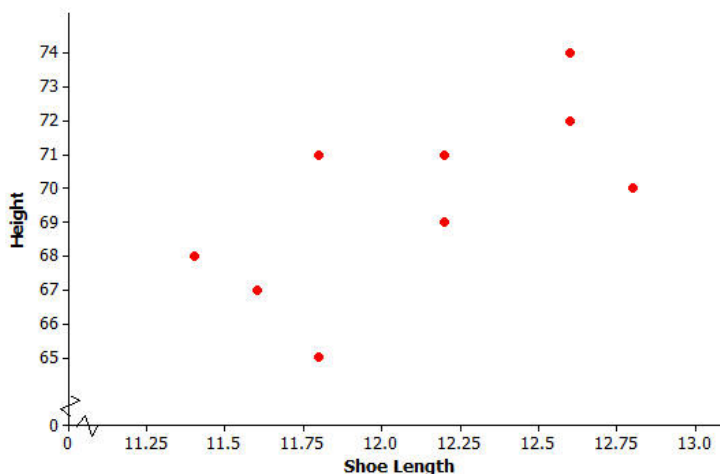


Using a Line to Describe a Relationship

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 = Shoe Length	y = 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. Is there 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?

- Write the linear regression equation for this data, rounding values to the nearest hundredth.
- 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?

Once you have found the regression equation, 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 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.

- 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.

- 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.

Example: 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.

You can calculate the prediction error by subtracting the predicted value from the actual value. This prediction error is called a *residual*. For each data point, the residual is calculated as follows:

$$\text{Residual} = \text{actual } y \text{ value} - \text{predicted } y \text{ value}$$

A. Use your regression equation to calculate the missing values and add them to complete the table.

$x = \text{Shoe Length}$	$y = \text{Height}$	Predicted y -value	Residual	Square of the Residual
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			

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

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

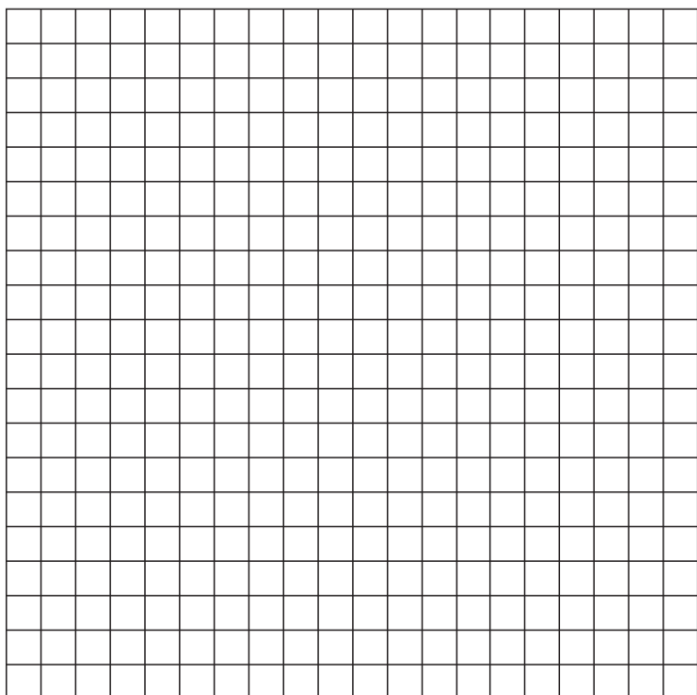
D. What is the sum of the residuals? Why did you get a number close to zero for this sum? Does this mean that all of the residuals were close to 0?

E. What is the sum of the squared residuals?

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. The most common choice for the *best* line is the line that makes the sum of the *squared* residuals as small as possible. The regression equation on your calculator is the line that makes these squares the least.

F. Why do we use the sum of the squared residuals instead of just the sum of the residuals (without squaring)?

G. Create a residual plot of the data.



H. How might a residual plot help you determine whether or not a regression equation is a good model for the data?

I. Based on your residual plot, do you think this regression equation is a good fit for the data?
Justify your reasoning.

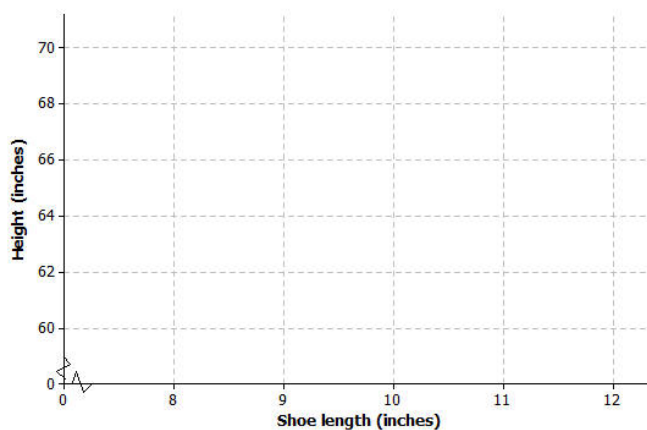
Name _____ Date _____

Lesson 14: Residuals - Modeling Relationships with a Line **CW/HW**

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

- Construct a scatter plot of these data.



- Is there a relationship between shoe length and height for these 12 women?

- 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?

5. Determine the predicted values and the residuals on the table below.

x = Shoe Length (Women)	y = Height (Women)	Predicted Height	Residual
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		

6. Create a residual plot of the data.

