

Lesson 14: Modeling Relationships with a Line

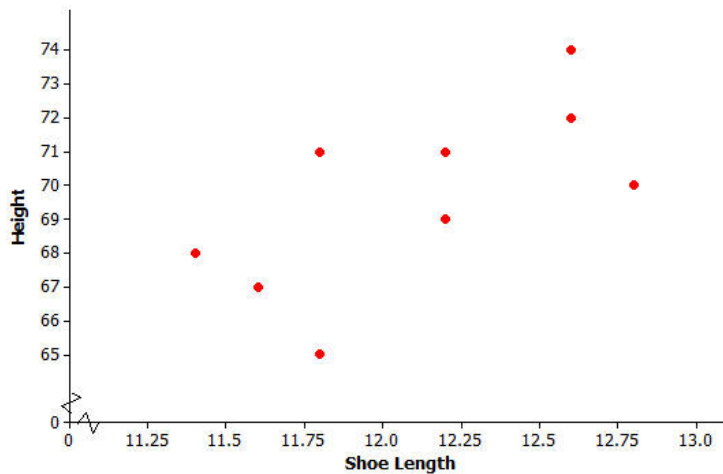
Classwork

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

- Is there a relationship between shoe length and height?
- How would you describe the relationship? Do the men with longer shoe lengths tend to be taller?

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.

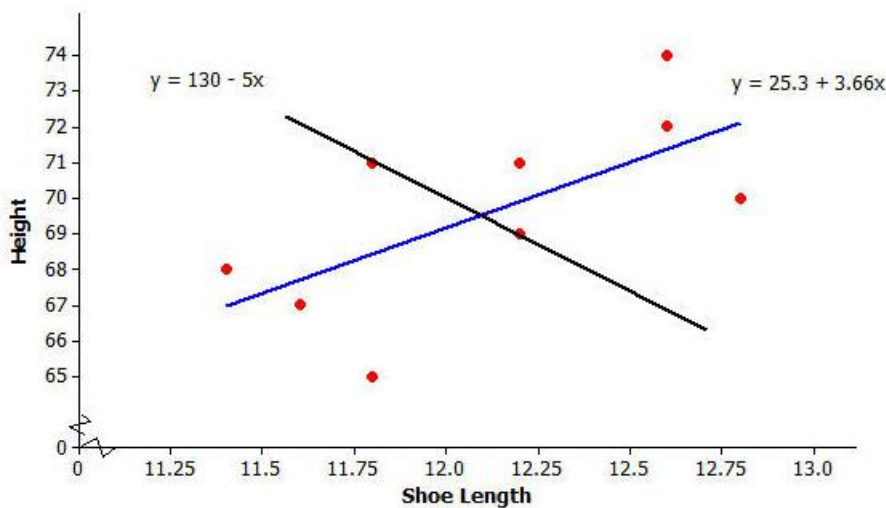
For example, the line $y = 25.3 + 3.66x$ might be used to describe the relationship between shoe length and height, where x represents shoe length and y represents height. To predict the height of a man with a shoe length of 12, you would substitute 12 in for x in the equation of the line and then calculate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

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.



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.

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

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

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		-3.49
12.2	71		
11.6	67	67.76	-0.76
12.2	69	69.95	-0.95
11.4	68	67.02	
12.8	70	72.15	-2.15
12.2	69		-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?
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?

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?

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?

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

Example 4: The Least squares Line (Best-Fit Line)

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.

Exercises 14–17

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.)

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?

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 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.

Lesson Summary

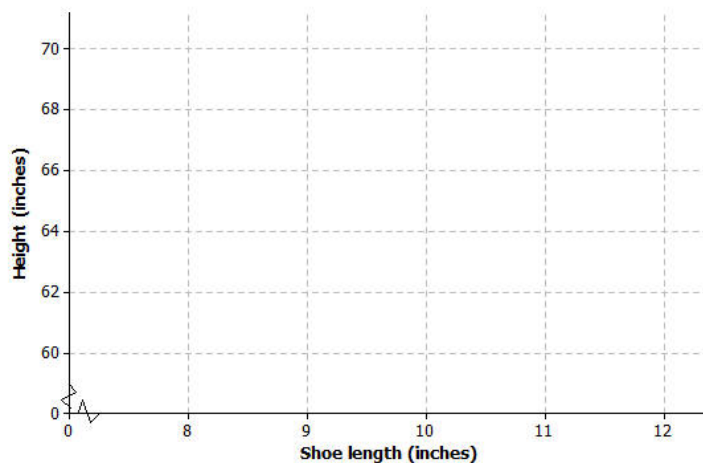
When the relationship between two numerical variables x and y is linear, a straight line can be used to describe the relationship. Such a line can then be used to predict the value of y based on the value of x . When a prediction is made, the prediction error is the difference between the actual y -value 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.

Problem Set

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.

$x = \text{Shoe Length (Women)}$	$y = \text{Height (Women)}$
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

1. Construct a scatter plot of these data.
2. Is there a relationship between shoe length and height for these 12 women?
3. Find the equation of the least squares line. (Round values to the nearest hundredth.)

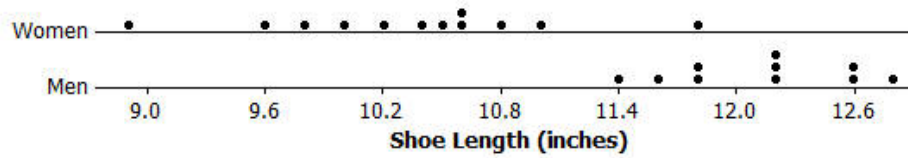


4. Suppose that these 12 women are representative of adult women in general. Based on the least squares line, what would you predict for the height of a woman whose shoe length is 10.5 inches? What would you predict for the height of a woman whose shoe length is 11.5 inches?
5. One of the women in the sample had a shoe length of 9.8 inches. Based on the regression line, what would you predict for her height?
6. What is the value of the residual associated with the observation for the woman with the shoe length of 9.8?
7. Add the predicted value and the residual you just calculated to the table below. Then, calculate the sum of the squared residuals.

$x = \text{Shoe Length (Women)}$	$y = \text{Height (Women)}$	Predicted Height	Residual
8.9	61	60.72	0.28
9.6	61	62.92	-1.92
9.8	66		
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.
9. Does it make sense to interpret the y -intercept of the least squares line in this context? Explain why or why not.
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.
11. For the men, the least squares line that describes the relationship between $x = \text{shoe length (in inches)}$ and $y = \text{height (in inches)}$ 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.

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.



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.