



Lesson 14: Modeling Relationships with a Line

Student Outcomes

- Students use technology to determine the least squares regression line from a given set of data.
- Students use the least squares regression line to make predictions.

Lesson Notes

Lesson 12 introduced the concept of using linear functions to model the relationship between two numerical variables. In this lesson, students will use technology to determine a least squares regression line to make predictions. Residuals are introduced in this lesson as the rationale for choosing the best-fit line. Students will further evaluate models based on residuals in Lessons 15–16.

As students work with problems in this lesson, encourage them to consider these examples as a potential context for their poster or presentation of Lesson 20.

Classwork

Example 1 (2 minutes): Using a Line to Describe a Relationship

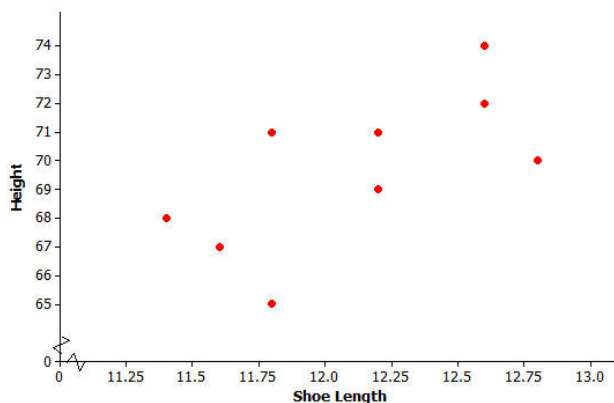
Briefly summarize the data and the graph.

Example 1: Using a Line to Describe a Relationship

Kendra likes to watch crime scene investigation shows on television. She watched a show where investigators used a shoe print to help identify a suspect in a case. She questioned how possible it is to predict someone's height is from his shoe print.

To investigate, she collected data on shoe length (in inches) and height (in inches) from 10 adult men. Her data appear in the table and scatter plot below.

$x = \text{Shoe Length}$	$y = \text{Height}$
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71



Exercises 1–2 (2 minutes)

Discuss the questions as a class.

Exercises 1–2

1. Is there a relationship between shoe length and height?

Yes, it looks like there is a relationship between shoe length and height.

2. How would you describe the relationship? Do the men with longer shoe lengths tend to be taller?

As shoe length increases, height tends to increase. Men with longer shoe lengths tend to be taller.

Example 2 (3 minutes): Using Models to Make Predictions**Example 2: Using Models to Make Predictions**

When two variables x and y are linearly related, you can use a line to describe their relationship. You can also use the equation of the line to predict the value of the y variable based on the value of the x variable.

For example, the line $y = 25.3 + 3.66x$ might be used to describe the relationship between shoe length and height, where x represents shoe length and y represents height. To predict the height of a man with a shoe length of 12, you would substitute 12 in for x in the equation of the line and then calculate the value of y .

$$y = 25.3 + 3.66x = 25.3 + 3.66(12) = 69.22$$

You would predict a height of 69.22 inches for a man with a shoe length of 12 inches.

Remind students that a model (equation) can be used to describe the relationship between variables.

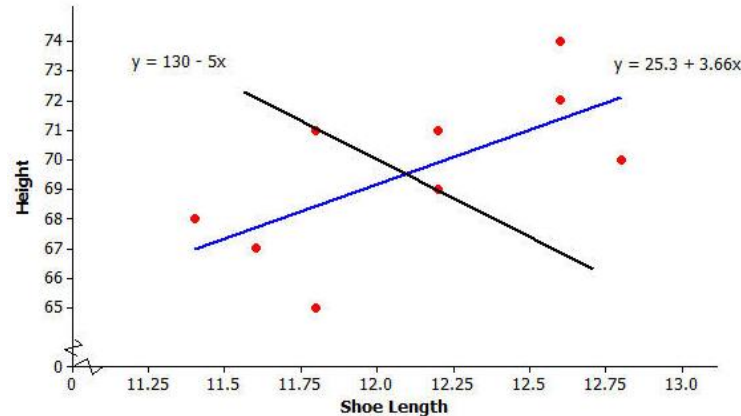
- Explain that the model can be used to make predictions by substituting a value for one variable to solve for the other.
- Discuss how to use the model given in the student lesson to predict a man's height using a shoe length of 12 inches.

Exercises 3–7 (7 minutes)

Introduce the scatter plot and models in Exercise 3. Discuss why neither line is a perfect model. Let students work in pairs to answer Exercises 3–7. After they complete their work, confirm as a class.

Exercises 3–7

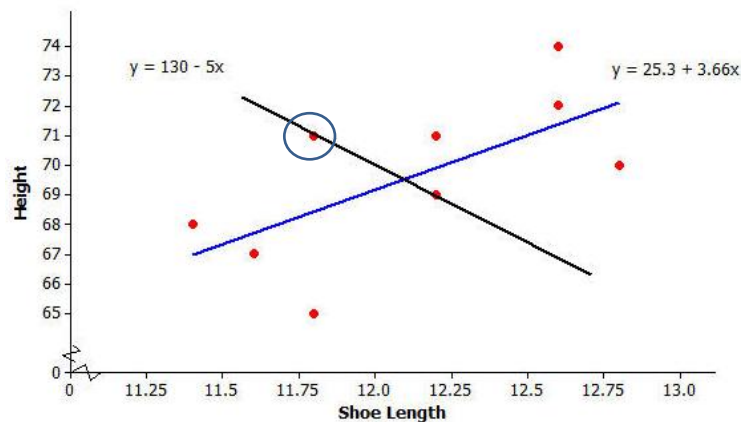
3. Below is a scatter plot of the data with two linear models, $y = 130 - 5x$ and $y = 25.3 + 3.66x$. Which of these two models does a better job of describing how shoe length (x) and height (y) are related? Explain your choice.



The line of $y = 25.3 + 3.66x$ is a better choice because as shoe length increases, height tends to increase.

4. One of the men in the sample has a shoe length of 11.8 inches and a height of 71 inches. Circle the point in the scatter plot in Question 3 that represents this man.

The point (11.8, 71) should be circled on the scatter plot.



5. Suppose that you do not know this man's height, but do know that his shoe length is 11.8 inches. If you use the model $y = 25.3 + 3.66x$, what would you predict his height to be? If you use the model $y = 130 - 5x$, what would you predict his height to be?

For the first line: $y = 25.3 + 3.66(11.8) = 68.488$ inches, or approximately 68 inches.

For the second line: $y = 130 - 5(11.8) = 71$ inches.

6. Which model was closer to the actual height of 71 inches? Is that model a better fit to the data? Explain your answer.

The model $y = 130 - 5x$ was closer to the actual value, but that is not the better model because $y = 25.3 + 3.66x$ is closer to more points.

7. Is there a better way to decide which of two lines provides a better description of a relationship (rather than just comparing the predicted value to the actual value for one data point in the sample)?

The better line (or the line with the best fit) is the one in which the distances from the points to the line are as short as possible. Highlight a few of the points and the distances from the points to the line. Highlight the vertical distances from various points to the line to show students how the values of residuals help us evaluate a line's fit.

Example 3 (3 minutes): Residuals

Explain the definition of residual. Work through the example in the text with students. Note that the first row of the table in Exercise 8 represents the example. If students struggle, work through another row of the table.

Example 3: Residuals

One way to think about how useful a line is for describing a relationship between two variables is to use the line to predict the y -values for the points in the scatter plot. These predicted values could then be compared to the actual y -values.

For example, the first data point in the table represents a man with a shoe length of 12.6 inches and height of 74 inches. If you use the line $y = 25.3 + 3.66x$ to predict this man's height, you would get:

$$\begin{aligned} y &= 25.3 + 3.66x \\ &= 25.3 + 3.66(12.6) \\ &= 71.42 \text{ in.} \end{aligned}$$

Because his actual height was 74 inches, you can calculate the prediction error by subtracting the predicted value from the actual value. This prediction error is called a *residual*. For the first data point, the residual is calculated as follows:

$$\begin{aligned} \text{Residual} &= \text{actual } y\text{-value} - \text{predicted } y\text{-value} \\ &= 74 - 71.42 \\ &= 2.58 \text{ in.} \end{aligned}$$

Exercises 8–10 (7 minutes)

Let students work in pairs to complete Exercises 8–10. Then discuss and confirm Exercises 9–10 as a class.

Exercises 8–10

8. For the line $y = 25.3 + 3.66x$, calculate the missing values and add them to complete the table.

x = Shoe Length	y = Height	Predicted y -value	Residual
12.6	74	71.42	2.58
11.8	65	68.49	−3.49
12.2	71	69.95	1.05
11.6	67	67.76	−0.76
12.2	69	69.95	−0.95
11.4	68	67.02	0.98
12.8	70	72.15	−2.15
12.2	69	69.95	−0.95
12.6	72	71.42	0.58
11.8	71	68.49	2.51

9. Why is the residual in the table's first row positive, and the residual in the second row negative?

The actual value is larger than the predicted value in the first row, so the difference is positive. In the second row, the actual value is less than the predicted value, so the difference is negative.

10. What is the sum of the residuals? Why did you get a number close to zero for this sum? Does this mean that all of the residuals were close to 0?

The sum of the residuals is -0.6 . This number is close to 0 because the positive and negative values of the residuals balance each other out. Not all residuals are close to zero. (Refer to the table.)

Exercises 11–13 (5 minutes)

Work on the exercises as a class. Allow for multiple student responses:

Explain that the line of best fit will be the model where the absolute values of the residuals (vertical distances to the line) are small. Refer back to the two linear models in Exercise 3. The residuals for the line of best fit ($y = 25.3 + 3.66x$) are much smaller than those for the other line ($y = 130 - 5x$) because in general, the points lie closer to the better model. Pose the question in Exercise 11 to the class.

Exercises 11–13

When you use a line to describe the relationship between two numerical variables, the *best* line is the line that makes the residuals as small as possible overall.

11. If the residuals tend to be small, what does that say about the fit of the line to the data?

If the residuals are small, the points tend to be close to the line.

Explain that it is the sum of the *squared* residuals that will lead to choosing the line of best fit. Tell students to square the residuals in the table (Exercise 8) and place their answers in a column they construct to the right of the table. Pose the following questions one at a time to the class.

The most common choice for the *best* line is the line that makes the sum of the *squared* residuals as small as possible. Add a column on the right of the table in Exercise 8. Calculate the square of each residual and place the answer in the column.

12. Why do we use the sum of the squared residuals instead of just the sum of the residuals (without squaring)? Hint: Think about whether the sum of the residuals for a line can be small even if the prediction errors are large. Can this happen for squared residuals?

If you just add all of the residuals, positive and negative residuals can offset one another and you can get a small sum even if some of the residuals are "big." Squaring the residuals gives positive values that still reflect how far the data points are from the line.

13. What is the sum of the squared residuals for the line $y = 25.3 + 3.66x$ and the data of Exercise 11?

The sum of the squared residuals (rounded to the nearest hundredth) is 34.54.

Example 4 (5 minutes): The Least Squares Line (Best-Fit Line)

The line that has a smaller sum of squared residuals than any other line is called the *least squares line*. Tell students that there are formulas for determining the least squares line, but that these formulas are tedious and time-consuming, so they will instead use a graphing calculator or statistical software to generate the best-fit line.

Example 4

The line that has a smaller sum of squared residuals for this data set than any other line is called the *least squares line*. This line can also be called the *best-fit line* or the *line of best fit* (or regression line).

For the shoe-length and height data for the sample of 10 men, the line $y = 25.3 + 3.66x$ is the least squares line. No other line would have a smaller sum of squared residuals for this data set than this line.

There are equations that can be used to calculate the value for the slope and the intercept of the least squares line, but these formulas require a lot of tedious calculations. Fortunately, a graphing calculator can be used to find the equation of the least squares line.

Your teacher will show you how to enter data and obtain the equation of the least squares line using your graphing calculator or other statistics program.

The following outlines how to determine the least squares regression using a TI-84 Plus. Most graphing calculators use a similar process. Having students use software that emulates a graphing calculator is also an option. Use this outline as a guide for working with students. Modify these steps for the available technology.

Finding the Regression Line (TI-84 Plus)

Step 1: From your home screen, press STAT.

Step 2: From the STAT menu, select the EDIT option. (EDIT enter)

Step 3: Enter the x -values of the data set in L1.

Step 4: Enter the y -values of the data set in L2.

Step 5: Select STAT. Move cursor to the menu item CALC and then move the cursor to option 4: LinReg($ax + b$) or option 8: LinReg($a + bx$). Press enter. (Discuss with students that both options 4 and 8 are representations of a linear equation. Anticipate that most students will be familiar with option 4, or the slope y -intercept form. Option 8 is essentially the same representation using different letters to represent slope and y -intercept. Option 8 is the preferred option in statistical studies.)

Step 6: With option 4 or option 8 on the screen, enter L1, L2, and Y1 as described in the following notes.

LinReg($a + bx$) L1, L2, Y1

Select enter to see results. The least squares regression will be stored in Y1. Work with students in graphing the scatter plot and Y1.

Note: L1 represents the x -values of the regression function, L2 the y -values, and Y1 represents the least squares regression function.

To obtain Y1, go to VARS, move cursor to Y-VARS, and then Functions (enter). You are now at the screen highlighting the y -variables. Move cursor to Y1 and hit enter.

Y1 is the linear regression line and will be stored in Y1.

Exercises 14–17 (5 minutes)

Let students work independently on Exercises 14–17. Assist students who need help using their calculators.

14. Enter the shoe-length and height data and then use your calculator to find the equation of the least squares line. Did you get $y = 25.3 + 3.66x$? (The slope and y-intercept here have been rounded to the nearest hundredth.)

The exact equation of the least squares line is $Y1 = 25.276859504133 + 3.6570247933884x$. When rounded to the nearest tenth for the y-intercept and hundredth for the slope, the equation is $y = 25.3 + 3.66x$.

15. Assuming that the 10 men in the sample are representative of adult men in general, what height would you predict for a man whose shoe length is 12.5 inches? What height would you predict for a man whose shoe length is 11.9 inches?

The predicted height of a man whose shoe length is 12.5 inches is 71.05 inches. The predicted height of a man whose shoe length is 11.9 inches is 68.86 inches.

Once you have found the equation of the least squares line, the values of the slope and y-intercept of the line often reveals something interesting about the relationship you are modeling.

The slope of the least squares line is the change in the predicted value of the y-variable associated with an increase of one in the value of the x-variable.

16. Give an interpretation of the slope of the least squares line $y = 25.3 + 3.66x$ for predicting height from shoe size for adult men.

The slope is 3.66; for every one-inch increase in a man's shoe length, you would estimate an increase of 3.66 inches in height.

The y-intercept of a line is the predicted value of y when x equals zero. When using a line as a model for the relationship between two numerical variables, it often does not make sense to interpret the y-intercept because an x-value of zero may not make any sense.

17. Explain why it does not make sense to interpret the y-intercept of 25.3 as the predicted height for an adult male whose shoe length is zero.

The y-intercept is (0, 25.3). Since x represents shoe length, it is impossible for the shoe length to be 0 inches when a man is 25.3 inches tall.

MP.4

Closing (2 minutes)

Lesson Summary

When the relationship between two numerical variables x and y is linear, a straight line can be used to describe the relationship. Such a line can then be used to predict the value of y based on the value of x . When a prediction is made, the prediction error is the difference between the actual y-value and the predicted y-value. The prediction error is called a residual, and the residual is calculated as residual = actual y-value – predicted y-value. The least squares line is the line that is used to model a linear relationship. The least squares line is the *best* line in that it has a smaller sum of squared residuals than any other line.

Exit Ticket (4 minutes)

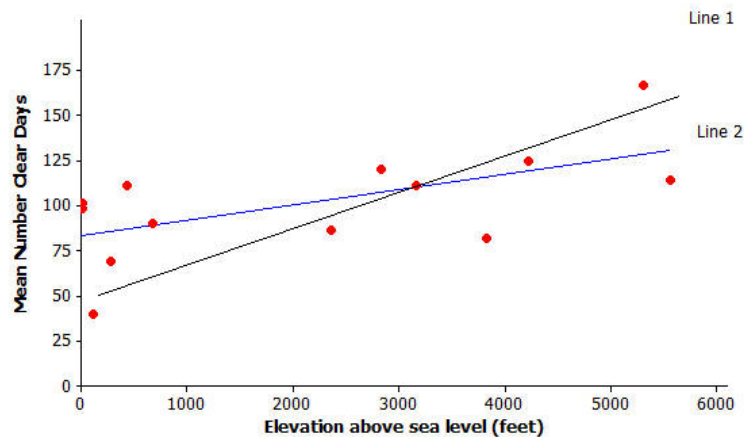
Name _____

Date _____

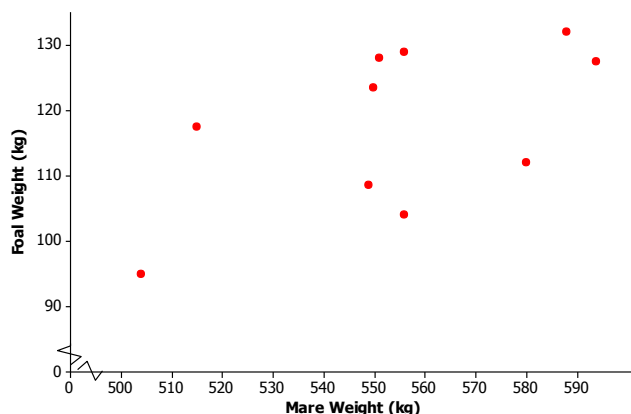
Lesson 14: Modeling Relationships with a Line

Exit Ticket

1. The scatter plot below displays the elevation and mean number of clear days per year of 14 U.S. cities. Two lines are shown on the scatter plot. Which represents the least squares line? Explain your choice.



2. Below is a scatter plot of foal birth weight and mare's weight.



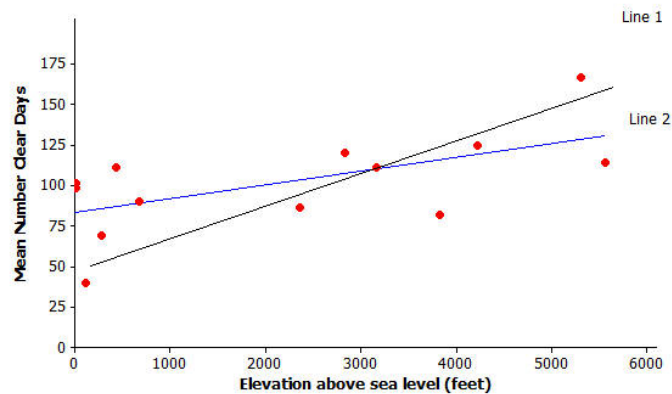
- a. The equation of the least squares line for the data is $y = -19.6 + 0.248x$, where x = mare's weight (in kg) and y = foal's birth weight (in kg).
What foal birth weight would you predict for a mare that weighs 520 kg?

- b. How would you interpret the value of the slope in the least squares line?

- c. Does it make sense to interpret the value of the y -intercept in this context? Explain why or why not.

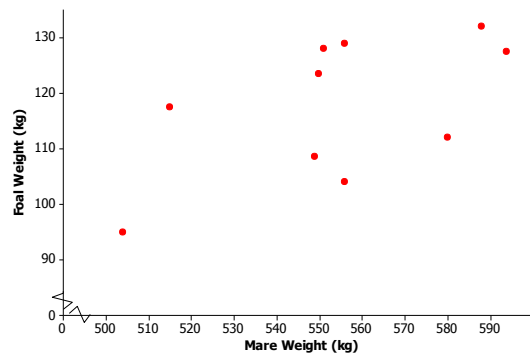
Exit Ticket Sample Solutions

1. The scatter plot below displays the elevation and mean number of clear days per year of 14 U.S. cities. Two lines are shown on the scatter plot. Which represents the least squares line? Explain your choice.



Line 2, because it appears to have smaller residuals overall.

2. Below is a scatter plot of foal birth weight and mare's weight.



- a. The equation of the least squares line for these data is $y = -19.6 + 0.248x$, where x = mare's weight (in kg) and y = foal's birth weight (in kg).

What foal birth weight would you predict for a mare that weighs 520 kg?

$$y = -19.6 + 0.248(520) \Rightarrow y = 109.36 \text{ kg to the nearest hundredth.}$$

- b. How would you interpret the value of the slope in the least squares line?

The value of the slope is 0.248, indicating that predicted foal weight increases by 0.248 kg for each one kg increase in the mare's weight.

- c. Does it make sense to interpret the value of the y-intercept in this context? Explain why or why not.

It does not make sense to interpret the value of the y-intercept because it is equivalent to predicting the foal's weight for a mare's weight of zero, which does not make sense.

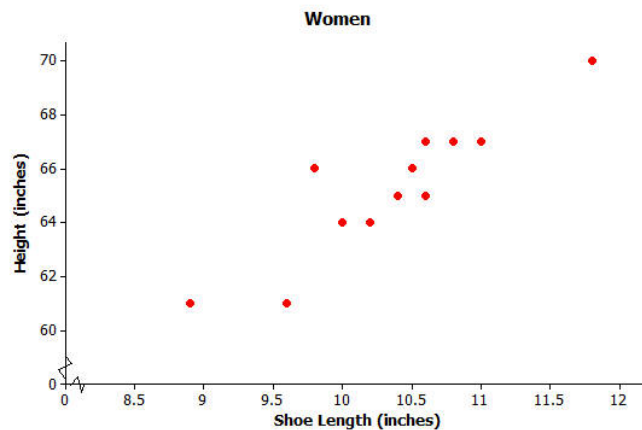
Problem Set Sample Solutions

If time permits, the following data set and questions provide an opportunity to give hands-on examples. Consider providing actual shoe prints from students or individuals (children and adults). Create a scenario in which the shoe print is “evidence,” and have the class figure out “who done it?”

Kendra wondered if the relationship between shoe length and height might be different for men and women. To investigate, she also collected data on shoe length (in inches) and height (in inches) for 12 women.

x = Shoe Length (Women)	y = Height (Women)
8.9	61
9.6	61
9.8	66
10.0	64
10.2	64
10.4	65
10.6	65
10.6	67
10.5	66
10.8	67
11.0	67
11.8	70

- Construct a scatter plot of these data.



- Is there a relationship between shoe length and height for these 12 women?
Yes, as shoe length increases, height also tends to increase.
- Find the equation of the least squares line. (Round values to the nearest hundredth.)
The least squares line is $y = 32.68 + 3.15x$.

4. Suppose that these 12 women are representative of adult women in general. Based on the least squares line, what would you predict for the height of a woman whose shoe length is 10.5 inches? What would you predict for the height of a woman whose shoe length is 11.5 inches?

The predicted height for a woman whose shoe length is 10.5 inches would be 65.76 inches, or 66 inches to the nearest inch. The predicted height for a woman whose shoe length is 11.5 inches would be 68.91 inches, or 69 inches to the nearest inch.

5. One of the women in the sample had a shoe length of 9.8 inches. Based on the regression line, what would you predict for her height?

Her predicted height is $y = 32.68 + 3.15(9.8) = 63.55$ inches.

6. What is the value of the residual associated with the observation for the woman with the shoe length of 9.8?

The residual would be the actual value minus the predicted value, or $66 - 63.55 = 2.45$ inches.

7. Add the predicted value and the residual you just calculated to the table below. Then, calculate the sum of the squared residuals.

The sum of the squared residuals is 12.94. (Note: The predicted heights and residuals come from the rounded values in the model, not the exact values generated from the regression line using the calculator.)

x = Shoe Length (Women)	y = Height (Women)	Predicted Height	Residual
8.9	61	60.72	0.28
9.6	61	62.92	-1.92
9.8	66	63.55	2.45
10.0	64	64.18	-0.18
10.2	64	64.81	-0.81
10.4	65	65.44	-0.44
10.6	65	66.07	-1.07
10.6	67	66.07	0.93
10.5	66	65.76	0.24
10.8	67	66.7	0.3
11.0	67	67.33	-0.33
11.8	70	69.85	0.15

8. Provide an interpretation of the slope of the least squares line.

We would predict an increase of 3.15 inches in height for an increase of 1 inch in shoe length.

9. Does it make sense to interpret the y -intercept of the least squares line in this context? Explain why or why not.

The intercept would be based on a shoe length of zero inches. This would not make sense.

10. Would the sum of the squared residuals for the line $y = 25 + 2.8x$ be greater than, about the same as, or less than the sum you computed in Question 7? Explain how you know this. You should be able to answer this question without calculating the sum of squared residuals for this new line.

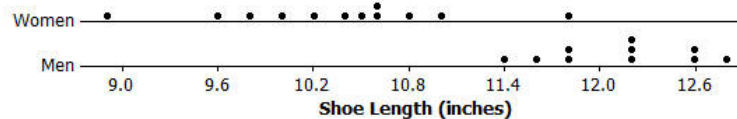
The sum of squared residuals will be greater for the line $y = 25 + 2.8x$. The equation $y = 32.68 + 3.15x$ is the approximate equation of the least squares line, so it has a smaller sum of squared residuals than any other line.

11. For the men, the least squares line that describes the relationship between x = shoe length (in inches) and y = height (in inches) was $y = 25.3 + 3.66x$. How does this compare to the equation of the least squares line for women? Would you use $y = 25.3 + 3.66x$ to predict the height of a woman based on her shoe length? Explain why or why not.

The slope is greater for men; a one-inch increase in shoe length translates to a greater increase in height for men compared to women. The y -intercept is greater for women. To predict the height of a woman, the least squares line for women must be used because it is the best fit.

12. Below are dot plots of the shoe lengths for women and the shoe lengths for men. Suppose that you found a shoe print and that when you measured the shoe length, you got 10.8 inches. Do you think that a man or a woman left this shoe print? Explain your choice.

It is more likely that a woman left the shoe print with length 10.8 inches based on the clustering around the value.



13. Suppose that you find a shoe print and the shoe length for this print is 12 inches. What would you predict for the height of the person who left this print? Explain how you arrived at this answer.

The print most likely belongs to a man. Use the least squares line for men to predict a height, or $y = 25.3 + 3.66(12)$, which is 69.22 inches, to the nearest hundredth of an inch.