down into two or more classes or divisions
* **data** – factual information (such as measurements or statistics) used as basis for reasoning, discussion, or calculation
* **numerical data** – the data represented in terms of a number, quantity, or measurement
* **two-way frequency table** – a type of table that displays data from one source organized into two separate variables
* **univariate data** – a type of data characterized by or depending on only one random variable
* **Venn diagram** – a graph that employs closed curves and especially circles to represent logical relations between and operations on sets and the terms of propositions by the inclusion, exclusion, or intersection of the curves

**Objective 1:** In this section, you willdistinguish the differences between numerical and categorical data.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* *Categorical data* is any feature that can be ordered together to create a group.
  + For example: hair color, birth month, favorite ice cream, etc.
* *Numerical data* is a quantity that provides a measurement.
  + For example: height, age, number of pets, etc.

|  |  |
| --- | --- |
| **Example:** Milosh compiled a list of his favorite games. Use the table to answer the following questions. | |
| **Which columns are considered categorical data?** | Video Game, Platform, Genre are categorical data. |
| **Which columns are considered numerical data?** | Players and Cost are numerical data. |
| **The cost of the game Massive Car Barrowing at $29.99 can be classified as which type of data?** | The $29.99 game cost can be classified as numerical data. This represents the value of a game in dollars. |
| **Sandbox can be classified as which type of data?** | Sandbox can be classified as categorical data. Sandbox can be put in a genre group. |
| **Describe a change that can be made to the players column to make it categorical.** | The number of players could be changed to two options, either single player or multiplayer. |

**Objective 2:** In this section, you will use two-way frequency tables to compare, organize, and interpret data.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* *Two-way frequency tables* are best for when you need to compare two different categories of information. They can be used to summarize data and make decisions.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Example:** As a member of the student council at your high school, you are tasked with planning the school’s autumn festival. You must decide whether to include a dunking booth or a pie-throwing booth. You are given the following data with the preferences of both juniors and seniors.  Use the two-way frequency table to answer the following questions.   |  |  |  |  | | --- | --- | --- | --- | |  | **Dunking Booth** | **Pie-Throwing Booth** | **Totals** | | **Juniors** | 22 | 18 | 40 | | **Seniors** | 16 | 24 | 40 | | **Totals** | 28 | 42 | 80 | | |
| **How many students in total were surveyed? How many were juniors and how many were seniors?** | * 80 students were surveyed: 40 juniors and 40 seniors |
| **Which game do the juniors prefer to have at the festival?** | * 22 juniors want a dunking booth. * 18 juniors want a pie-throwing booth. * Overall, the juniors prefer the dunking booth. |
| **Which game do the juniors and seniors prefer overall?** | * There are 28 total votes for the dunking booth. * There are 42 total votes for the pie-throwing booth. * Overall, students prefer to have a pie-throwing booth at the festival. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | At a dog show, dogs are organized and shown by breed. Is this an example of categorical or numerical data? Enter 1 for categorical data or 2 for numerical data. | 1 |
| P 2 | Ahmad is involved in a project that tracks weather data for the week. He wants to use only categorical data. If he decides to include the measure of the lowest temperature for each day, would his project still use categorical data only? Enter 1 for yes or 2 for no. | 2 |
| P 3 | Use the two-way frequency table to answer the question. What are the missing values? | 1. 55 2. 68 |
| P 4 | Use the table to answer the question.    Review the two-way frequency table. How many third-grade students did **not** go on a field trip last year? | 104 |
| P 5 | Use the table to answer the question.    According to the table of attendance at a recent event, which night had greater adult attendance? Enter 1 for Friday or 2 for Saturday. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Which statement explains how you can distinguish between numerical and categorical data? | Numerical data sorts by number values such as measurements, whereas categorical data sorts by shared features or attributes. |
| Q 2 | Use the table to answer the question.    Consider the data provided in the table. Which of the columns shows categorical data? | dog breed |
| Q 3 | Use the table to answer the question.    Use the two-way frequency table to summarize how many total students participated in a survey about whether they like rock music. | 538 |
| Q 4 | Use the table to answer the question.    Review the table, which includes data about whether people of different age groups like pineapple on pizza. Which of the following statements is true based on the data in the table? | More adults responded positively than children. |
| Q 5 | Use the table to answer the question.    Riana surveyed people on the street about whether they preferred cats or dogs, and then asked for their age. How many people who are 40 years old or younger said they preferred dogs, according to the table? | 983 |

**Lesson 3 – Relative Frequencies**

**Key Words:**

* **association** – a group of factors in a set of data that indicates a relationship between two or more categorical variables
* **causation** – an act or an agent that creates another event (also known as cause-and-effect)
* **conditional relative frequency** – finding the ratio of frequency in a particular category to the total number of data points in a particular category
* **joint frequency** – the number of times a combination of two conditions occur together
* **joint relative frequency** – the ratio of frequency in a particular category to the total number of data points
* **marginal relative frequency** – the ratio of a sum of joint relative frequencies to the total number of data points
* **relative frequency table** – a table that records relative frequency outcomes

**Formulas:**

* Joint Relative Frequency:
* Marginal Relative Frequency:
* Conditional Relative Frequency:

**Objective 1:** In this section, you will use two-way frequency tables to obtain and interpret relative frequency tables.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* From a *two-way frequency table*, you can create a *relative frequency table.* 
  + *Relative frequency tables* are made of of the *joint relative frequency* and the *marginal relative frequency*.
* *Joint relative frequency* represents the frequency for one particular category.
  + Each entry except the “Total” sections in a two-way frequency table are considered *joint frequency* values.
  + To calculate j*oint relative frequency*, you need to divide the *joint frequency* by the total number of entries in the survey.
    - joint relative frequency =
* *Marginal relative frequency* represents the frequency of all the *joint relative frequencies* in one row or column.
  + The “Total” row and “Total” column make up the *relative marginal frequency* of a table.
  + To calculate *marginal relative frequency*, divide the total for a given category by the total number of entries in the survey.
    - Marginal Relative Frequency =

|  |  |
| --- | --- |
| **Example:** Calculate the joint relative frequencies for each entry in the following two-way frequency table and interpret the results. | |
| **Step 1:** Identify the joint frequency values and the total number surveyed. |  |
| **Step 2:** Calculate each joint relative frequency by dividing the category by the total entries. Interpret each result. | * Football players and sports drink:   + 41 percent of the athletes surveyed are football players who also prefer a sports drink during games. * Soccer players and sports drink:   + 21 percent of the athletes surveyed are soccer players who also prefer a sports drink during games. * Football players and water:   + 11 percent of the athletes surveyed are football players who also prefer water during games. * Soccer players and water:   + 27 percent of the athletes surveyed are soccer players who also prefer water during games. |
| **Step 3:** Identify the marginal frequency values and the total number surveyed. |  |
| **Step 4:** Calculate the marginal relative frequencies and interpret each result. | * Football players:   + 52 percent of all athletes surveyed were football players. * Soccer players:   + 48 percent of all athletes surveyed were soccer players. * Sports Drink:   + 62 percent of all athletes, regardless of sport, prefer sports drinks over water. * Water:   + 38 percent of all athletes, regardless of sport, prefer water over sports drinks. |
| **Step 5:** Create a relative frequency table from the joint and marginal relative frequency values calculated. |  |

**Objective 2:** In this section, you will use two-way frequency tables to calculate and interpret conditional relative frequency.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* *Conditional relative frequency* is useful when you want to interpret only a certain part of the data.
* *Conditional relative frequency* is found by dividing the *joint frequency* of the desired category by the total for that category.
  + Conditional Relative Frequency =

|  |  |
| --- | --- |
| **Example:** Use the two-way frequency table to answer the following questions. | |
| What is the conditional frequency of students who prefer **sneakers** given they also prefer **pants**? | * Students who prefer sneakers given they also prefer pants:   + 76 percent of the surveyed students who prefer pants also prefer sneakers. |
| What is the conditional frequency of students who prefer **pants** given they also prefer **sneakers**? | * Students who prefer pants given they also prefer sneakers:   + 56 percent of the surveyed students who prefer sneakers also prefer pants. |

**Objective 3:** In this section, you will interpret conditional relative frequencies to determine whether variables or categories are the result of association and not causation.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* *Association* in *categorical data* occurs when two variables in a set of data relate to each other.
* For an *association* to exist, the difference in the *conditional relative frequencies* between the two variables must be significant.
* *Causation* is when one event causes another. However, *association* in a data set does not mean *causation* can be found.
* [Recall](#Bookmark1) how to calculate *conditional relative frequencies*.

|  |  |
| --- | --- |
| **Example:** The following table displays the results from a survey of favorite outdoor activities of people who can swim and of people who cannot swim. Does an association exist between the categorical variables? | |
| **Step 1:** Calculate the conditional relative frequencies for a favorite activity given that a person can or can’t swim. | To calculate the conditional relative frequency for each variable, divide each joint frequency by the total for its row or column.   * People who prefer the beach given they can swim: * People who prefer the park given they can swim: * People who prefer the beach given they cannot swim: * People who prefer the park given they cannot swim: |
| **Step 2:** Organize the data into a conditional relative frequency table and interpret the data. | * Given a person can swim, 74% prefer to go to the beach and 26% prefer the park. * Given a person can’t swim, 18% prefer to go to the beach and 82% prefer the park. |
| **Step 3:** Determine the association. | The frequency table shows a noticeable difference between going to the beach and park given the ability to swim. Therefore, there is a strong association between a person’s ability to swim and their favorite outdoor activity. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the table to answer the question.    The table shows survey data about which season students prefer. What is the joint relative frequency of elementary age students who prefer winter? Round to the nearest whole number.  \_\_\_\_\_% | 33 |
| P 2 | Use the table to answer the question.    Using data from the table, what is the marginal relative frequency for drivers? Round your answer to the nearest whole percent value.  \_\_\_\_\_percent | 65 |
| P 3 | Use the table to answer the question.    Interpret the two-way table for students’ drink and food preferences. Given the students that prefer soda, what is the conditional relative frequency that they like hot dogs? Round your answer to the nearest whole percentage.  \_\_\_\_\_% | 33 |
| P 4 | Use the table to answer the question.    Use the two-way frequency table to complete the conditional frequency table for students who prefer soda or water given they prefer hamburgers or hot dogs. Round each percentage to the nearest whole number. | 1. 67 2. 33 |
| P 5 | Use the table to answer the question.    A ninth-grade algebra teacher gave her class a test. The table displays the conditional relative frequency data on whether a student studied and whether the student earned an A on the math test. Select the correct statement about the data.  Option #1: There is no association between studying and earning an A on the test.  Option #2: There is an association between studying and earning an A on the test.  Option #\_\_\_\_ is the correct statement. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    Use the relative frequency table to decide which of the following is a correct interpretation of the data. | More fourth graders responded than third graders. |
| Q 2 | Use the table to answer the question.    Sal kept track of how many people visited the local library and recorded their purpose for visiting. Using marginal relative frequency, what interpretation can you make from the data in the two-way table? | The weekend is the most popular time to visit. |
| Q 3 | Use the table to answer the question.    At Fairmount Elementary School, 108 students were polled about their mode of transit to school and on the distance they live from school. This two-way table summarizes the information. Interpret the table to find the conditional relative frequency of students walking to school, given they live more than a mile away. | 10% |
| Q 4 | Use the table to answer the question.    High school juniors were asked about how many hours a week they practice their musical instruments to prepare for a musical competition. The study included 60 participants, and its intent was to see if there was a difference between competition levels of 1 to 3 and levels 4 or higher. The results are summarized in the table. Which statement is a correct association that can be made based on the data provided? | There is an association between the competition level and the number of hours of practice. |
| Q 5 | Use the table to answer the question.    At a local amusement park, parents were asked whether they were season pass holders and the ages of their children. The results of the survey are in the table. Which statement describes a correct association for the data? | There is an association between the ages of children and whether a parent is a season pass holder. |

**Lesson 4 – Relationships Between Two Numerical Variables**

**Key Words:**

* **exponential equation** – a mathematical statement in which the exponent is a variable
* **line of best fit** – a straight line that minimizes the distance between it and the data, also called the least squares regression line
* **linear relationship** – a pattern created when the relationship between the independent and dependent variables remains constant
* **nonlinear relationship** – a pattern created when the relationship between the independent and dependent variables does not remain constant
* **parabola** – a curve where any point is equidistant from a fixed point (the focus) and a fixed straight line (the directrix); the focus may not lie on the directrix
* **quadratic equation** – any equation containing one term in which the unknown is squared and no term in which it is raised to a higher power
* **scatterplot** – a two-dimensional graph in rectangular coordinates consisting of points whose coordinates represent values of two variables under study

**Objective 1:** In this section, you will distinguish between linear and nonlinear relationships in scatterplots.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* A *scatterplot* is a type of chart designed to represent relationships between variables. *Scatterplots* tell us the following:
  + whether the data represent a linear or nonlinear association
  + whether the association is negative or positive
  + whether the data has any association at all
* When a *scatterplot* has a *linear relationship*, a *line of best fit* can be drawn through the data to show the closest linear model of the data.
  + If most of the points are not close to the line, then the *scatterplot* does not represent a *linear relationship*.

|  |  |
| --- | --- |
| **Linear Relationship** |  |
| **Nonlinear Relationship** |  |

**Objective 2:** In this section, you will use data, models, and scatterplots that model nonlinear relationships to analyze situations.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* A *scatterplot* represents a *nonlinear relationship* when the data cannot be modeled with a straight line.

|  |  |
| --- | --- |
| **Nonlinear Relationships** | |
| Quadratic Relationship   * A *scatterplot* with a U-shaped line and contains a maximum or a minimum indicates a *quadratic relationship*. |  |
| Exponential Relationship   * A *scatterplot* that shows a curved line with a rapid increase or decrease models an *exponential relationship*. |  |

|  |  |
| --- | --- |
| **Example:** In the following exponential relationship, approximately how much would you expect a 7-year-old to be worth? | |
| **Step 1:** Locate a data point on the scatterplot that represents a 7-year-old baseball card. | * The *x-*axis represents years and the *y-*axis represents dollars. * Locate the data point with an *x-*value of 7. |
| **Step 2:** Interpret the data point. | Using the point (7,25) we can approximate that a 7-year-old baseball card has a value of about $25. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the image to answer the question.  A scatter plot in first quadrant of coordinate plane. Both axes range from 0 to 12 in increments of 2.  Distinguish what, if any, type of association is represented in the scatterplot. | a negative linear relationship |
| P 2 | Use the image to answer the question.  A scatter plot in first quadrant of coordinate plane. X-axis ranges from 0 to 20 in increments of 2 and y-axis ranges from 0 to 20,000 in increments of 2000.  Distinguish what, if any, type of association is represented in the scatterplot. | No association is represented. |
| P 3 | Use the data sets to answer the question.  Data Set A: (0,0), (1,1), (2,2), (3,3), (4,4), (5,5)  Data Set B: (0,4), (1,6), (2,8), (3,10), (4,12)  Data Set C: (0,1), (3,3), (6,9), (9,27), (12,81)  Data Set D: (10,0), (11,−1), (12,−2), (13,−3), (14,−4)  Create a scatterplot for each of the data sets shown. Distinguish which set of data represents a non-linear relationship. | Data Set C |
| P 4 | Use the image to answer the question.  A scatterplot has 11 closed points plotted on a coordinate plane.  Two balls are thrown from the top of a 35-foot-tall building. Approximately how many seconds did it take the balls to hit the ground? (Round to the nearest whole number.)  It took the balls approximately \_\_\_\_ seconds to reach the ground. | 5 |
| P 5 | Use the image to answer the question.  A scatterplot has 17 closed points plotted on a coordinate plane.  A soccer player is practicing kicking the ball. The coach captures one data point for each time she kicks the ball, noting the height of the ball at some time after it was kicked. The scatterplot shows the data points from 17 trials. Using the data, estimate the amount of time it takes for the soccer ball to return to the ground after it is kicked. Answer with a whole number.  The ball reaches the ground in \_\_\_\_ seconds. | 8 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the image to answer the question.  A graph shows a series of upward slope diagonal plots from origin.  Distinguish what, if any, type of association is represented in the scatterplot. | a positive linear relationship |
| Q 2 | Use the image to answer the question.  Shows 4 scatter plots.  Distinguish which scatterplot represents a negative linear relationship. | Scatterplot D |
| Q 3 | Use the data sets to answer the question.  Data Set A: (0,2), (1,4), (2,8), (3,16), (4,32), (5,64)  Data Set B: (3,4), (4,6), (5,8), (6,10), (7,12)  Data Set C: (0,1), (3,3), (6,9), (9,27), (12,81)  Data Set D: (2,4), (−1,2), (7,16), (−4,32), (1,64)  Create a scatterplot for each of the data sets shown. Which data set can reasonably be modeled by a linear relationship? | Data Set B |
| Q 4 | Use the image to answer the question.  Twenty points are plotted on a scatterplot titled Hours Playing with Toys. The x-axis is labeled Age, and the y-axis is labeled Hours.  What conclusion is plausible from analyzing the scatterplot? | The age group that spends the most amount of time playing with toys is 8- to 10-year-olds. |
| Q 5 | Which of the following scatterplots models an exponential relationship? | A scatterplot has 9 closed points plotted on a coordinate plane. |

**Lesson 5 – Linear Models**

**Key Words:**

* **extrapolate** – to predict by projecting beyond the known values of the explanatory or independent variable
* **interpolate** – to estimate values of data between two known values
* **least squares regression** – a method of fitting a curve to a set of points representing statistical data in such a way that the sum of the squares of the distances of the points from the curve is a minimum
* **line of best fit** – a straight line that minimizes the distance between it and the data, also called the least squares regression line
* **linear model** – an equation that describes the relationship between two quantities that show a constant rate of change
* **overestimate** – to estimate as being more than the actual size, quantity, or number
* **residual** – the difference in measurement between a data point and the least squares regression line
* **scatter plot** – a graph in which the values of two quantitative variables are plotted along two axes, the pattern of the resulting points reveals any correlation present
* **slope-intercept form** – the equation that represents linear relationships (), where *m* is the slope and *b* is the *y*-intercept
* **underestimate** – to estimate as being less than the actual size, quantity, or number

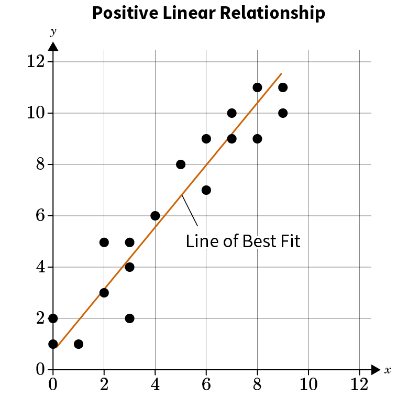
**Formulas:**

* Slope-Intercept Form:

**Objective 1:** In this section, you will use technology to determine the least squares regression line from a given dataset.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* Recall that the *line of best fit* is a straight line that minimizes the distance between it and the data. It is an approximate line used to model the data of a *scatterplot*.
  + 
* The most accurate way to calculate the exact placement for the *line of best fit* is using *least squares regression*.
  + *Least squares regression* determines the best fit for a set of data by squaring the distance each point is from the *line of best fit*.
  + The least squares regression line models a linear function and can be written in slope-intercept form: .
* You can calculate the *least squares regression line* by hand, but the best way is to use technology.

|  |  |
| --- | --- |
| **Least Squares Regression Using** [**GeoGebra**](https://www.geogebra.org/classic#spreadsheet) | |
| **Example:** Determine the least squares regression line from the table.   |  |  | | --- | --- | | ***x*** | ***y*** | | 0 | 1.5 | | 1 | 1 | | 2 | 3 | | 3 | 3 | | 4 | 5 | | 5 | 5.5 | | |
| **Step 1:** Open the [GeoGebra Spreadsheet Calculator](https://www.geogebra.org/classic#spreadsheet) and insert the given table of data into columns A and B. |  |
| **Step 2:** Highlight all values in columns A and B and select “Two Variable Regression Analysis”. |  |
| **Step 3:** A scatterplot will appear. Find the label “Regression Model” and select the option “Linear” in the dropdown menu. |  |
| **Step 4:** A line of best fit will appear on the scatterplot, along with an equation in slope-intercept form.  **This equation is the line of regression.** |  |

**Objective 2:** In this section, you will interpret the slope and *y*-intercept of linear models in the context of the data.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* Recall that the *least squares regression line* is a *linear model* in the form .
  + Recall that the slope, *m*, represents the steepness and direction of a function, or the rate of change.
  + Recall that the *y-*intercept, *b*, represents the point in the graph that intersects the *y-*axis, when *x* is zero.

|  |  |
| --- | --- |
| **Example 1:** Consider the following scatterplot that details the relationship between hours spent studying and grades on an exam. Interpret the slope and *y-*intercept of the least squares regression line.  Least Squares Regression: | |
| **Step 1:** Interpret the slope of the line. | * The slope is positive 6.65 which represents a positive relationship. * For every hour spent studying, a student’s grade will increase by approximately 6.65 points. |
| **Step 2:** Interpret the *y-*intercept. | * The *y-*intercept for this line is 57.88. * When a student spends no time studying, their exam score will be approximately 57.88. |

|  |  |
| --- | --- |
| **Example 2:** Given the following scatterplot that displays the relationship between rainy days and days spent visiting the beach per month, interpret the slope and *y-*intercept from the linear model.  Least Squares Regression: | |
| **Step 1:** Interpret the slope of the line. | * The slope is –0.90, so for every additional rainy day, the number of days spent at the beach will decrease by approximately 0.90. |
| **Step 2:** Interpret the *y-*intercept. | * The *y-*intercept is 31. This means that when a month has zero rainy days, people will spend approximately 31 days at the beach. |

**Objective 3:** In this section, you will use linear models to make predictions.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* Using the graph analysis skills learned in previous lessons, you can:
  + *interpolate* - estimate values of data between two known values
  + *extrapolate* - predict by projecting beyond the known values
  + Determine if a line of best fit is an *overestimate* or *underestimate* for any given data point

|  |  |
| --- | --- |
| **Example:** The probability that a school will win its first few games in the tournament is based on its rank going into the tournament. Use the scatterplot to answer the questions.  Line of Best Fit: | |
| **Interpolate**  What is the probability of a school ranked #6 winning its first game? Round your answer to the nearest percent. | This question is using interpolation because a ranking of #6 is within the known values.   * Method 1: Use the scatterplot to approximate a value.    + When the school is ranked #6, the probability of winning the first game is approximately 67%. * Method 2: Substitute into the line of best fit.   + When the school is ranked #6, the probability of winning the first game is about 67%. |
| **Extrapolate**  What is the probability of a school ranked #13 winning its first game? Round your answer to the nearest percent. | This question uses extrapolation because the data in question is beyond the known values and cannot be estimated using the scatterplot.   * Substitute into the line of best fit equation.   + When the school is ranked #13, the probability of winning the first game is about 22%. |
| **Over/Underestimate**  Does the line of best fit over or underestimate the school ranked #3? | Evaluate the line of best fit for and compare it to the known value from the scatterplot.  Compare this value to a known value using the scatterplot:    When rounding to the nearest percent, the line of best fit is accurate for the school ranked #3. It does not over or underestimate. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the table to answer the question.    Devin creates a table showing five points. He uses a spreadsheet calculator to calculate the slope and *y*-intercept of the least squares regression line based on the points in the table. What is the slope of this line?  The slope of the least squares regression line is \_\_\_\_\_. | 2 |
| P 2 | Use the image to answer the question.  Thirty-eight points are plotted on a scatterplot. The x-axis is labeled Number of Hours Studied, and the y-axis is labeled Test Scores.  Which option could be the slope for this data set and the correct interpretation in this context?  Option #1: The slope is 10, meaning for each hour of study the score goes up ten points.  Option #2: The slope is 10, meaning for each point on the test you must study for ten hours.  Option #3: The slope is , meaning for each hour of study, the score will go up ten points.  Option #4: The slope is , meaning for each test point, ten hours of study are needed.  Option #\_\_\_ is the correct option. | 1 |
| P 3 | Use the image to answer the question.  A line is plotted on a coordinate plane. The x-axis representing months ranges from negative 1 to 12 in unit increments. The y-axis representing total amount paid in dollars ranges from negative 50 to 600 in increments of 50, but marked at every 100.  Claire got a part-time job pet sitting. She received a $150 sign-on bonus and then gets paid each month depending on how many hours she works. The graph shows the line of best fit displaying the number of months she has worked compared to the total amount of money she has earned. Which of the following options is a correct interpretation of the *y*-intercept (0,150)?  Option #1: Claire earns $150 for every job she completes.  Option #2: Claire earns an average of $150 a month.  Option #3: The initial amount of money Claire earns is $150.  Option #\_\_\_\_ is the correct option. | 3 |
| P 4 | The loan amount is a function of time and can be represented by the line of best fit , where *x* is the number of years. How much is left on the loan after 9 years?  The loan amount is $\_\_\_after 9 years. | 698 |
| P 5 | Use the linear model to make a prediction. The line of best fit  represents the tolls you pay after driving a specific number of miles. How much would it cost, in exact change, to travel 70 miles?  It would cost $\_\_\_ to travel 70 miles. | 30.90 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    Use a spreadsheet calculator to determine the equation of the line of best fit for the data in this table. Round to the nearest hundredth, if needed. |  |
| Q 2 | Use the table to answer the question.    Ava is a middle hitter on her club volleyball team. The table displays the number of hitting attempts she had in the last six games compared to the number of kills, which is when a hitting attempt hits the ground and scores a point. If the slope for the line of best fit is , interpret the following representations of the slope and identify which statement is accurate. | Ava has about 1 kill for every 2 hitting attempts. |
| Q 3 | Use the image to answer the question.  Four points are plotted on a coordinate plane, and a line is graphed through them. The x-axis is labeled Days and ranges from 0 to 600 in increments of 20. The y-axis is labeled Weight in Pounds and ranges from 250 to 2,500 in increments of 250.  An elephant calf is born weighing 282 pounds and gains about 2.5 pounds each day for its first year of life. After one year, the calf weighs approximately 1,195 pounds. The graph represents the line of best fit for the data. Which of the following interpretations of the slope is accurate? | The calf gains 5 pounds every 2 days. |
| Q 4 | Use the linear model to make a prediction. The line of best fit  represents the tolls you pay after driving a specific number of miles. What prediction can be made based on this line of best fit? | You can spend $42 to drive 100 miles. |
| Q 5 | Use the table to answer the question.    The weight of a cheetah can be based on the number of miles they run a day. The line of best fit for this dataset is . The table represents the scatter plot of data. Find the weight of a cheetah that runs 7 miles, and state whether extrapolation or interpolation is occurring. | 143.9 pounds and interpolation |

**Lesson 6 – Residuals**

**Key Words:**

* **line of best fit** – a straight line that minimizes the distance between it and the data, also called the least squares regression line
* **linear regression** – the process of finding the line of best fit to model a set of data
* **residual** – the vertical distance between a data point and the line of best fit; the result of subtracting the predicted y-value from the actual y-value
* **residual plot** – the graph of the residuals of data on which the x-axis is the same as the graph of the data and the y-axis is the residuals for the data
* **scatter plot** – a graph in which the values of two quantitative variables are plotted along two axes, the pattern of the resulting points reveals any correlation present

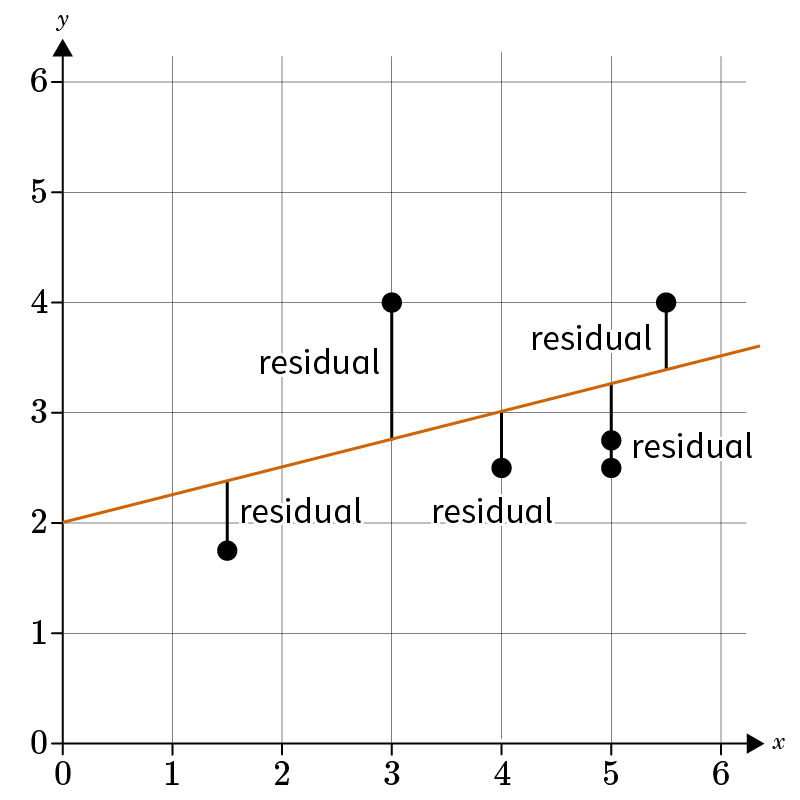
**Formulas:**

* Residual = actual value – predicted value, or

**Objective 1:** In this section, you will graph residuals between data and corresponding linear models.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* A *residual* is the vertical distance between a data point and the *line of best fit*.
  + 
* You can calculate the r*esidual* value by subtracting the predicted *y-*value from the actual *y-*value.
  + *y* represents the actual *y-*value, given in the data or on the scatterplot.
  + represents the predicted *y-*value, calculated using the line of best fit.
  + Residual = actual value (*y*) - predicted value ()

|  |  |
| --- | --- |
| **Example:** A group of meteorologists took measurements from a hurricane over several days as it developed. The variable *x* represents the air pressure in millibars(kPa) and *y* represents the maximum sustained wind speed in knots (nautical miles per hour). Graph the residuals when the line of best fit is . | |
| **Step 1:** Use the line of best fit to find the predicted value for each data point. | Substitute each *x-*value (air pressure) into the line of best fit to get the predicted value. |
| **Step 2:** Find the residual value for each data point by subtracting each predicted value from the corresponding actual value. | |  |  |  |  | | --- | --- | --- | --- | | **Air Pressure (kPa)** | **Wind Speed (knots)** |  |  | | 921 | 131 | 129.43 | 131 − 129.43 = 1.57 | | 945 | 92.3 | 101.35 | 92.3 − 101.35 = −9.05 | | 955 | 89.2 | 89.65 | 892 − 89. 65 = −0.45 | | 967 | 77.7 | 75.61 | 77.7 − 75.61 = 2.09 | | 992 | 40.3 | 46.63 | 40.3 − 4636 = −6. 06 | | 1,002 | 36 | 34.66 | 36 − 34.66 = 1.34 | |
| **Step 3:** Graph the residual values on the residual plot for the data. | Graph each data point . |

**Objective 2:** In this section, you will use residual plots to determine the reliability of linear models and the accuracy of predictions made by using them.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* [Recall](#Bookmark2) how to create a residual plot. You can interpret a *residual plot* to determine whether a linear model is the best fit for the data.

|  |  |
| --- | --- |
| **How to Interpret a Residual Plot** | |
| **Reliable Line of Best Fit**  If the residuals are randomly scattered around the line where residual=0 for the entire data, then the linear model that was used is appropriate. |  |
| **Not Reliable Line of Best Fit**  If there is any sort of pattern to the residuals, then the linear model that was used is **not** appropriate. |  |

|  |  |
| --- | --- |
| **Example:** Determine the reliability of each linear model given the scatterplot. | |
|  | The given residual plot clearly has a pattern. Therefore, a linear model is not reliable and will not provide accurate predictions for the original data. |
|  | The given residual plot is equally and randomly spaced around the line where residual=0. Therefore, the linear regression model is a good choice for this data and will be accurate when making predictions. |

**Objective 3:** In this section, you will connect shapes of scatter plots to shapes of corresponding residual plots.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* To connect a given scatter plot to a *residual plot*, you must analyze the scatter plot and consider the following:
  + The further a data point is from the line of best fit, the further the *residual point* will be from the on the *residual plot*.
  + It is also important to keep in mind that the *x-*values of the data points on a scatter plot will match the *x-*values of the points on the *residual plot*. Only the *y-*values will be different.

|  |  |
| --- | --- |
| **Connecting Scatter Plots and Residual Plots** | |
| Data points that are above the line of best fit will be positive values on the residual plot. |  |
| Data points that are below the line of best fit will be negative values on the residual plot. |  |
| Data points that are on the line of best fit will lie on the on the residual plot. |  |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Find the residual value that would need to be plotted if the *y*-value is 195 and the  is 192.  The plotted residual value is \_\_\_. | 3 |
| P 2 | The line of best fit is . A data point in the scatterplot is (7,90). What is the  value? Enter your answer to the nearest tenth.  The value is\_\_\_\_\_. | 92.9 |
| P 3 | Use the image to answer the question.  Six points are plotted on a scatterplot. The x-axis is labeled Height in feet, and the y-axis is labeled Residual Value.  Determine the reliability of the linear model given the residual plot. Enter the number of the correct option.  Option #1: The given residual plot clearly has a pattern. Therefore, a linear model is not reliable and will not provide accurate predictions.  Option #2: The given residual plot clearly has a pattern. Therefore, a linear model is reliable and will provide accurate predictions for the original data.  Option #3: The residual plot is equally and randomly distributed around the line where residual=0. Therefore, the linear regression model is a good choice for this data and will provide accurate results when making predictions.  Option #\_\_\_\_. | 1 |
| P 4 | Use the image to answer the question.  Six points are plotted on a scatterplot. The x-axis is labeled Air Pressure in kilopascals, and the y-axis is labeled Residual Value.  Determine the reliability of the linear model given the residual plot. Enter the number of the correct option.  Option #1: The residual plot is equally and randomly distributed around the line where residual=0. Therefore, the linear regression model is a good choice for this data and will provide accurate results when making predictions.  Option #2: The given residual plot clearly has a pattern. Therefore, a linear model is not reliable and will not provide accurate predictions for the original data.  Option #3: The given residual plot clearly has a pattern. Therefore, a linear model is reliable and will provide accurate predictions for the original data.  Option #\_\_\_\_\_. | 1 |
| P 5 | Use the image to answer the question.  A coordinate plane for cost versus distance.  How many points would be below the residual=0 line in the residual plot of the function shown?  There would be \_\_\_\_ dots above the residual=0 line. | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    What value point would you graph for residual at a height of 6.08 given the data in the table? | -14.4 |
| Q 2 | The line of best fit is . This line shows the comparison of house size in comparison to the cost of the house. If the size of the house, 1,800 square feet, gives the cost of $250,000, what is the value given the line of best fit? | $244,804 |
| Q 3 | Use the image to answer the question.  Ten points are plotted on a scatterplot. The x-axis is labeled Minutes Played, and the y-axis is labeled Residual Value.  Determine the reliability of the linear model given the residual plot. | The residual plot is equally and randomly distributed around the line where residual=0. Therefore, the linear regression model is a good choice for this data and will provide accurate results when making predictions. |
| Q 4 | Use the image to answer the question.  A coordinate plane for cost versus distance.  Based on the scatterplot, connect the shape of the scatterplot to the accurate description of the shape of a residual plot. Which residual plot description is the most accurate? | There are two dots that will fall on or close to the residual=0 line. |
| Q 5 | Use the image to answer the question.  A line and ten points are plotted in quadrant 1 of a coordinate plane. The x-axis is labeled Minutes Played, and the y-axis is labeled Points Scored.  Based on the scatterplot, which residual plot description would be the best match? | a residual plot with 3 dots below the *x*-axis, 3 dots on the *x*-axis, and 4 dots above the *x*-axis |

**Lesson 7 – Correlation Coefficient**

**Key Words:**

* **coordinate plane** – a two-dimensional plane formed by the intersection of the *x*-axis and the *y*-axis
* **correlation** – the relationship between variables
* **correlation coefficient** – a number (usually represented with the variable *r*) in the range of that indicates how accurately an equation represents the data being examined
* **line of best fit** – a straight line that minimizes the distance between it and the data, also called the least squares regression line
* **slope** – the steepness of a line, found by dividing the change in the *y*-value by the change in the *x*-value

**Formulas:**

* Correlation Coefficient:

**Objective 1:** In this section, you will use technology to determine the correlation coefficient (*r*-value) for a dataset.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* The *correlation coefficient* is a number that indicates how well an equation represents the data.
* The variable ***r*** represents *correlation coefficient* and is always between –1 and +1 ().
* Using GeoGebra, you can calculate the *correlation coefficient* in two ways, depending on your data:

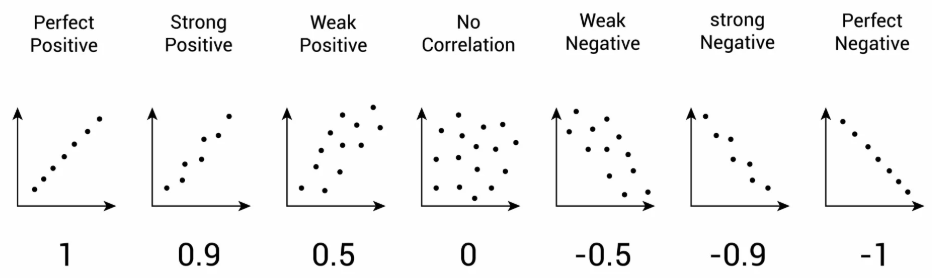
|  |  |
| --- | --- |
| **Correlation Coefficient in GeoGebra: Table Feature** | |
| If you are given and xy table of data, use the “Table” feature in GeoGebra.   * Select the “Table” feature in GeoGebra to enter the *xy* data into the appropriate columns. * Select the three dots next to *y1*. * Select *xy1* Statistics. |  |
| * Several statistics will appear, scroll to find “Correlation Coefficient”, *r*. |  |

|  |  |
| --- | --- |
| **Correlation Coefficient in GeoGebra: Algebra Feature** | |
| If you are given an equation, use the “Algebra” feature in GeoGebra.   * Select the “Algebra” feature in GeoGebra. * Type the equation into the field. * Select the three dots next to the equation. * Select “Table of Values”. |  |
| * A window will pop up asking you to set your start and end *x-*values. You can leave them as is and select OK. |  |
| * A table of values will appear. * Add a third column that is exactly the same as the second column, *f(x)*. This will automatically be named *y1.* |  |
| * Select the three dots next to *y1*. * Select *xy1* Statistics. |  |
| * Several statistics will appear, scroll to find “Correlation Coefficient”, *r*. |  |

**Objective 2:** In this section, you will interpret what the correlation coefficient of a data set means in terms of its strength and direction on the coordinate plane.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* The *correlation coefficient* (*r*) tells us both the **strength** and **direction** of a relationship between two variables.
* Strength:
  + The closer your *r-*value is to **1**, the **stronger the positive** correlation of the data.
  + The closer your *r-*value is to **-1**, the **stronger the negative** correlation of the data.
  + The closer your *r-*value is to **zero**, the **less correlation** there is in the data.
* Direction:
  + If *r* is **negative** (between -1 and 0), the correlation has a **negative slope** (as one variable increases, the other decreases).
  + If *r* is **positive** (between 0 and 1), the correlation has a **positive slope** (both variables increase together).
  + 

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the table to answer the question.    Find the correlation coefficient of the data using technology (such as GeoGebra). Round your r-value to two decimal places, if necessary.  *r* = \_\_\_\_ | 0.99 |
| P 2 | Find the correlation coefficient of the following linear equation using technology (such as GeoGebra). Round your r-value to two decimal places, if necessary.  *r* = \_\_\_\_ | -1 |
| P 3 | Is the correlation coefficient  strong or weak?  Type 1 for strong.  Type 2 for weak. | 2 |
| P 4 | Does the correlation coefficient  show a decrease or increase from left to right on a coordinate plane with the line of best fit?  Type 1 for decrease.  Type 2 for increase. | 2 |
| P 5 | Which of the following correlation coefficients is the strongest:  or ?  Type 1 for .  Type 2 for . | 1 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    Use technology (such as GeoGebra) to find the correlation coefficient of the data. Round your r-value to two decimal places, if necessary. | 0.93 |
| Q 2 | Find the correlation coefficient of the following linear equation using technology (such as GeoGebra). For your r-value, use the first two decimal places. | 1.00 |
| Q 3 | Interpret the following correlation coefficients and list them in order from weakest to strongest.   * −0.89234651 * 0.72643969 * −0.99872348 * 0.69846237 | 0.69846237, 0.72643969, −0.89234651, and −0.99872348 |
| Q 4 | What is the correlation coefficient of the following data? Is it a strong or weak correlation? | 0.97300135; strong |
| Q 5 | Use the image to answer the question.  An unlabeled coordinate plane has a line with an arrow that begins at the point of origin and moves diagonally upward from left to right. Twenty-five closed points are plotted along the entire length of the line, above, below, and on it.   Which of the following statements best describes the data illustrated in the graph? | The graph illustrates a strong positive correlation between the data. |

**Lesson 8 – Correlation Between Variables**

**Key Words:**

* **correlation coefficient** – a number (usually represented with the variable *r*) in the range of that indicates how accurately an equation represents the data being examined
* **line of best fit** – a straight line that minimizes the distance between it and the data, also called the least squares regression line
* **residual** – the difference between the *y*-value for a point on a scatterplot and the value predicted/estimated by a linear model (line of best fit)
* **scatterplot** – a two-dimensional plane with coordinate points that display values for two variables of a dataset
* **slope** – the steepness of a line, found by dividing the change in the y-value by the change in the *x*-value

**Objective 1:** In this section, you will estimate the correlation coefficients (*r*-values) of scatterplots in a variety of forms.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* [Recall](#Bookmark3) what you have learned about the *r-*value, or *correlation coefficient.* This information can be used to estimate the *r-*value from a scatterplot.
* The farther the data points lie outside of the line of best fit (spread out), the closer the estimate of the *r-*value is to zero.
  + The closer the r-value is to zero, the weaker the correlation.
* The closer the data points are to the line of best fit (tighter fit), the closer the estimate of the *r-*value is to 1 or –1.
  + The closer the *r-*value is to 1 or –1, the stronger the correlation.

|  |  |
| --- | --- |
| **Example:** What is the best estimated *r-*value of the following scatterplot? 0.2, -0.5, -0.9, or 0.8? | |
| **Step 1:** Identify the direction of the correlation. | * The scatterplot has an upward, positive direction. * The correlation coefficient will be a positive number. * This eliminates the options of –0.5 and –0.9.   0.2, ~~-0.5~~, ~~-0.9~~, 0.8 |
| **Step 2:** Identify the strength of the correlation. | * The data points are tighter together and closer to the line of best fit. * This indicates a stronger correlation. * The correlation coefficient 0.2 represents a weak positive correlation, so this cannot be the correlation coefficient of this data.   ~~0.2~~, ~~-0.5~~, ~~-0.9~~, 0.8 |

**Objective 2:** In this section, you will pair residual analysis and the correlation coefficient (*r*-value) of a dataset to determine whether the linear model (line of best fit) is appropriate.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* Recall how to calculate residuals.
  + [residuals](#Bookmark4) - The residual values indicate how far the data points are from the line of best fit.
* Recall how to analyze a dataset using GeoGebra and how they relate to residuals.
  + [line of best fit](#Bookmark5) - The line of best fit should have data points both above and below to most accurately represent the data if the correlation coefficient does not equal 1 or –1.
  + [correlation coefficient](#Bookmark6) - Positive and negative correlation coefficients should yield both positive and negative residual values.

|  |  |
| --- | --- |
| **Example:** You decide to examine a small dataset comparing the number of hours you have spent practicing the piano (*x-*values and input) with the number of notes missed while playing the song for your upcoming piano recital (*y-*values and output). | |
| **Step 1:** Use GeoGebra to create a scatterplot. |  |
| **Step 2:** Use GeoGebra to [calculate the line of best fit](#Bookmark5). |  |
| **Step 3:** Use GeoGebra to [calculate the correlation coefficient, *r*](#Bookmark6). |  |
| **Step 4:** [Calculate the residuals](#Bookmark4). | Based on *r*, the correlation coefficient, you can predict that each of the residuals will be small—a size indicating that each data point sits very close to the linear model. |
| **Step 5:** Analyze the residuals and determine if the linear model accurately represents the data. | * The table indicates that three of the residual values are positive.   + This positive value means that they sit above the line of best fit. * The table also reveals that three of the residual values are negative.   + This negative value means that they sit below the line of best fit. * A line of best fit (linear model) should be the best possible representation of the data with some points above and some points below that line. |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Use the image to answer the question.  Thirteen randomly spaced points are plotted in a diagonal pattern in the first quadrant of a coordinate plane with x and y axes. The first point starts almost at the origin and then the plots show an increasing trend.  True or false, the scatter plot has an estimated correlation coefficient of −1.  Type 1 for true.  Type 2 for false. | 2 |
| P 2 | Use the image to answer the question.  A scatterplot shows eight randomly spaced points on the first quadrant of a coordinate plane with x and y axes. The points start on a higher y-axis value and decrease as the x-axis value increases.  Which is a better r-value estimate for the scatter plot, −0.8 or −1?  Type 1 for −0.8.  Type 2 for −1. | 1 |
| P 3 | For an activity in class, a team of students was given the line of best fit, . If one of the actual data points is (1,16.88), what is the residual of the data point? Round the answer to the nearest hundredths.  The residual of the data point (1,16.88) is \_\_\_\_. | -0.22 |
| P 4 | Use the image to answer the question.  Four points are plotted on the first quadrant of a coordinate plane. The x-axis ranges from 0 to 10 in increments of 1 and the y-axis ranges from 0 to 17 in increments of 1. A line is drawn that almost passes through all the points.  Find the residual of the data point (4,14.59). Round the answer to the nearest thousandths.  The residual of the data point (4,14.59) is \_\_\_\_. | 0.146 |
| P 5 | A team of students collected data on the growth of a plant. They plotted the height of the plant every day and found the line of best fit  for the growth of the plant. If the actual data point for day six is (6,7.82), what is the residual of the data point for day six? Round the answer to the nearest hundredths.  The residual of the data point (6,7.82) is \_\_\_\_\_. | -3.46 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Use the table to answer the question.    Use the data table to create a scatter plot. Estimate the correlation coefficient based on the data. | 1 |
| Q 2 | Use the image to answer the question.  Eleven points are plotted in a diagonal pattern in the first quadrant of a coordinate plane with x and y axes. The first point starts almost at the origin and shows an increasing trend.  Based on the data, which of the following choices would be the best estimated correlation coefficient? | 0.9 |
| Q 3 | Use the image to answer the question.  Twelve points are plotted in a scattered increasing curved pattern in the first quadrant of a coordinate plane with x and y axes. The x-axis is labeled as Time Spent Studying and y-axis as Scores on a Test.  Based on the data, which of the following choices would be the best estimated correlation coefficient? | 0.98 |
| Q 4 | Jes used GeoGebra to find a linear model (trend line) of the data collected for their science project. They calculated the following residuals for their data 0.26, 0.93, 0.5489, and 0.659. What conclusion can Jes make about the appropriateness of the linear model based on the residuals? | The linear model does not accurately represent the data since all residuals are positive. |
| Q 5 | Jade was working on her statistics homework. One of the questions gave her a trend line and asked her to find the residuals. She found the following residuals 2.6, 9.3,−5.489, and −6.59. What conclusion can Jade make about the linear model based on the residuals? | The linear model seems to accurately represent the data since half of the residuals are positive and half of the residuals are negative. |

**Lesson 11 – Analyzing a Dataset**

**Key Words:**

* **exponential function** – a function in the form  where *x* is a variable and *a* and *b* are real numbers, and
* **linear function** – a function in the form where *x* is a variable and *m* and *b* are real numbers
* **quadratic function** – a function in the form where *x* is a variable; *a*, *b*,and *c* are real numbers; and

**Formulas:**

* Exponential Function:
* Linear Function:
* Quadratic Function:

**Objective 1:** In this section, you will recognize various datasets that represent linear functions, quadratic functions, and exponential functions.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas**:

* You can identify the differences between *linear*, *quadratic*, and *exponential functions* based on their models/equations.
* Data can be represented in the form of a graph, a table, or an equation.

|  |  |
| --- | --- |
| **Linear, Quadratic, and Exponential Functions from a Table** | |
| **Linear Functions:**   * Linear functions have data points that follow a straight path with a positive or negative slope. | * To verify that a table of values represents a linear function you can calculate the first differences by subtracting consecutive *y-*values. * If the first differences are constant, it is linear.   Example: |
| **Quadratic Functions:**   * Quadratic functions have data points that create a U-shaped curve called a parabola. | * To verify that a table of values represents a quadratic you can calculate the second differences by subtracting consecutive first differences. * If the second differences are constant, it is exponential.   Example: |
| **Exponential Functions:**   * Exponential functions show growth or decay.   + Growth starts flat and quickly increases.   + Decay starts high and quickly declines. | * To verify that a table of values represents an exponential function you can find the common ratio between the *y-*values. * To find the common ratio, divide the *y-*value by the previous *y-*value. * If the ratio is constant, it is exponential.   Example: |

**Objective 2:** In this section, you will create models of datasets that represent linear functions, quadratic functions, and exponential functions.

*Mathematical Practice Standard: Model with mathematics.*

**Big Ideas:**

* When given a table of data, you can use GeoGebra to plot the graph and find the function that best models the data.

|  |  |
| --- | --- |
| **GeoGebra: Create a Function Model for a Dataset** | |
| **Step 1:** Use the “Table” feature in the GeoGebra calculator.   * Enter the *x* and *y* values into the table. * Select the gear icon in the upper right. * Select Zoom to Fit. |  |
| **Step 2:** The table and graph will now be displayed side-by-side. |  |
| **Step 3:** Use regression to find the equation of the function that best fits the dataset.   * Select the three dots next to the y1 header. * Select Regression. * From the dropdown menu, select the type of regression, linear, quadratic, or exponential based on the shape of the graph. |  |
| **Step 4:** [Based on the shape of the data](#Bookmark7), select the type of model from the dropdown menu. (Linear, Quadratic, or Exponential)   * Substitute the values of the parameters into the Formula to find the equation. | Formula:  Parameters:  Equation: |

**Practice Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| P 1 | Which dataset represents a quadratic function? |  |
| P 2 | Which type of function **best** models the data in the table?  **Time and Distance Data** | exponential |
| P 3 | Use the table to answer the question.    Which algebraic model best represents the dataset on the *xy*-chart?  Option #1: . This is a quadratic function.  Option #2:  This is a linear function.  Option #3: . This is an exponential function.  Option #\_\_\_\_ is the correct response. | 1 |
| P 4 | Which function best models the dataset in the table?    Option #1:  Option #2:  Option #3:  Option #\_\_\_\_ is the correct response. | 2 |
| P 5 | Which exponential function best models the data in the table?    Option #1:  Option #2:  Option #3: | 2 |

**Quick Check Questions and Answers**

|  |  |  |
| --- | --- | --- |
|  | Question | Answer |
| Q 1 | Which of the following **best** models the data in the table?  **Type of Function** | quadratic |
| Q 2 | Which type of function **best** models the data in the table?  **Time and Growth Data** | quadratic |
| Q 3 | Use the table to answer the question.    Create a quadratic model for the dataset in the *xy*-chart. |  |
| Q 4 | Use the table to answer the question.    Select the function that best models the data and determine the value of the function at . |  |
| Q 5 | Use the table to answer the question.    Create an exponential model of the dataset shown in the *xy*-chart. |  |