LESSON 4: Residuals

[Objective]
The student will use residuals to predict values based on a regression line and draw conclusions about the appropriate use of regression equations.

[Prerequisite Skills]
Scatter plots, line regression, correlation coefficient, coefficient of determination, regression equations

[Materials]
Student pages **S1–S19**

[Essential Questions]
1. What are residuals?
2. If residuals are all zero, then what do we know about the best-fit line?
3. How can we use residuals to analyze how good a best-fit line represents a set of data?

[Words For Word Wall]
Scatter plots, regression, correlation coefficient, residuals, coefficient of determination, line of best fit, quadratic regression, linear regression, minimum value, maximum value, scale, line of regression

[Grouping]
Cooperative Pairs (CP), Whole Group (WG), Individual (I)
*For Cooperative Pairs (CP) activities, each student should be responsible for designated tasks within the lesson.

[Levels of Teacher Support]
Modeling (M), Guided Practice (GP), Independent Practice (IP)

[Multiple Representations]
SOLVE, Verbal Description, Pictorial Representation, Graphic Organizer, Algebraic Formula, Graph, Table

[Warm-Up] (IP, WG) S1 (Answers on T14.)
Have students turn to S1 in their books to begin the Warm-Up. Students will answer questions about a scatter plot, line of best fit and correlation coefficient. Monitor students to see if any of them need help during the Warm-Up. After students have completed the warm-up, review the solutions as a group. **{Algebraic Formula, Table}**

[Homework]
Take time to go over the homework from the previous night.

[lesson] [2 - 3 Days (1 day – 80 minutes) – M, GP, WG, CP, IP]

SOLVE Problem (WG, GP) S2 (Answers on T15.)

Have students turn to S2 in their books. The first problem is a SOLVE problem. You are only going to complete the S step with students at this point. Tell students that during the lesson they will learn how to use residuals to determine the appropriateness of a regression equation for a specific situation. They will use this knowledge to complete this SOLVE problem at the end of the lesson. **{SOLVE, Verbal Description, Graphic Organizer, Table}**

****Teacher Note: Students will need to understand and use both linear and quadratic regression equations in this lesson. Teachers may want to offer students an opportunity to review and practice the specific types of regression equations with their graphing calculators before beginning the lesson.**

Identifying Residuals (M, GP, WG, CP, IP) S2, S3, S4, S5 (Answers on T15, T16, T17, T18.)

M, GP, WG, CP, IP: Have students turn to S2 in their books. Students will be applying what they know about scatter plots, line of best fit, and correlation coefficient to identify the residuals. **{Graph, Pictorial Representation, Verbal Description, Graphic Organizer, Table, Algebraic Formula}**

MODELING

Identifying Residuals

Step 1: Have students review the data given in the table on S2 for Example 1.

- What is the relationship that we are examining? (We are looking at the number of police cars on a given highway and the average speed of motorists on that highway.)
- Have students make a prediction about the relationship between the two values. (Students may have varying answers, but one can make a prediction that the more police cars on a highway, the more cautious people will be about their speed.)

Step 2: Direct students to Question 1.

- What do we need to do when creating a **scatter plot**? (Determine the appropriate **scale** and label the axes.)
- Which value will be the x-value on the graph? (Number of police cars.)
- Which value will be the y-value on the graph? (Average speed of motorists.)
- What would be an appropriate scale for the x-axis? Explain your thinking. (An appropriate scale would be units of 1. Our **minimum value** is 1 and our **maximum value** is 9 on the x-axis.)
- What would be an appropriate scale for the y-axis? Explain your thinking. (An appropriate scale would be by units of 5 starting with 50. Our minimum value is 55 and our maximum value on the y-axis is 75.)

Step 3: Have students plot the points from the table on the coordinate plane and then turn to S3.

- Does there appear to be a linear relationship? (Yes)
- Find the **line of best fit** with your graphing calculator. What is the line of best fit? (y = $^-$2.45 + 74.69) Record.
- Draw the line of best fit on the scatter plot on S2.
- What is the correlation coefficient? (r = $^-$0.96) Record.
- Explain what this means in terms of the data. [Because the correlation coefficient is ($^-$0.96), x and y have a (strong negative linear correlation.)] Record.

Step 4: Use the line of best fit from Question 2. Determine the predicted speed if there were 8 patrol cars on the highway. (y = 55.09) Record.

- This is the value we determined using the line of best fit, called the (predicted value). Record.
- Have students look at Question 4. What is the data given in our table when there were 8 patrol cars on the highway? [The average speed was (57).] Record.
- This is the actual data from the table. It is called the (actual value). Record.

Step 5: Have students look at the chart in Question 5.

- What do you notice about this chart (It has a third row called "Predicted Value".)
- How can we use the information from the line of best fit to determine the predicted values for the chart? (Substitute in the x-value which is the number of police cars on a given highway into the line of best fit to find the predicted value.)
- Have students work with a partner to complete the chart and review the answers as a whole group.

Step 6: Have students turn to page S4 and look at Question 6.

- This is the line of best fit of our data. Draw a dotted line from each point vertically up or down to the line of best fit.
- What do you notice about the points? (Some of the points are above the line and some are below the line.)

Step 7: Have students look at Question 7 on the bottom of S4.

- What is the label on the additional row on the chart on S4? (Actual Data Value – Predicted Value.)
- How can we find the difference between the two values? (Subtract 68 – 67.34)
- What is the difference in the two values? (0.66) Record.

- Is the value positive or negative? Explain your thinking. (The value is positive because the actual value is greater than the predicted value. The actual value is above the line of best fit.)
- What is the difference between the second actual value and the second predicted value? (2.76) Record.
- Is the value positive or negative? (positive)
- What is the difference between the next two values? ($^{-}2.89$) Is the value positive or negative? (negative) Record the sign. Explain your thinking. (The actual value is less than the predicted value.)
- Have students complete the chart on the bottom of page S4 and review their answers as a whole group.

Step 8: Have students turn to page S5 and graph the values from the bottom row of the chart on S4 for Question 8.

- Do you notice any pattern with the data values that you plotted on the graph? (No) Record.
- The data values that we plotted on the graph in Question 8 are known as **residuals**.

****Teacher Note: Students will be making observations about each of the examples during the lesson. On page S14, students will compile the information from each of the six example problems and look for patterns and draw conclusions.**

Step 9: Have students look at the box below Question 9 that contains the information about residuals.

- How can we define a residual? The (distance) between a data point and its predicted value on the line of best fit. Record.
- How do we determine if the residual is negative? The residual is negative if the point is (below) the line of best fit. Record.
- How can we determine if the residual is positive? The residual is positive if the point is (above) the line of best fit. Record.
- In what case would the residual be zero? The residual is 0 if the point is (on) the line of best fit. Record.
- How do we represent the residual on a graph? The (vertical) distances on the examples below represent residuals. Record

Step 10: Have students look at the Observation Chart on the bottom of S5.

- What is the **correlation coefficient**? ($r = {}^{-}0.96$) Record.
- What does this mean? (The line of best fit is a good predictor of the data. Record.
- Is there a pattern with the residual graph? (No) Record.
- The (correlation coefficient) measures how well our predication data matches with the actual model. Although this is an excellent match, it does not tell us whether the model is (appropriate) only whether it is a good (prediction). Record.

Plotting a Residual Graph Using a Linear Regression on the Graphing Calculator

(M, GP, WG) S6 (Answers on T19.)

M, GP, WG: Have students turn to S6 in their books. Students will plot a residual graph with the graphing calculator using a linear regression. **{Graph, Pictorial Representation, Verbal Description, Table, Graphic Organizer, Algebraic Formula}**

****Teacher Note: On the graphing calculator, the leading zero is not shown. However, it is included in the values in the lesson so that students will not be confused or omit a decimal point.**

MODELING

Plotting a Residual Graph Using a Linear Regression on the Graphing Calculator

Step 1: Have students answer the following questions.

- Another name for a line of best fit is a (**regression**) equation. Record.
- We can find other types of (regressions) based on the data. Record.
- Types of regression equations include: (linear, quadratic or exponential) Record.
- Have students look at the data table for Example 2 on S6. They will be using this data to complete the activity on the graphing calculator.
- What is the relationship that we are examining? (We are looking at the data of the relationship between age and height.)

Step 2: Model steps 1 – 6 for the students on S6.

- Make sure your Diagnostic is On.
- Store your data in lists. STAT and 1: Edit Enter the *x*-values (age) in L1 and the *y*-values (height) in L2.
- Perform a linear regression. STAT, arrow right to CALC, choose 4:LinReg (*ax*+*b*)
- Now go back to STAT and 1: Edit. Arrow over to the top of L3. Press LIST (2nd STAT) and choose #7 RESID. Press ENTER.
- Go to the STAT PLOT menu (2nd *Y*=). Choose Plot1, hit ENTER. Turn the Plot ON (by hitting ENTER). Arrow down to Type: choose the first option (scatter plot). Make *X* List: L1 (the *x*-values) Make Y List: by hitting LIST 7 (the residuals)
- Go to ZOOM. Choose ZOOM 9:ZoomStat

Step 3: At this point, you should see your Residual Plot. Graph the residual plot on the coordinate grid. (You will also see the residuals in your L3 Column.)

- What is the **linear regression** equation? ($y = 1.87x + 31.7$) Record.
- What is your r? (0.96) Record.
- By looking at the r do we have a good line? (Yes) Record
- Explain how you know the line is a good regression equation. (The value of the r is close to 1.)
- Does your residual plot make a pattern? (No) Record.

Interpreting Residuals (M, GP, WG) S7, S8 (Answers on T20, T21.)

M, GP, WG: Have students turn to S7 in their books. Students will be applying what they know about scatter plots, line of best fit, correlation coefficient to identify and interpret the residual. **{Graph, Pictorial Representation, Verbal Description, Table, Graphic Organizer, Algebraic Formula}**

MODELING

Interpreting Residuals

Step 1: Have students read the real world problem for Example 3 on page S7.

- What is the relationship that we are examining? (We are looking at the data from an experiment with a play car and the relationship between the time and distance traveled.)
- Have students make a prediction about the relationship between the two values. (Students may have varying answers, but one can make a prediction that the more time a play car traveled, the longer distance it traveled.)

Step 2: Have students plot the points from the example on the coordinate plane in Question 1.

- Which value will be the x-value on the graph? (Time traveled in seconds) Record.
- Which value will be the y-value on the graph? (Distance traveled in feet) Record.
- What do we need to do when creating a scatter plot? (Determine the appropriate scale and label the axes.)
- What would be an appropriate scale for the x-axis? Explain your thinking. (An appropriate scale would be by units of 0.5. Our minimum value is 0 and our maximum value on the x-axis is 4 with increments of 0.5)
- What would be an appropriate scale for the y-axis? Explain your thinking. (An appropriate scale would be by units of 5, starting with 0. Our minimum value is 0 and our maximum value on the y-axis is 22.8.)

Step 3: Does there appear to be a linear relationship? (Yes)

- Find the line regression with your graphing calculator. What is the line of best fit? ($y = 5.64x - 3.3$) Record.
- What is the correlation coefficient? ($r = 0.96$) Record.
- Explain what this means in terms of the data. [The correlation coefficient is (0.96). Because the correction coefficient is (0.96), x and y have a (strong positive linear correlation.)] Record.
- Draw the line of best fit on the scatter plot on S7.

Step 4: Have students turn to page S8. Model the steps for determining the residuals.

- Make sure your Diagnostic is On.
- Store your data in lists. STAT and 1: Edit Enter the *x*-values (age) in L1 and the *y*-values (height) in L2.
- Perform a linear regression. STAT, arrow right to CALC, choose 4:LinReg (*ax*+*b*)
- Now go back to STAT and 1: Edit. Arrow over to the top of L3. Press LIST (2nd STAT) and choose #7 RESID. Press ENTER.
- Go to the STAT PLOT menu (2nd *Y*=). Choose Plot1, hit ENTER.

  Turn the Plot ON (by hitting ENTER). Arrow down to Type: choose the first option (scatter plot). Make *X* List: L1 (the *x*-values) Make Y List: by hitting LIST 7 (the residuals)
- Go to ZOOM. Choose ZOOM 9:ZoomStat

Step 5: Use the information on your graphing calculator to record the Residuals in the table above Problem 1 on S7.

- Have students look at Question 3 on S2. What does the residual represent? (It is the difference between the predicted value and the observed value.) Record.

Step 6: Have students turn to page S8 and graph the residuals from the chart on S7 for Question 4.

- Discuss the observations in the chart below the graph.
- What is the correlation coefficient? (0.96) Record.
- What does this mean? (The line of best fit is a good predictor of the data.) Record.
- Do you notice any sort of pattern with your residual graph? (Yes) Record.
- Describe the pattern. (The data points form a curve.) Record.

Step 7: Based on the pattern in the residual graph at the top of the page, what type of regression is modeled by the data points? **(quadratic regression)** Record.

- If students do not recognize that this is a quadratic equation you may need to go back review the graph of a quadratic equation.
- Explain your answer. (The data pattern is in the form of a curve that models a quadratic relationship.) Record.
- Another word for the line of best fit is called a **line of regression**

****Teachers: Remember that at this point, we are only making and charting observations that we will use for our conclusions after all the examples.**

Using a Regression Equation **(M, GP, WG) S9, S10, S11, S12 (Answers on T22, T23, T24, T25.)**

M, GP, WG: Have students turn to S9 in their books. Students will determine the appropriate use of a regression equation using real world data and comparing the model from Example 2. **{Graph, Pictorial Representation, Verbal Description, Graphic Organizer, Algebraic Formula, Table}**

MODELING

Using a Regression Equation

Step 1: Have students look at the data table on S9. This is the same data table from Example 3.

- What is the relationship that we are examining? (We are looking at the data from an experiment with a play car and the relationship between the time and distance traveled.)

Step 2: Have students plot the points from the example on the coordinate grid in Question 1.

- Which value will be the x-value on the graph? (Time traveled in seconds) Record.
- Which value will be the y-value on the graph? (Distance traveled in feet) Record.
- What do we need to do when creating a scatter plot? (Determine the appropriate scale and label the axes.)
- What would be an appropriate scale for the x-axis? Explain your thinking. (An appropriate scale would be by units of 0.5. Our minimum value is 0 and our maximum value on the x-axis is 4 with increments of 0.5)
- What would be an appropriate scale for the y-axis? Explain your thinking. (An appropriate scale would be by units of 5, starting with 0. Our minimum value is 0 and our maximum value on the y-axis is 22.8.)

Step 3: Have students plot the points from the table in Example 4 on the grid.

- In the previous example problem, what relationship did we identify? (linear)
- Have students go back and look at the graph on S7.
- Explain the relationship between the line of best fit and the data points on page S7. (The line was not going through many points.)
- Have students determine the regression equation using the quadratic function using the graphing calculator.
- What is the regression equation? ($y = 1.46x^2 - 0.21x + 0.12$) Record.
- Draw the line on the coordinate grid in Question 1.
- How does this regression look compared to the linear graph? (The curved line is going through more of the points.) Record.

Step 4: Have students use the graphing calculator to determine the residuals and record in the table on the top of S9.

Step 5: Have students turn to page S10 and graph the residuals on the graph.

- Have students direct their attention to the Observation Chart on the bottom of S10.
- Do you notice any sort of pattern with your residual graph? (No) Record.
- Have students read Question 4. What is the r^2 of your quadratic regression equation from your graphing calculator? (0.9997) Record.

Step 6: What is the meaning of r^2? The (**coefficient of determination**) is another measure of how well the best fit line performs as a predictor of y. Record.

- What does r^2 mean? (r^2 takes on values between 0 and 1) Record.
- What is the meaning of a value of 1 for r^2? (A value of 1, indicates a perfect fit and a good predictor of future data.) Record.
- What is the meaning of a value of 0 for r^2? (A value of 0 indicates that the model fails to accurately model the data set.) Record.

Step 7: Have students turn to page S11 and work with a partner to complete the linear regression with the graphing calculator.

- Write the residuals in the table and graph them on the graph.
- Work as a group to complete the questions in the Observation Chart below the graph.
- What is the linear regression equation? (y = 6.04x + 13.07) Record.
- What is your r^2 value? (0.88) Record.
- What is your r value? (0.94) record.
- By looking only at these two values, would you conclude that this line is a good predictor of future data? (Yes) Record.
- Explain your thinking. (Both values are close to 1) Record.
- Does your residual plot make a pattern? (Yes) Record.

Step 8: Have students turn to page S12.

- Have students look at the chart in Example 6. This is the same data from Example 5. This time, we will graph this as a quadratic regression to compare the linear and quadratic graphs.
- Have students use their graphing calculators to determine the quadratic regression, and include this information in the Observation Chart below the graph.
- Have students determine the residual values in the chart from the graphing calculator and plot them on the coordinate graph.

Step 9: Have students finish the questions in the Observation Chart.

- What is the quadratic regression equation? ($y = {}^{-}0.63x^2 + 13.61x + .45$) Record.
- What is your r^2? (0.998) Record.
- What is your r? (It did not give us one) Record.
- By looking at the r^2 and r do we have a good line? (Yes, even though it did not have an r.) Record
- Does your residual plot make a pattern? (No) Record.

Drawing Conclusions about Patterns with Regression Equations and Coefficient of Determination

(M, GP, WG, CP, IP) S13, S14, S15 (Answers on T26, T27, T28.)

M, GP, WG: Have students turn to S13 in their books. Students will look at the equations and patterns between the six example problems the have completed. **{Graph, Pictorial Representation, Verbal Description, Graphic Organizer, Algebraic Formula, Table}**

MODELING

Drawing Conclusions about Patterns with Regression Equations and Coefficient of Determination

Step 1: Have students look at the chart on S13. Students will use the information from the six example problems and situations.

Step 2: Let's look at Example 1.

- What is the regression equation for Example 1? ($y = {}^-2.45x + 74.69$) Record.
- What is the correlation coefficient? ($r = {}^-0.96$) Record.
- Use the graphing calculator to determine the Coefficient of Determination
  ($r^2 = 0.91$) Record.
- Look at the graph of the residuals for Example 1. Is there a pattern? (No)

Step 3: Use the questions from Step 2 and Examples 2 – 6 to complete the chart.

- Review the answers as a whole group.
- Have students work with a partner or in small groups to write down some conclusions about linear and quadratic regression equations.

Step 4: Have students turn to page S14. Work with the students to complete the Conclusions Chart.

- Which examples showed no pattern with the graphed residuals? (Examples 1, 2, 4 and 6) Record.
- Let's Look at Examples 1 and 2.
- What do you notice about the correlation coefficient in each of those examples? (The correlation coefficient or "*r*" is very close to 1 or negative 1.) Record.
- What type of line of best fit or regression equation was used to model those situations? (Linear model) Record.
- Did these examples show a pattern when the residuals were graphed? (No) Record.

Step 5: Let's look at Examples 3 and 5.

- Did these examples show a pattern when the residuals were graphed? (Yes) Record.
- What did we have you do after we saw that they had a pattern? (tried a quadratic regression) Record.

Step 6: Let's look at Examples 4 and 6.

- Did these examples show a pattern when the residuals were graphed? (No) Record.
- On the graphing calculator quadratic equations do not give us a "*r*" correlation of coefficient but do give us a r^2 (coefficient of determination)
- What can you say about how close r^2 is to 1. (The coefficient of determination (r^2) was very close to 1.

Step 7: Let's compare the coefficient of determination for Examples 3 and 4 (time and distance) and Examples 5 and 6 (time and speed)

- Example 3 uses a linear regression and Example 4 uses a quadratic regression. What can we say about the r^2 and about if there was a pattern when graphing the patterns? (The r^2 is closer to one in the quadratic regression and there was no pattern when graphing the residuals.) Record.
- Example 5 uses a linear regression and Example 6 uses a quadratic regression.

  What can we say about the r^2 and about if there was a pattern when graphing the patterns? (The r^2 is closer to one in the quadratic regression and there was no pattern when graphing the residuals.) Record.

Step 8: Why do you think we asked you to look at another regression equation for Examples 3 and 5? (The residuals formed a pattern, a curve.) Record.

- What can you conclude from those examples? (If the residuals show a pattern with a linear model, then a quadratic model is the better regression model. You can also compare the r^2 coefficient of determination and see if that is higher on the quadratic regression model.) Record.

Step 9: Have students turn to page S15.

- Have students look at Question 1.
- We do not want residuals to fall long a curve or make a distinct (pattern). Record. If so, then it is likely that a (linear model) is not appropriate to fit the data and perhaps an (exponential or quadratic model) is better. Record.
- Have students sketch an example of a good fit and a bad fit on the grids below Question 1.

CP, IP, WG: Have students complete Questions 2 and 3 on the bottom of S15. Review the answers as a whole group. **{Graph, Pictorial Representation, Verbal Description, Graphic Organizer, Algebraic Formula, Table}**

SOLVE Problem

(WG, GP) S16 (Answers on T29.)

Remind students that the SOLVE problem is the same one from the beginning of the lesson. Complete the SOLVE problem with your students. Ask them for possible connections from the SOLVE problem to the lesson. (Students learn how to assess the fit of the line for residuals and graph the line.) **{SOLVE, Verbal Description, Graphic Organizer, Table, Graph}**

If time permits...

S17 (Answers on T30.)

Have students create the scatter plots and create residual plots.

[Closure]

To wrap up the lesson, go back to the essential questions and discuss them with students.

- What are residuals? *(A residual is the distance between a data point and its predicted value on the line of best fit.)*
- If residuals are all zero, then what do we know about the best-fit line? *(The regression is an accurate fit.)*
- How can we use residuals to analyze how good a best-fit line represents a set of data? *(If the graph of the residuals forms a pattern, then we know it is not a good regression equation.)*

[Homework] Assign S18-S19 for homework. (Answers on T31-T32.)

[Quiz Answers] T33 – T36

1. **C** 2. **D** 3. **C** 4. **B** 5. **A** 6. **B** 7. **B** 8. **A** 9. **D** 10. **D**

The quiz can be used at any time as extra homework or to see how students progress with the use of residuals.

Here is the key to **S1**.

Warm–Up

Directions: Answer each of the follow questions.

The table shows the cost of visiting a working ranch for one day and night for different numbers of people.

1. Make a scatter plot of the data in the table below using your graphing calculator.

Number of people (x)	5	7	9	11	13
Cost in dollars (y)	200	340	410	490	560

2. What is the line of best fit?

$y = 43.5x + 8.5$

3. Find the correlation coefficient, r.

$r = 0.99$

4. Is this a good line of best fit?

It is a good line because the r of 0.99 is close to 1.

LESSON 4: Residuals

Here is the key to **S2**.

Directions: Complete the following SOLVE problem with your teacher. You will only complete the S Step.

The table shows the number of women (in millions) in the U.S. work force at various times during the past century. Is the line of best fit, $y = 0.552x - 1052.3$, a good model for this data?

Year, x	1900	1920	1930	1950	1970	1990
Number, y	5	8	10	16	31	57

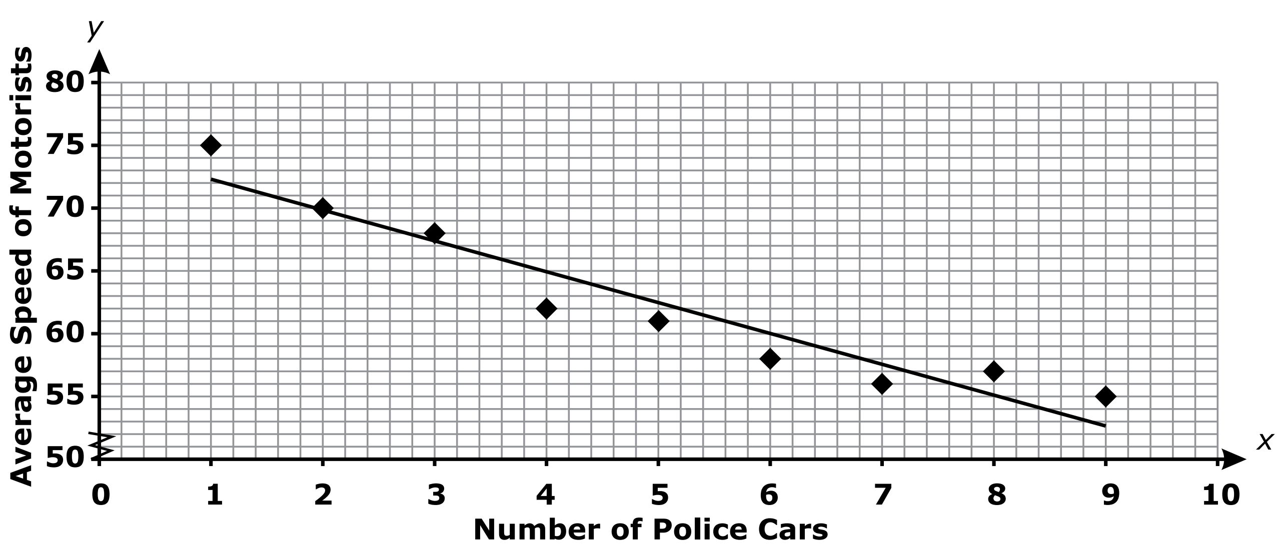
S Underline the question.
This problem is asking me to find **if the equation given is appropriate for this data.**

Directions: Complete this page with your teacher and partner.

Example 1:

Number of Police Cars on a Given Highway	3	1	4	5	7	2	8	6	9
Average Speed of Motorists	68	75	62	61	56	70	57	58	55

1. Draw a scatter plot of the data on the grid provided. Be sure to label and scale axes appropriately.



Here is the key to **S3**.

Directions: Complete this page with your teacher and partner.

Example 1 continued

2. Find the line of best fit and the correlation coefficient using your graphing calculator. Draw your line of best fit in the scatter plot on page S2. (Round to the nearest thousandth.

$y =$ $\mathbf{{}^{-}2.45}\boldsymbol{x}\mathbf{ + 74.69}$ $r =$ $\mathbf{{}^{-}0.96}$

What does this mean in terms of the predicting our data?

The correlation coefficient (r)is $\mathbf{{}^{-}0.96}$ and x and y have a **strong negative linear correlation**.

3. Using the line of best fit, what would be the predicted speed if there were 8 patrol cars on the highway? Show your work.

$\boldsymbol{y} \mathbf{= {}^{-}2.45}\boldsymbol{x} \mathbf{+ 74.69}$

$\boldsymbol{y} \mathbf{= {}^{-}2.45(8) + 74.69}$

$\boldsymbol{y} \mathbf{= 55.09}$

This is our **predicted value**.

4. From our given data what was the average speed given that there are 8 patrol cars out? **57**

This is our **actual value**.

5. Let's calculate the predicted value for each relationship in the chart.

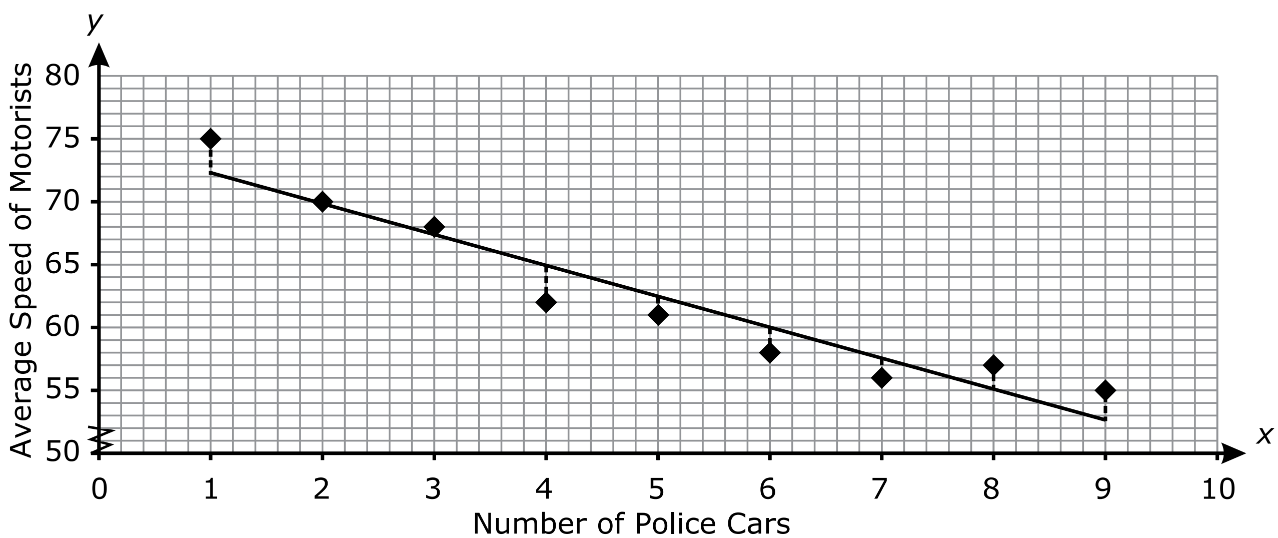
Number of Police Cars on a Given Highway	3	1	4	5	7	2	8	6	9
Average Speed of Motorists	68	75	62	61	56	70	57	58	55
Predicted Value	**67.34**	**72.24**	**64.89**	**62.44**	**57.54**	**69.79**	**55.09**	**59.99**	**52.64**

Here is the key to **S4**.

Directions: Complete this page with your teacher and partner.

Example 1 continued

6. On the graph below. Connect the line of best fit to each point with a vertical line. What do you notice? **Some of the points are above and some of the points are below the line.**



7. What is the difference between our actual value and predicted values for each point?

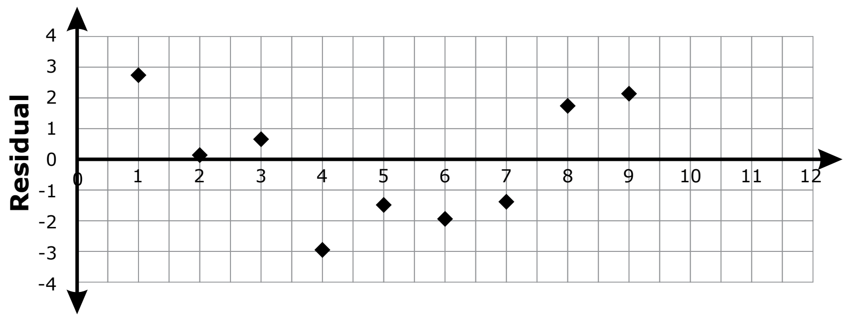
Number of Police Cars on a Given Highway	3	1	4	5	7	2	8	6	9
Average Speed of Motorists	68	75	62	61	56	70	57	58	55
Predicted Value	**67.34**	**72.24**	**64.89**	**62.44**	**57.54**	**69.79**	**55.09**	**59.99**	**52.64**
Actual Data Value – Predicted Value	**0.66**	**2.76**	**¯2.89**	**¯1.44**	**¯1.54**	**0.21**	**1.91**	**¯1.99**	**2.36**

Here is the key to **S5**.

Directions: Complete this page with your teacher and partner.

Example 1 continued

8. Graph the data points from the fourth row of the graph on page S4. These data points represent the difference between the actual data value and the predicted value.

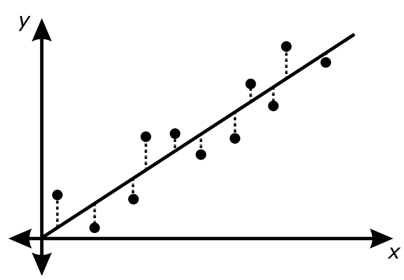


9. Do you notice any pattern with the data values that you plotted on the graph? **No**

The data values that we plotted on the graph in Question 8 are known as **residuals**.

Residual: The **distance** between a data point and its predicted value on the line of best fit.

- The residual is negative if the point is **below** the line of best fit.
- The residual is positive if the point is **above** the line of best fit.
- The residual is 0 if the point is **on** the line of best fit.
- The **vertical** distances on the example below represent residuals.



Observations	
10. What is the correlation coefficient? (Refer back to Question 2 on S3.)	***r* = ⁻0.96**
11. What does this mean?	**The line of best fit is a good predictor of the data.**
12. Is there a pattern with the residual graph?	**No**

The **correlation coefficient** measures how well our prediction data matches with the model. Although this is an excellent match, it does not tell us whether the model is **appropriate**, only whether it is a good **prediction**.

LESSON 4: Residuals

Here is the key to **S6**.

Directions: Complete this page with your teacher and partner.

- Another name for a line of best fit is a **regression** equation.
- We can find other types of **regressions** based on the data.
- Types of regression equations can include: **linear, quadratic** or **exponential**.

Example 2: Given the data below, follow the instructions on how to plot a residual graph in your calculator using a linear regression.

Age	6	8	10	11	13	15	17	20
Height (in.)	41	45	51	52	60	63	64	65

1. Make sure your Diagnostic is On.

2. Store your data in lists. STAT and 1: Edit Enter the x-values (age) in L1 and the y-values (height) in L2.

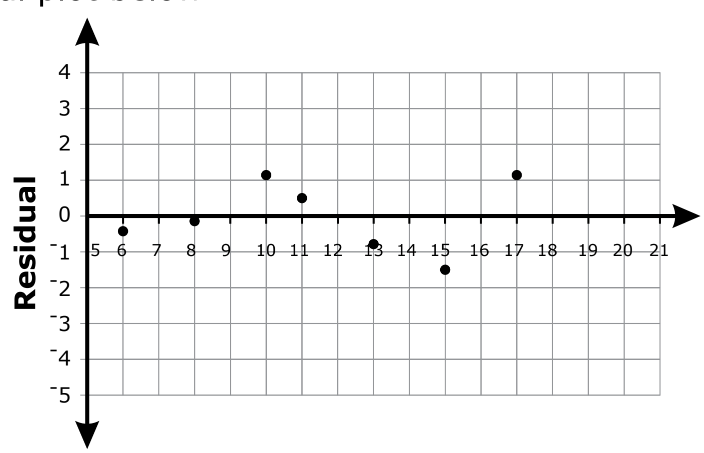
3. Perform a linear regression. STAT, arrow right to CALC, choose either 4:LinReg ($ax+b$)

4. Now go back STAT and 1: Edit. Arrow over to the top of L3. Press (2nd STAT) LIST and choose 7: RESID. Press ENTER.

5. Go to the STAT PLOT menu (2nd $Y=$). Choose Plot1, hit ENTER. Turn the Plot ON (by hitting ENTER). Arrow down to Type: choose the first option (scatter plot). Make X List: L1 (the x-values) Make the Y List: L3, by hitting LIST 7 (the residuals)

6. Go to ZOOM. Choose ZOOM 9:ZoomStat You should see your Residual Plot.

Graph your residual plot below.



6. What is your linear regression equation?	$\boldsymbol{y = 1.87x + 31.7}$
7. What is your r?	**0.96**
8. Just by looking r do you have a good line?	**Yes**
9. Explain how you know the line is a good regression equation.	**The *r* value is close to 1.**
10. Does your residual plot make a pattern?	**No**

Here is the key to **S7**.

Directions: Complete this page with your teacher and partner.

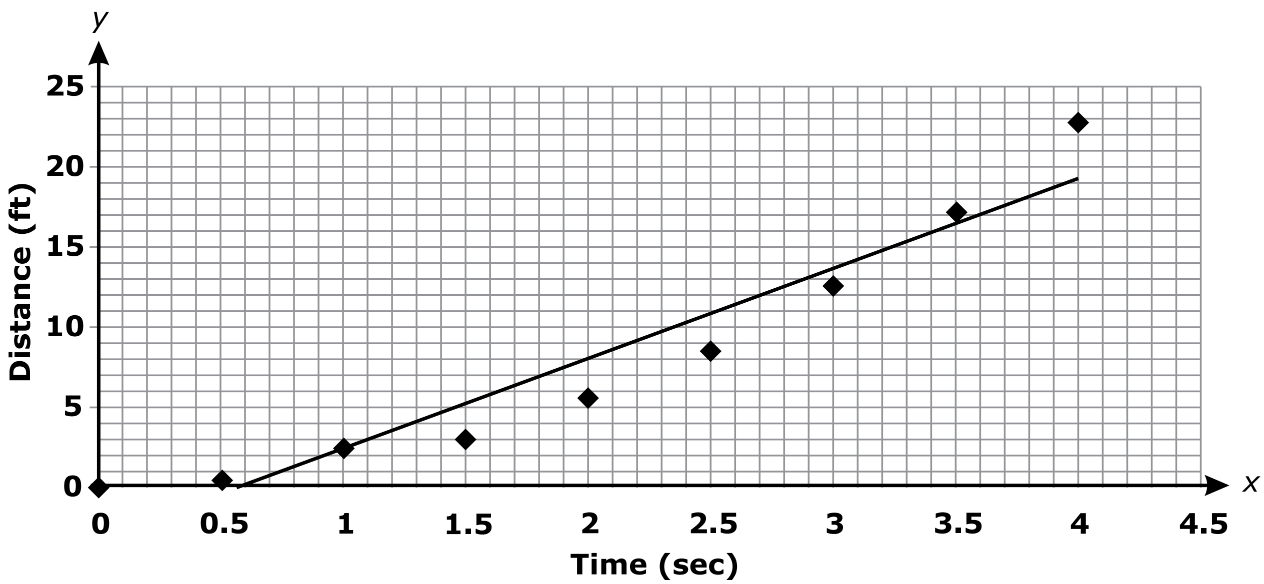
Example 3

In this example, we will look at residuals and residual graphs to determine if a linear model is appropriate.

A group of students in a stem class is performing an experiment where they allow a play car to travel down a ramp and record the distance that it has gone versus the time it has been moving. The data for one such experiment are shown below.

Time (sec)	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Distance (ft)	0	0.4	1.5	3.2	5.6	8.5	12.6	17.2	22.8
Residuals	**3.3**	**0.08**	**$^{-}$0.84**	**$^{-}$1.96**	**$^{-}$2.38**	**$^{-}$2.3**	**$^{-}$1.02**	**0.76**	**3.54**

1. Draw a scatter plot of the data on the grid provided. Label and scale the axes.



2. Find the line of best fit and the correlation coefficient using your graphing calculator. Draw your line of best fit in the scatter plot above.

Line of best fit: y = **5.64*x* – 3.3** Correlation coefficient: r = **0.96**

What does this mean in terms of the predicting our data?

The correlation coefficient is **0.96**. Because the correlation coefficient is **0.96**, x and y have a **strong positive linear** correlation.

3. What does the residual represent? **It is the difference between the predicted value and the observed value.**

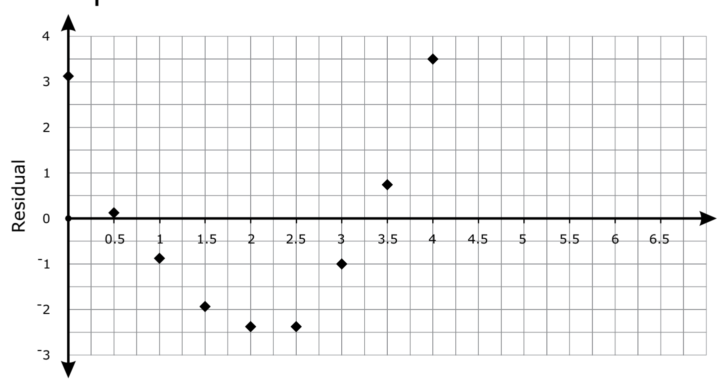
Here is the key to **S8**.

Directions: Complete this page with your teacher and partner.

Example 3 continued

4. Now let's graph our residuals on a separate graph using the time on the *x*-axis and the residual on the *y*-axis.

- Make sure your Diagnostic is On.
- Store your data in lists. STAT and 1: Edit Enter the *x*-values (age) in L1 and the *y*-values (height) in L2.
- Perform a linear regression. STAT, arrow right to CALC, choose either 4:LinReg (*ax*+*b*)
- Now go back STAT and 1: Edit. Arrow over to the top of L3. Press (2nd STAT) LIST and choose 7: RESID. Press ENTER.
- Go to the STAT PLOT menu (2nd *Y*=). Choose Plot1, hit ENTER. Turn the Plot ON (by hitting ENTER). Arrow down to Type: choose the first option (scatter plot). Make *X* List: L1 (the *x*-values) Make the *Y* List: L3, by hitting LIST 7 (the residuals)
- Go to ZOOM. Choose ZOOM 9:ZoomStat You should see your Residual Plot. Graph your residual plot below.



Observations	
5. What is the correlation coefficient?	**0.96**
6. What does this mean?	**The line of best fit is a good predictor of the data**
7. Is there a pattern with the residual graph?	**Yes**
8. Explain the pattern	**The data points form a curve**

Based on the pattern in the residual graph above, what type of relationship is modeled by the data points? **Quadratic** Explain your answer. **The data pattern is in the form of a curve which models a quadratic relationship.** Another word for the line of best fit is called a **line of regression.**

Here is the key to **S9**.

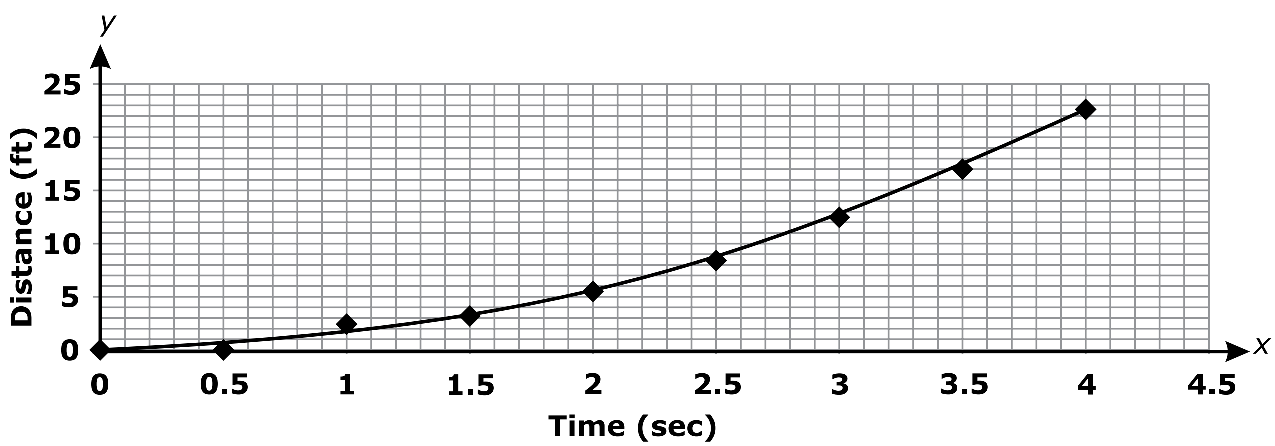
Directions: Complete this page with your teacher and partner.

Example 4

Let's look at the same data and determine if there is a more appropriate regression equation.

Time (sec)	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Distance (ft)	0	0.4	1.5	3.2	5.6	8.5	12.6	17.2	22.8
Residuals	**⁻0.12**	**0.02**	**0.13**	**0.11**	**0.06**	**⁻0.22**	**⁻0.03**	**⁻0.07**	**0.16**

1. Draw a scatter plot of the data on the grid provided. Be sure to label and scale axes appropriately.



2. Find the regression equation using the quadratic function on your graphing calculator. Instead of hitting 4:LinReg hit 5:QuadReg.

y = $\mathbf{1.46x^2 - 0.21x + 0.12}$

Draw your regression equation in the scatter plot above.

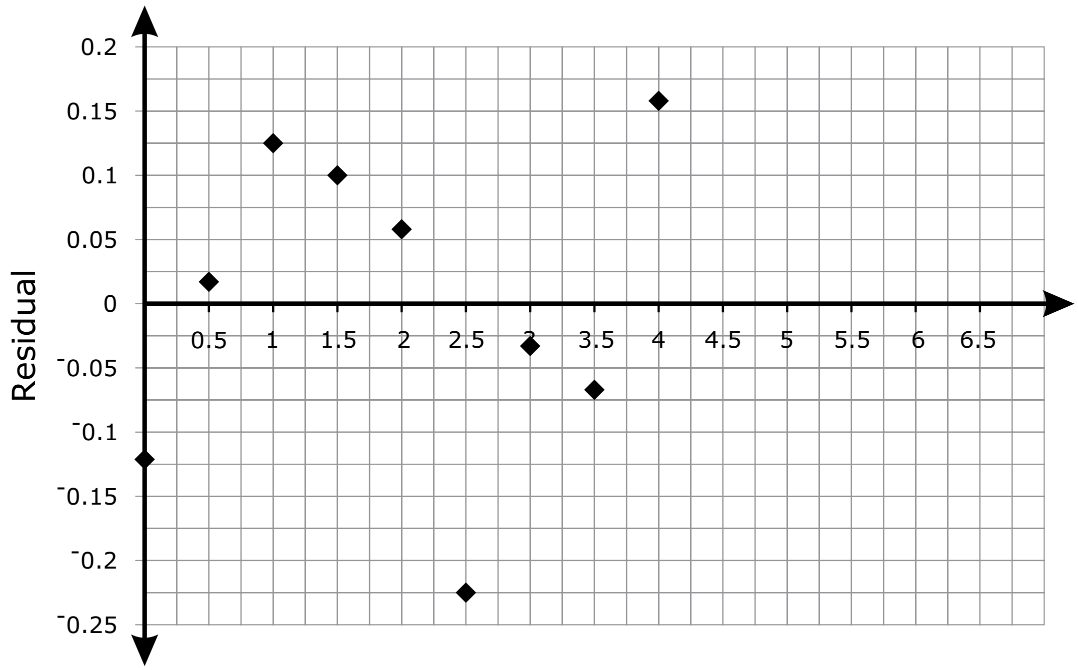
How does this regression look compared to the linear graph on page S7? **The curved line is going through more of the points.**

Here is the key to **S10**.

Directions: Complete this page with your teacher and partner.

Example 4 continued

Graph the residuals from the chart on S10 on the graph below.



Observations	
3. Is there a pattern with the residual graph?	**No**
4. Use the graphing calculator to determine the value of r^2 using the quadratic function from page S9.	$\mathbf{r^2 = 0.9997}$
5. What does r^2 mean?	The **coefficient of determination** is another measure of how well the best fit line performs as a predictor of *y*.
6. What is the range of values for r^2?	**r^2 takes on values between 0 and 1.**
7. What is the meaning of a value of 1 for r^2?	**1 indicates a perfect fit and a good predictor of future data.**
8. What is the meaning of a value of 0 for r^2?	**0 indicates that the model fails to accurately model the data set.**

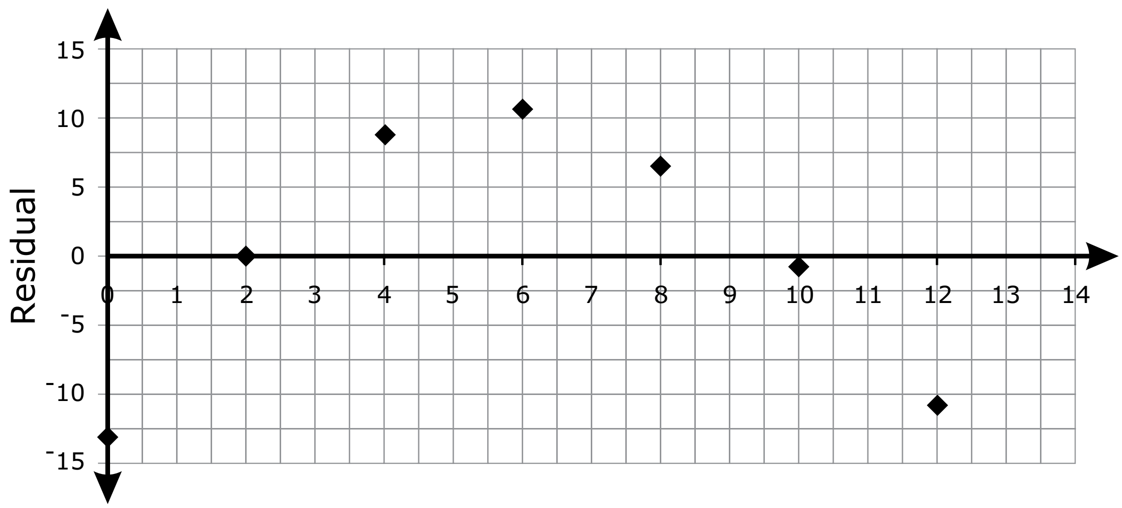
Here is the key to **S11**.

Directions: Complete this page with your partner.

Example 5

Use the data below to find a linear regression and plot the residuals in the graph with your graphing calculator.

Time (sec)	0	2	4	6	8	10	12
Speed (ft/sec)	0	25	46	60	68	72	74
Residuals	**¯13.07**	**¯0.14**	**8.78**	**10.71**	**6.64**	**¯1.43**	**¯11.5**



Use the graphing calculator to answer the following questions.

Observations	
1. What is your linear regression equation?	$\boldsymbol{y = 6.04x + 13.07}$
2. What is your r^2?	**0.88**
3. What is your r?	**0.94**
4. Just by looking at the r^2 and r do you have a good line?	**Yes**
5. Explain your thinking.	**Both values are close to 1.**
6. Does your residual plot make a pattern?	**Yes**

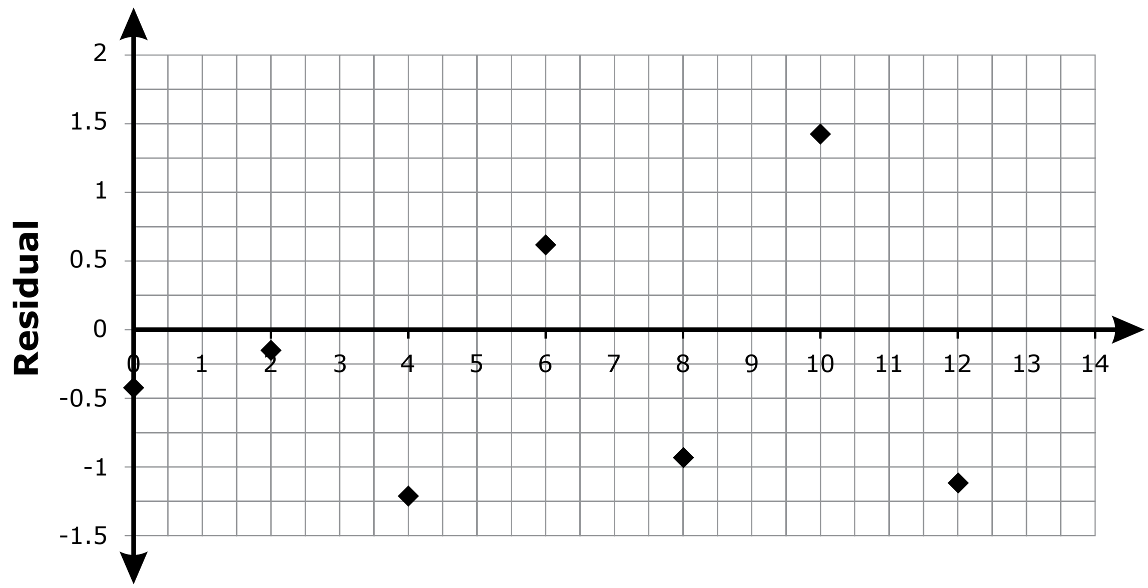
Here is the key to **S12**.

Directions: Complete this page with your teacher and partner.

Example 6

Now let's try the same real-world situation as a quadratic regression and plot the residuals on the graph.

Time (sec)	0	2	4	6	8	10	12
Speed (ft/sec)	0	25	46	60	68	72	74
Residual	**--0.45**	**-0.14**	**-1.21**	**0.62**	**-0.93**	**1.43**	**-1.12**



Use the graphing calculator to answer the following questions.

Observations	
1. What is your quadratic regression equation?	**$y = {}^-0.63x^2+13.61x+ 0.45$**
2. What is your r^2?	**0.998**
3. What is your r?	**It did not give us one.**
4. Just by looking at the r^2 and r do you have a good line?	**Yes, even though it did not give a *r*.**
5. Does your residual plot make a pattern?	**No**

LESSON 4: Residuals

Here is the key to **S13**.

Directions: Complete this page with your teacher and partner.

Problem	Situation	Regression Equation	Correlation Coefficient	Coefficient of Determination	Pattern when the Residuals were graphed?
Example 1	Police Cars and Speed	$y = {}^{-}2.45x + 74.69$	$r = {}^{-}0.96$	$r^2 = 0.91$	**No**
Example 2	Number of Years Old and Height	$y = 1.87x + 31.7$	$r = 0.96$	$r^2 = 0.91$	**No**
Example 3	Time and Distance (Linear)	$y = 5.64x - 3.3$	$r = 0.96$	$r^2 = 0.920$	**Yes**
Example 4	Time and Distance (Quadratic)	$y = 1.46x^2 - 0.21x + 0.12$		$r^2 = 0.9997$	**No**
Example 5	Time and Speed (Linear)	$y = 6.04x + 13.07$	$r = {}^{-}0.95$	$r^2 = 0.88$	**Yes**
Example 6	Time and Speed (Quadratic)	$y = {}^{-}0.63x^2 + 13.61x + 0.45$		$r^2 = 0.998$	**No**

Directions: Before you move on, use the space below to write down some conclusions about linear and quadratic regression equations. (Answers will vary.)

Here is the key to **S14**.

Directions: Complete this page with your partner.

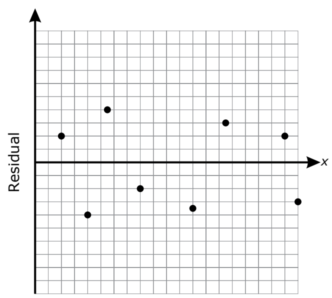
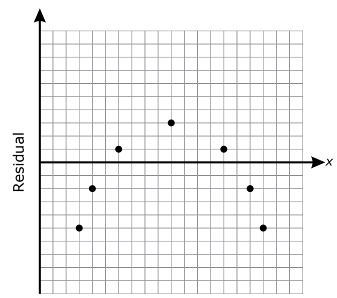
Conclusions:	
1. Which examples showed no pattern with the graphed residuals?	**Examples 1, 2, 4, and 6**
Let's look at Examples 1 and 2.	
2. What do you notice about the correlation coefficient (*r*) for Examples 1 and 2?	**The correlation coefficient (*r*) is very close to 1 or negative 1 for Example 1 and 2.**
3. What type of regression equation was used to model those situations?	**Linear model**
4. Did these examples show a pattern when the residuals were graphed?	**No**
Let's look at Examples 3 and 5	
5. Did these examples show a pattern when the residuals were graphed?	**Yes**
6. What did we have you do after we saw that they had a pattern?	**Tried a quadratic regression.**
Let's look at Examples 4 and 6.	
7. Did these examples show a pattern when the residuals were graphed?	**No**
8. Quadratic Equations do not give us a r (correlation of coefficient) but does give us a r^2 (coefficient of determination). What can you say about how close the r^2 is to 1.	**The coefficient of determination (r^2) was very close to 1.**
Comparing the Coefficient of Determination for Examples 3 and 4 (Time and Distance) Examples 5 and 6 (Time and Speed)	
9. Example 3 is using a Linear Regression Example 4 is using a Quadratic Regression What can you say about the r^2 and about if there was a pattern when graphing the patterns?	**The r^2 is closer to one in the quadratic regression and there was no pattern when graphing the residuals.**
10. Example 5 is using a Linear Regression Example 6 is using a Quadratic Regression What can you say about the r^2 and about if there was a pattern when graphing the patterns?	**The r^2 is closer to one in the quadratic regression and there was no pattern when graphing the residuals.**
11. Why do you think we asked you to look at another regression equation for Examples 3 and 5?	**The residuals formed a pattern, a curve.**
12. What can you conclude from those examples?	**If the residuals show a pattern with a linear model, then a quadratic model is the better regression model. You can also compare the r^2 coefficient of determination and see if that is higher.**

LESSON 4: Residuals

Here is the key to **S15**.

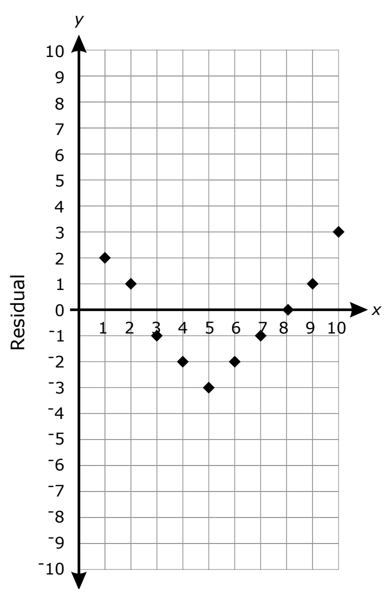
Directions: Complete this page with your partner.

1. We do not want residuals to fall along a curve or make a distinct **pattern.** If so, then it is likely that a **linear model** is not appropriate to fit the data and perhaps an **exponential or quadratic** model is better. Sketch a graph of a good fit and a bad fit on the grids below. (Answers will vary for sketches. Suggested examples are shown.)

Plot is a GOOD Fit	Plot is a BAD Fit
	

2. The table below represents the residuals for a line of best fit. Plot these residuals on the set of axes below.

x	1	2	3	4	5	6	7	8	9	10
Residual	2	1	$^-1$	$^-2$	$^-3$	$^-2$	$^-1$	0	1	3



3. Using the plot, assess the fit of the line for these residuals and justify your answer. **The residuals form a pattern, therefore the line of best fit was not appropriate.**

Here is the key to **S16**.

Directions: Complete this page with your teacher and partner.

Year, x	1900	1920	1930	1950	1970	1990
Number, y	5	8	10	16	31	57

The table shows the number of women (in millions) in the U.S. work force at various times during the past century. | <u>Is the line of best fit, $y = 0.552x - 1052.3$, a good model for this data?</u>|

S Underline the question.

This problem is asking me to find **if the equation given is appropriate for this data.**

O Identify the facts**.**

Eliminate the unnecessary facts.

List the necessary facts. **Information in table, $y = 0.552x - 1052.3$**

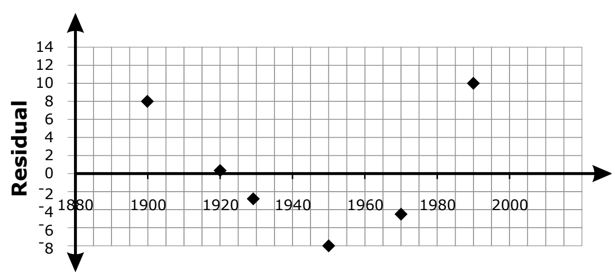
L Write in words what your plan of action will be. **Use the line of best fit to find the residual, the actual value minus the predicted value.**

Choose an operation or operations. **Multiplication, subtraction**

V Estimate your answer. **Equation could be appropriate; they are all increasing**

Carry out your plan.

Year, x	1900	1920	1930	1950	1970	1990
Number, y	5	8	10	16	31	57
Residual	**8.5**	**0.26**	**$^-$3.06**	**$^-$8.1**	**$^-$4.14**	**10.82**



This is not a good fit because the residuals form a pattern.

E Does your answer make sense? (Compare your answer to the question.) **Yes, because I checked the line of best fit compared to the data**

Is your answer reasonable? (Compare your answer to the estimate.) **Yes, because I estimated based on what I think the scatter plot would look like.**

Is your answer accurate? (Check your work.) **Yes.**

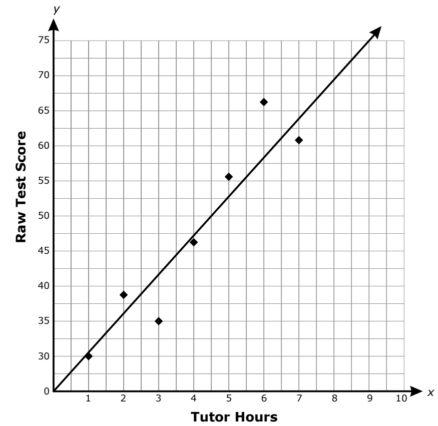
Write your answer in a complete sentence. **The line of best fit was not a good fit because when I graphed the residuals it formed a pattern.**

Here is the key to **S17**.

Directions: Complete this page with your partner.

1. Create a scatterplot of the data set below. Put Tutor Hours on the *x*-axis and Raw Test Score on the *y*-axis. Sketch a best-fit line.

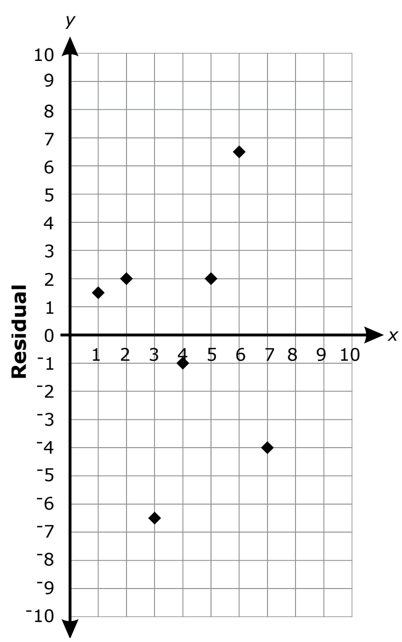
Tutor Hours, *x*	Raw Test Score	Residual (Actual – Predicted)
1	30	1.3
2	37	1.9
3	35	$^{-}6.4$
4	47	$^{-}0.7$
5	56	2.0
6	67	6.6
7	62	$^{-}4.7$



2. Use the data below to write the regression equation ($y = ax + b$) for the raw test score based on the hours tutored. Round all values to the nearest hundredth.

Equation: $\boldsymbol{y = 6.3x + 22.43}$

3. What is the value of the correlation coefficient? Please round to the *nearest thousandth*. Is the equation from part B a good fit? Why or why not? ***r* = 0.91** **Yes, it is a good fit since it is close to 1.**



Given the residual plot that you just made, is the regression line that we found in the warm-up a good fit to represent this table of data? Why or why not?
Yes, it is a good regression line because the residuals do not form a pattern.

Here is the key to **S18**.

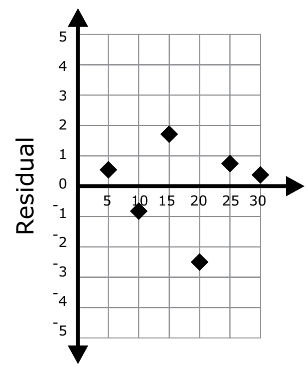
Homework

Name ______________________ **Date** ______________________

Complete each table by finding a linear regression and round your answer to one decimal place. Then construct a residual plot in each case on the graph to the right.

1-2.

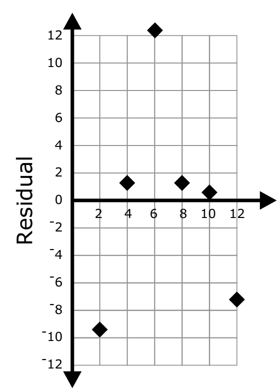
x	*y*	Residual Value
5	3	**0.6**
10	4	**$^{-}$0.9**
15	9	**1.7**
20	7	**$^{-}$2.7**
25	13	**0.9**
30	15	**0.4**



Does the residual plot suggest a linear relationship? Explain. **Yes, because no pattern was formed.**

3-4.

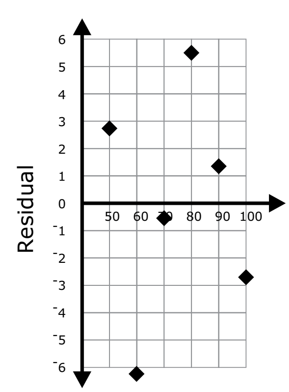
x	*y*	Residual Value
2	5	**$^{-}$9.7**
4	15	**1.3**
6	26	**13.3**
8	13	**1.3**
10	11	**0.3**
12	3	**$^{-}$6.7**



Does the residual plot suggest a linear relationship? Explain. **No, because it forms a pattern.**

5-6.

x	*y*	Residual Value
100	505	**$^{-}$2.9**
90	460	**1.3**
80	415	**5.4**
70	360	**$^{-}$0.4**
60	305	**$^{-}$6.3**
50	265	**2.9**



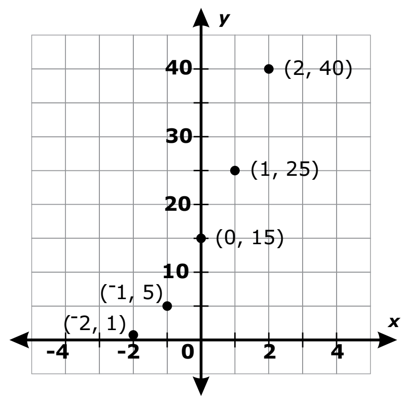
Does the residual plot suggest a linear relationship? Explain. **Yes, because no pattern was formed.**

Here is the key to **S19**.

Homework

Name ______________________ **Date** ______________________

$y = 9.8x + 17.2x$ is a possible regression model for the data set graphed on the scatterplot below.



7-9. Complete the table. Round to the nearest hundredth.

x	y observed	y predicted	residual
$^{-}2$	**1**	**$^{-}2.4$**	**3.4**
$^{-}1$	**5**	**7.4**	**$^{-}2.4$**
0	**15**	**7.4**	**$^{-}2.2$**
1	**25**	**17.2**	**$^{-}2$**
2	**40**	**27**	**3.2**

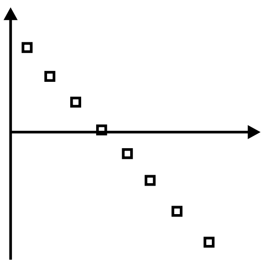
10. Describe how well the model fits the data based on the residuals.

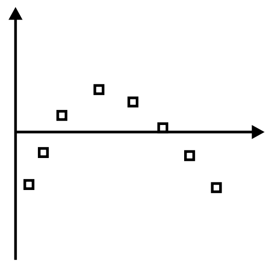
It is not a good fit because it will form a quadratic (curve) pattern.

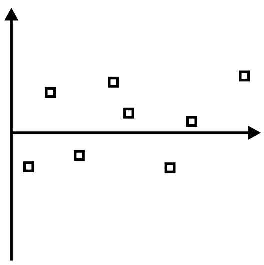
Name ____________________________________ Date ________________

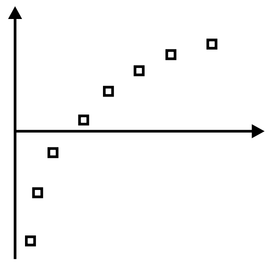
Quiz

1. Which of the following residual plots indicates a model that is most appropriate?

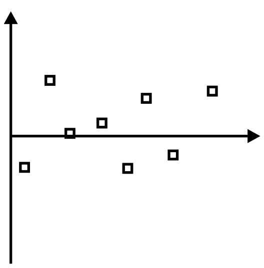
A.


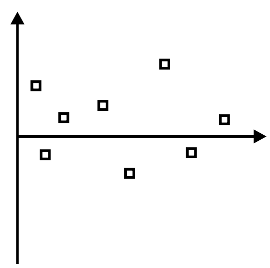
B.


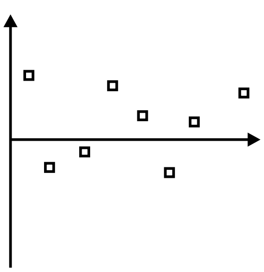
C.


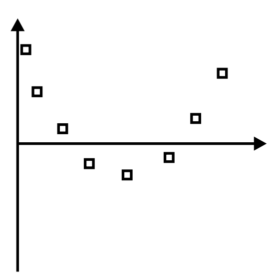
D.


2. Which of the following residual plots would indicate the linear model used to produce it was an inappropriate choice?

A.


B.


C.


D.


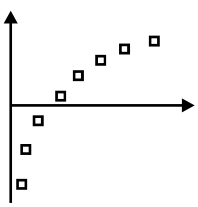
LESSON 4: Residuals

3. A set of data is fit with linear regression. The equation for the line of best fit was $y = 2.5x + 18$. If the observed value when $x = 10$ was $y = 35$, then which of the following represents the value of the residual for this data point?

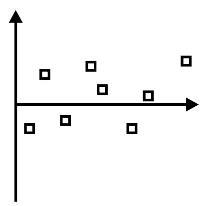
A. 52 B. 8 C. $^{-}8$ D. -10

4. Four different types of regression were used on a data set. Which of the following four residual plots shows the model that is the most appropriate to use?

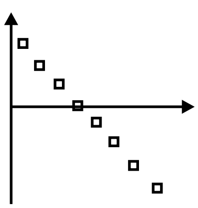
A.



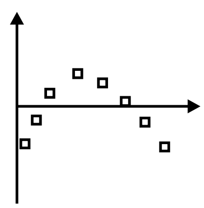
B.



C.



D.



5. Economists create a linear regression equation to predict the price of a gallon of gasoline, y, based on the price of a barrel of oil, x. The equation they find is $y = 0.035x + 0.95$. One of the data points they use is (60, 3.62). What is the residual for this data point?

A. 0.57 B. 0.68

C. 0.72 D. 0.84

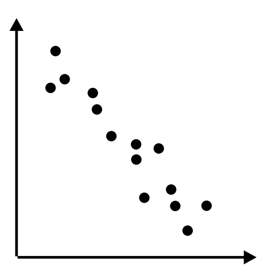
6. Which correlation coefficient is most appropriate for the scatterplot shown below?

A. $r = {}^{-}1.00$

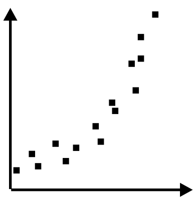
B. $r = {}^{-}0.92$

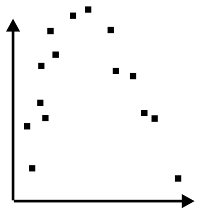
C. $r = 0.78$

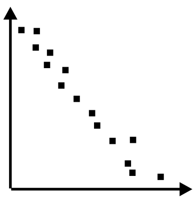
D. $r = 1.00$

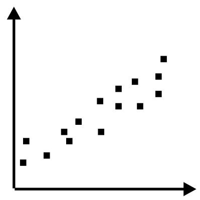


7. Which of the following scatter plots most likely indicates a quadratic regression equation is most appropriate for the data?

A.


B.


C.


D.


8. Regression equation A is $y = 2x^2$ Regression equation B is $y = 2x^3$. Which of the following describes a situation in which the exponential regression equation better fits the data?

A. A: $r = 0.87$; B: $r = 0.93$

B. A: $r = 0.89$; B: $r = 0.99$

C. A: $r = 0.97$; B: $r = 0.97$

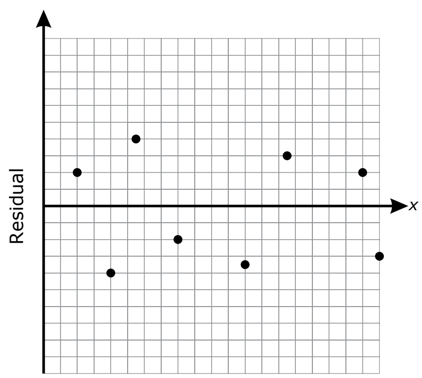
D. A: $r = 0.97$; B: $r = 0.90$

9. Which statistic would indicate that a linear function would not be a good fit to model a data set?

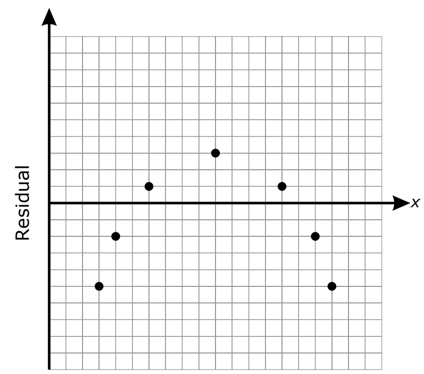
A. $r = {}^{-}0.93$

B. $r = 1$

C.



D.



10. The table below shows the number of grams of carbohydrates, x, and the number of Calories, y, of six different foods.

Carbohydrates (x)	Calories (y)
8	120
9.5	138
10	147
6	88
7	108
4	62

Which equation best represents the line of best fit for this set of data?

A. $y = 15x$

B. $y = 0.07x$

C. $y = 0.1x - 0.4$

D. $y = 14.1x + 5.8$