Linear Regression

Strand:	Algebra and Functions
Topic:	Determining lines of best fit
Primary SOL:	AFDA.3 The student will collect and analyze data, determine the equation of the curve of best fit in order to make predictions, and solve practical problems using models of linear, quadratic, and exponential functions.
Related SOL:	AFDA.1

Materials

- Up to Speed activity sheet (attached)
- See Starbuck Run activity sheet (attached)
- White-water Rafting on Silly Creek (attached)
- Graphing utility

Vocabulary

average rate of change, continuous, dependent variable, domain, function, general equation of a line, independent variable, line of best fit, linear regression, range, relation, residuals, scatterplot, slope, y-intercept

Student/Teacher Actions: What should students be doing? What should teachers be doing?

Time: 50 minutes

1. Review with students how to enter data points into a graphing utility, and scale the scatterplot to fully fit in the viewing window. Review how to access the linear-regression tools on students' graphing utilities.

2. Distribute the Up to Speed activity sheet. Have students complete the sheet in pairs or individually. (Note: Wording within the activity may need to be modified based on the graphing utility being used.)

3. Distribute the See Starbuck Run activity sheet. Have students complete the sheet in pairs or individually. Ideally, students should be in different groups for this activity, to facilitate discussion when comparing the Up to Speed and See Starbuck Run activities. (Note: Wording within the See Starbuck Run activity may need to be modified based on the graphing utility being used.)

4. Distribute the White-water Rafting on Silly Creek activity sheet. Have students complete the sheet in pairs or individually. Ideally, students should be in different groups for this activity than they were for the first two activities to help facilitate discussions comparing all the activities. (Note: Wording within this activity may need to be modified based on the graphing utility being used.)

Assessment

- **Questions**
  - Compare the graphs of the lines of best fit for all three activities. What characteristics did the graphs have in common? Were there any characteristics that were different for all three activities?
  - Lines of best fit are sometimes used to make predictions. How does the table of residual values give us insight about the accuracy of a prediction made using the line of best fit?

- **Journal/writing prompts**
  - A classmate tells you that because the sum of the residuals is small, the line of best fit matches the data closely. Explain why it is important to look at sums of the residuals squared and not only the sum of the residuals.
  - Imagine you are working for an airline carrier and you are in charge of plotting fuel consumption with distance traveled. Your boss mentions that your line of best fit does not seem to be accurate at the beginning and ending of the flights. Give possible explanations, as well as how you could demonstrate the use of residuals to show when the line of best fit is accurate.

- **Other Assessments**
  - As students are working on activity sheets, interview the pairs of students about how the data would need to change to increase/decrease the slope of the line of best fit or change the y-intercept.
  - Provide an exit slip where students are given a set of data and must determine the line of best fit, including whether the line is accurate using the sum of the residuals squared.

Extensions and Connections

- Have students collect their own data to plot. For example, have the class go outside and time students walking or running a set distance. Students with stopwatches can record intermediate times with the fixed distances. An example activity could include having students record the time it takes for a student to run/walk/jog 100 yards, in ten-yard intervals, on a football field.

- Have students gather data from a biology class on a specific animal population in the student's local county/city (e.g., deer population over time in Roanoke County). Use this data to create a line of best fit.

- Have students choose one data point in an activity and change its value by an order of magnitude. Next, have them discuss how one data point can affect a line of best fit.

Strategies for Differentiation

- Review the general form of a line, if needed.
- Use vocabulary cards for related vocabulary listed above.

- Have students pick one of the activities to complete. Afterward, have them pair up with students from the other two activities to compare what they learned from their data.

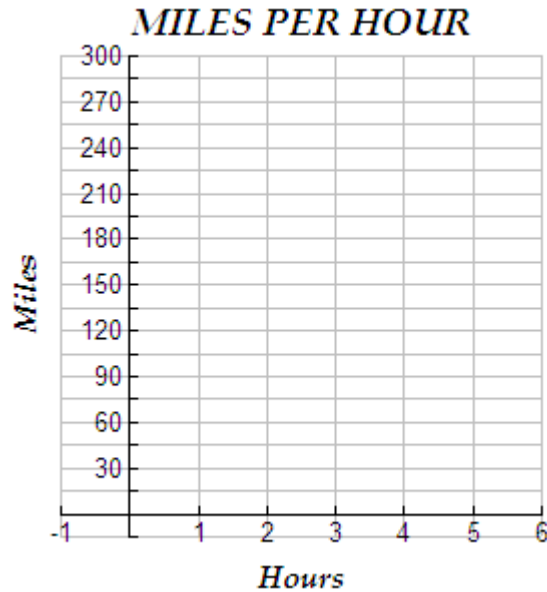- Allow the use of dictionaries for unfamiliar terms.

Note: The following pages are intended for classroom use for students as a visual aid to learning.

Up to Speed

Looking at the Data

1. Create a scatterplot using the data and coordinate plane provided below.

    a. What is the independent variable?

    b. What is the dependent variable?

Hours	Miles
0	0
1	64
2	118
3	178
4	241
5	298

MILES PER HOUR

(empty coordinate plane with y-axis labeled Miles, ranging 30 to 300 in increments of 30, and x-axis labeled Hours ranging -1 to 6)

2. Using a graphing utility, enter data into the lists. Graph the scatterplot in an appropriately sized window.

    a. What is the domain of the relation?
    b. What is the range of the relation?
    c. Is the relation continuous?
    d. Is the relation a function?
    e. What family of functions does the data most resemble?
    f. Write the general form of the equation that would represent the data.

3. What is the average rate of change in miles per hour for the entire trip?

    a. Show computations.

    b. Using your graphing utility's statistics function, determine the equation of the line of best fit using the general form of the equation representing the data. Record the equation for the line of best fit.

        $y =$ _____

    c. What do you notice about your answers in parts 3a and 3b?

4. What is the rate of change in miles per hour when driving, according to the table, from time = zero to the end of the first hour?
    a. Show computations.

    b. Enter the data for the indicated hours and distance into a graphing utility. Readjust the window and graph. Using the graphing utility's statistics function, determine the equation of the line of best fit using the general form of the equation representing the data. Record the equation for the line of best fit.

        y = _____

    c. What do you notice about your responses to parts 4a and 4b?

5. What is the rate of change in speed when driving, according to the table, from the 3rd to the 4th hour?
    a. Show computations.

    b. Enter the data for the indicated hours and distance into a graphing utility. Readjust window and graph. Using the graphing utility's statistics function, determine the equation of the line of best fit using the general form of the equation representing the data. Record the equation for the line of best fit.

        y = _____
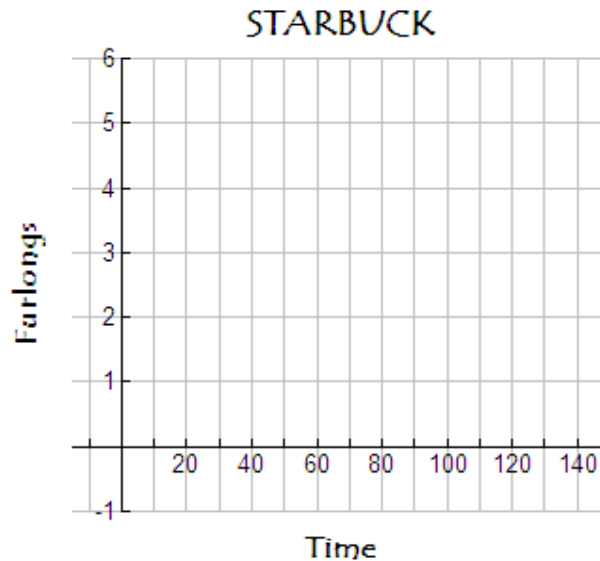
    c. What do you notice about your responses to parts 5a and 5b?

6. Compare each of the answers in 3c, 4c, and 5c. Explain what you noticed in the comparison.

7. What do you think the average rate of change in miles per hour was 2 hours after the start of the trip?

8. Calculate the distance traveled at
    a. 3 hours

    b. 7 hours

    c. 1.5 hours

See Starbuck Run

Mike Millionaire is watching a 10-furlong steeplechase near Charles Town, West Virginia. He is doing some research during the steeplechase in anticipation of attending to watch his favorite horse, Starbuck, at some future date. As Starbuck passes a furlong (F) marker, Mike records the time (t) elapsed in seconds since the beginning of the steeplechase. The data are shown in the table below.

Furlong	Time
0	0
1	23
2	33
3	46
4	59
5	73
6	86
7	100
8	112
9	124
10	135

STARBUCK

(Empty coordinate plane: Furlongs on vertical axis marked 1–6, Time on horizontal axis marked 20, 40, 60, 80, 100, 120, 140)

Looking at the Data

1. Create a scatterplot using the data and coordinate plane above.

    a. What is the independent variable?

    b. What is the dependent variable?

2. Using a graphing utility, enter data into the lists and graph in appropriately sized window.

    a. What is the domain of the relation?
    b. What is the range of the relation?
    c. Is the relation continuous?
    d. Is the relation a function?
    e. What family of functions does the data most resemble?
    f. Write the general form of the equation that would represent the data.

    g. What unit of measure would be appropriate for the average rate of change in furlongs over a given time?

3. How fast is Starbuck running from the start to the very end of his event?

4. How fast is Starbuck running from the exact moment he passes the 4th furlong marker to the moment he passes the 5th furlong marker?

5. How fast is Starbuck running from the moment he passes the 6th furlong marker to the moment he passes the 7th furlong marker?

6. How fast is Starbuck running during the last furlong?

7. How long does it take for Starbuck to finish the event?

8. Using a graphing utility, determine the equation of the line of best fit for the given data. Record the equation.

    $y =$ _____

    a. What does the **a** in the equation of the line of best fit represent?

    b. What does the **b** in the equation of the line of best fit represent?

9. How does the appearance of the data in this activity compare with the Up to Speed activity?

10. How does the equation of the line of best fit relate to the outcomes of questions 3 through 7? Explain.

11. What is the average rate of change when the horse was 46 seconds from the start of the steeplechase?

12. Calculate the total distance traveled at the end of:
    a. 46 seconds

    b. 150 seconds

    c. 90 seconds

Investigation

13. Between which two furlong markers is Starbuck running the fastest? Show your computations and explain in writing.

14. Compare the values recorded in the table with the graph, the average rate of change, and what is happening in the steeplechase. Explain what you notice in the comparison.

White-water Rafting on Silly Creek

White-water rafting enthusiasts in West Virginia enjoy a stretch of 3.60 miles on Silly Creek. A contour map of a section of the river shows a drop in elevation of more than 770 feet. Estimations of the elevations in feet (y) at various distances in miles down the creek (x) from the start of the rafting trip are shown in the table below.

Distance	Elevation
0	1575
0.55	1515
0.91	1445
1.2	1385
1.3	1321
1.42	1261
1.63	1199
1.73	1139
1.97	1041
2.24	997
2.45	931
2.78	862
3.6	801

WHITE WATER RAFTING
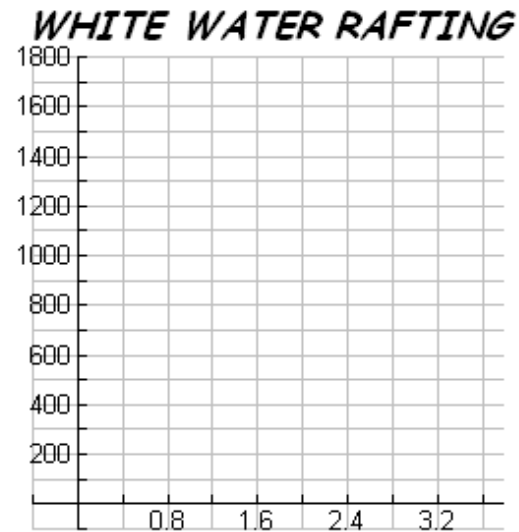
Displaying the Data

1. Create a scatterplot using the above data and coordinate plane provided.
    a. What is the independent variable?

    b. What is the dependent variable?

2. Using a graphing utility, enter data into lists and graph in appropriately sized window.
    a. What is the domain of the relation?
    b. What is the range of the relation?
    c. Is the relation continuous?
    d. Is the relation a function?
    e. What family of functions does the data most resemble?
    f. Write the general form of the equation that would represent the data.
    g. What unit of measure would be appropriate for the average rate of change in elevation over a given distance?

3. Distance is measured in _____.

4. Elevation is measured in _____.

5. What unit of measure would be appropriate for the average rate of change in elevation for a given distance?

6. Calculate the total distance traveled. Show calculations.

7. Calculate the total change in elevation.

8. What is the average rate of change in the elevation over the distance traveled from the start to the very end of the trip for the white-water rafter?

9. Explain how your answers to questions 6 and 7 relate to your findings in question 8.

10. What is the rate of change in the elevation over the distance traveled from the start of the trip, 0 miles, to the time when the white-water rafter passes the 0.55 mile marker?

11. What is the rate of change in the elevation over the distance traveled from the time when the rafter passes the 1.73 mile marker to the time when she passes the 1.97 mile marker?

12. Using a graphing utility, determine the line of best fit using linear regression.

    $y =$ _____

    a. What does ***a*** in the equation of the line of best fit represent?
    b. How is ***a*** related to white-water rafting on Silly Creek?
    c. What does the ***b*** in the equation of the line of best fit represent?
    d. How is ***b*** related to white-water rafting on Silly Creek?

13. How does the data in this activity compare to the data in the Up to Speed activity and the See Starbuck Run activity?

14. Compare the equation of the line of best fit (question 12) to your results in questions 10 and 11. Explain.

15. Determine the average rate of change in the elevation over the distance traveled when the white-water rafters pass the 1.42 mile marker.

16. Calculate the elevation at the specified distances from the beginning of the trip:
    a. 1.3 miles

    b. 1.8 miles

    c. 3.75 miles

Investigation

17. White-water rafting on Silly Creek will be most dangerous between which two points? Explain your reasoning.

18. Relate the values recorded in the table with the graph, the average rate of change in elevation over the distance traveled, and what the white-water rafters experience during their trip.