Lesson 9: Determining the Equation of a Line Fit to Data

Student Outcomes

- Students informally fit a straight line to data displayed in a scatter plot.
- Students determine the equation of a line fit to data.
- Students make predictions based on the equation of a line fit to data.

Lesson Notes

In this lesson, students informally fit a line to data by drawing a line that describes a linear pattern in a scatter plot and then use their lines to make predictions. They determine the equation of the line and informally analyze different lines fit to the same data. This lesson begins developing the foundation for finding an objective way to judge how well a line fits the trend in a scatter plot and the notion of a *best-fit* line in Algebra I.

Classwork

Example 1 (5 minutes): Crocodiles and Alligators

Discuss the data presented in the table and scatter plot. You might start by asking if students are familiar with crocodiles and alligators and how they differ. Ask students if they can imagine what a bite force of 100 pounds would feel like. Ask them if they know what body mass indicates. If students understand that body mass is an indication of the weight of a crocodilian and bite force is a measure of the strength of a crocodilian's bite, the data can be investigated even if they do not understand the technical definitions and how these variables are measured. Also, ask students if any other aspects of the data surprised them. For example, did they realize that there are so many different species of crocodilian? Did the wide range of body mass and bite force surprise them? If time permits, you may want to suggest that students do further research on crocodilian.

Example 1: Crocodiles and Alligators

Scientists are interested in finding out how different species adapt to finding food sources. One group studied crocodilian to find out how their bite force was related to body mass and diet. The table below displays the information they collected on body mass (in pounds) and bite force (in pounds).

Crocodilian Biting

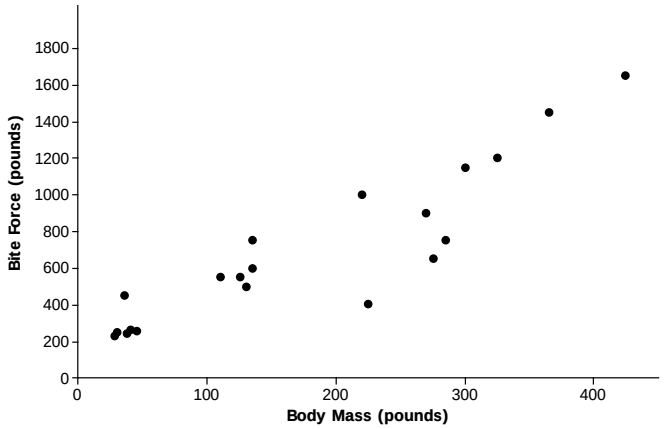
Species	Body Mass (Pounds)	Bite Force (Pounds)
Dwarf Crocodile	35	450
Crocodile F	40	260
Alligator A	30	250
Caiman A	28	230
Caiman B	37	240
Caiman C	45	255
Crocodile A	110	550
Nile Crocodile	275	650
Crocodile B	130	500
Crocodile C	135	600
Crocodile D	135	750
Caiman D	125	550
Indian Gharial Crocodile	225	400
Crocodile G	220	1,000
American Crocodile	270	900
Crocodile D	285	750
Crocodile E	425	1,650
American Alligator	300	1,150
Alligator B	325	1,200
Alligator C	365	1,450

Data Source: PLoS One Greg Erickson biomechanics, Florida State University

Scaffolding:

- The word *crocodilian* refers to any reptile of the order Crocodylia.
- This includes crocodiles, alligators, caimans, and gavials. Showing students a visual aid with pictures of these animals may help them understand.

As you learned in the previous lesson, it is a good idea to begin by looking at what a scatter plot tells you about the data. The scatter plot below displays the data on body mass and bite force for the crocodilian in the study.



Exercises 1–5 (14 minutes)

Exercises 1 through 5 ask students to consider the fit of a line. Each student (or small group of students) draws a line that would be a good representation of the trend in the data. Students evaluate their lines and the lines of the four students introduced in Exercise 4.

In Exercise 2, students draw a line they think will be a good representation of the trend in the data. Ask them to compare their line with other students. As a group, decide who might have the best line, and ask students why they made that choice. Have groups share their ideas. Point out that it would be helpful to agree on a standard method for judging the fit of a line. One method is to look at how well the line predicts for the given data or how often it is over or under the actual or observed value.

Scaffolding:

Point out to ELL students that the terms *body mass* and *weight* are used interchangeably in this lesson.

Exercises 1–6

1. **Describe the relationship between body mass and bite force for the crocodilian shown in the scatter plot.**

   As the body mass increases, the bite force tends to also increase.

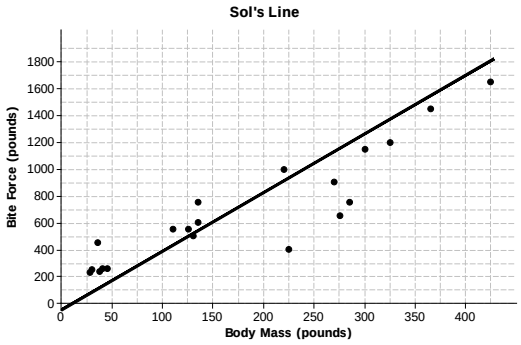
2. **Draw a line to represent the trend in the data. Comment on what you considered in drawing your line.**

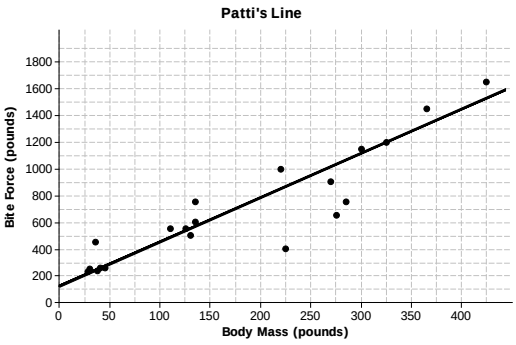
   The line should be as close as possible to the points in the scatter plot. Students explored this idea in Lesson 8.

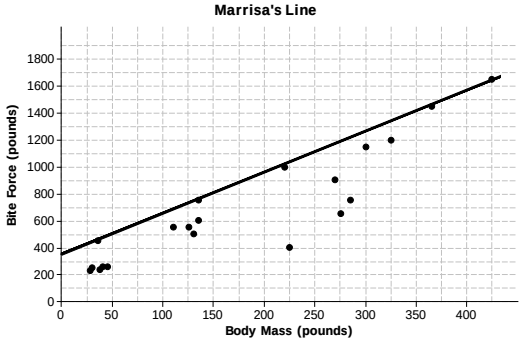
3. **Based on your line, predict the bite force for a crocodilian that weighs 220 pounds. How does this prediction compare to the actual bite force of the 220-pound crocodilian in the data set?**

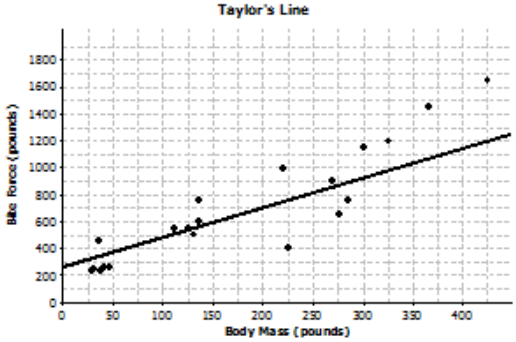
   Answers will vary. A reasonable prediction is around 650 to 700 pounds. The actual bite force was $1,000$ pounds, so the prediction based on the line was not very close for this crocodilian.

4. **Several students decided to draw lines to represent the trend in the data. Consider the lines drawn by Sol, Patti, Marrisa, and Taylor, which are shown below.**









MP.2

MP.2

For each student, indicate whether or not you think the line would be a good line to use to make predictions. Explain your thinking.

a. Sol's line

In general, it looks like Sol's line overestimates the bite force for heavier crocodilian and underestimates the bite force for crocodilian that do not weigh as much.

b. Patti's line

Patti's line looks like it fits the data well, so it would probably produce good predictions. The line goes through the middle of the points in the scatter plot, and the points are fairly close to the line.

c. Marrisa's line

It looks like Marrisa's line overestimates the bite force because almost all of the points are below the line.

d. Taylor's line

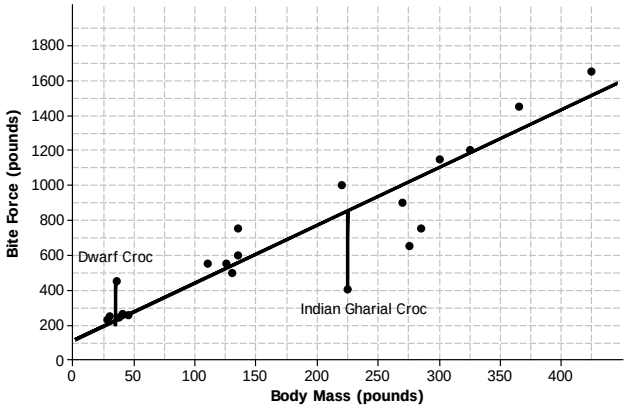
It looks like Taylor's line tends to underestimate the bite force. There are many points above the line.

5. What is the equation of your line? Show the steps you used to determine your line. Based on your equation, what is your prediction for the bite force of a crocodilian weighing 200 pounds?

Answers will vary. Students have learned from previous modules how to find the equation of a line. Anticipate that students will first determine the slope based on two points on their lines. Students then use a point on the line to obtain an equation in the form $y = mx + b$ (or $y = a + bx$). Students will use their line to predict a bite force for a crocodilian that weighs 200 pounds. A reasonable answer would be around 800 pounds.

Exercise 6 (5 minutes)

6. Patti drew vertical line segments from two points to the line in her scatter plot. The first point she selected was for a Dwarf Crocodile. The second point she selected was for an Indian Gharial Crocodile.



a. Would Patti's line have resulted in a predicted bite force that was closer to the actual bite force for the Dwarf Crocodile or for the Indian Gharial Crocodile? What aspect of the scatter plot supports your answer?

The prediction would be closer to the actual bite force for the Dwarf Crocodile. That point is closer to the line (the vertical line segment connecting it to the line is shorter) than the point for the Indian Gharial Crocodile.

b. Would it be preferable to describe the trend in a scatter plot using a line that makes the differences in the actual and predicted values large or small? Explain your answer.

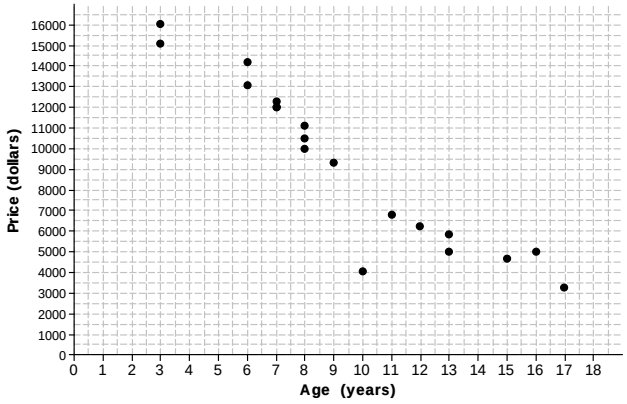
It would be better for the differences to be as small as possible. Small differences are closer to the line.

Exercise 7 (14 minutes): Used Cars

This exercise provides additional practice for students. Students use the equation of a line to make predictions and informally assess the fit of the line.

Exercise 7: Used Cars

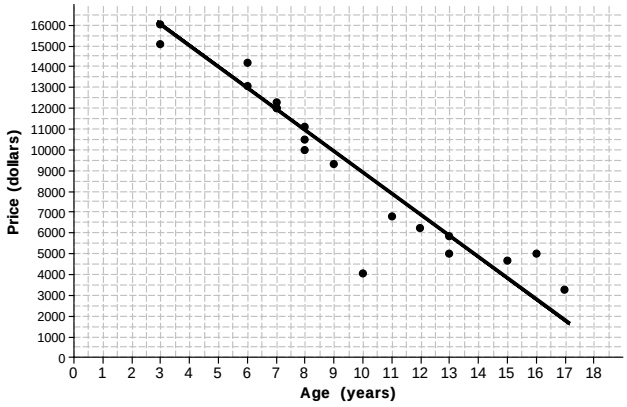
7. The plot below shows the age (in years) and price (in dollars) of used Honda Civic cars that were advertised in a local newspaper.



a. Based on the scatter plot above, describe the relationship between the age and price of the used cars.

The older the car, the lower the price tends to be.

b. Nora drew a line she thought was close to many of the points and found the equation of the line. She used the points $(13, 6000)$ and $(7, 12000)$ on her line to find the equation. Explain why those points made finding the equation easy.



The points are at the intersection of the grid lines in the graph, so it is easy to determine the coordinates of these points.

c. **Find the equation of Nora's line for predicting the price of a used car given its age. Summarize the trend described by this equation.**

Using the points, the equation is $y = -1000x + 19000$, or $\text{Price} = -1000(\text{age}) + 19000$. The slope of the line is negative, so the line indicates that the price of used cars decreases as cars get older.

d. **Based on the line, for which car in the data set would the predicted value be farthest from the actual value? How can you tell?**

It would be farthest for the car that is 10 years old. It is the point in the scatter plot that is farthest from the line.

e. **What does the equation predict for the cost of a 10-year-old car? How close was the prediction using the line to the actual cost of the 10-year-old car in the data set? Given the context of the data set, do you think the difference between the predicted price and the actual price is large or small?**

The line predicts a 10-year-old car would cost about $\$9,000$. $-1000(10) + 19000 = 9000$. Compared to $\$4,040$ for the 10-year-old car in the data set, the difference would be $\$4,960$. The prediction is off by about $\$5,000$, which seems like a lot of money, given the prices of the cars in the data set.

f. **Is $\$5,000$ typical of the differences between predicted prices and actual prices for the cars in this data set? Justify your answer.**

No, most of the differences would be much smaller than $\$5,000$. Most of the points are much closer to the line, and most predictions would be within about $\$1,000$ of the actual value.

Closing (2 minutes)

- When you use a line to describe a linear relationship in a data set, what are characteristics of a good fit?
  - *The line should be as close as possible to the points in the scatter plot. The line should go through the "middle" of the points.*

Lesson Summary

- **A line can be used to represent the trend in a scatter plot.**
- **Evaluating the equation of the line for a value of the independent variable will determine a value predicted by the line.**
- **A good line for prediction is one that goes through the middle of the points in a scatter plot and for which the points tend to fall close to the line.**

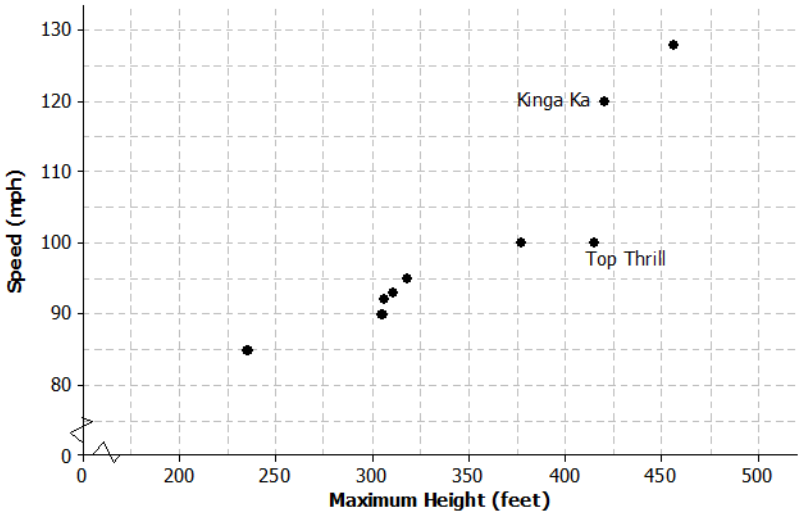
Exit Ticket (5 minutes)

Name _____ Date_____

Lesson 9: Determining the Equation of a Line Fit to Data

Exit Ticket

1. The scatter plot below shows the height and speed of some of the world's fastest roller coasters. Draw a line that you think is a good fit for the data.



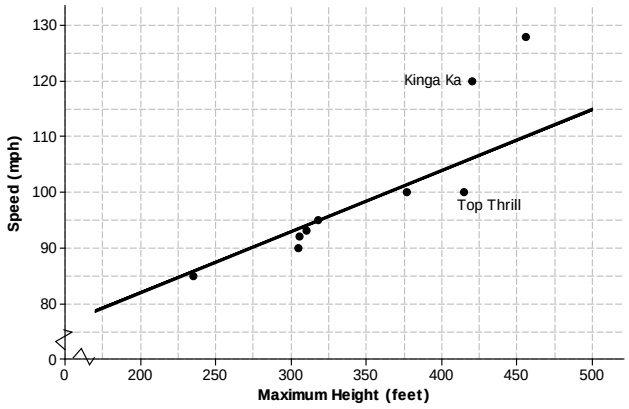
2. Find the equation of your line. Show your steps.

3. For the two roller coasters identified in the scatter plot, use the line to find the approximate difference between the observed speeds and the predicted speeds.

Exit Ticket Sample Solutions

1. The scatter plot below shows the height and speed of some of the world's fastest roller coasters. Draw a line that you think is a good fit for the data.

Students would draw a line based on the goal of a best fit for the given scatter plot. A possible line is drawn below.



2. Find the equation of your line. Show your steps.

Answers will vary based on the line drawn. Let S equal the speed of the roller coaster and H equal the maximum height of the roller coaster.

$$m = \frac{115 - 85}{500 - 225} \approx 0.11$$

$$S = 0.11H + b$$
$$85 = 0.11(225) + b$$
$$b \approx 60$$

Therefore, the equation of the line drawn in Problem 1 is $S = 0.11H + 60$.

3. For the two roller coasters identified in the scatter plot, use the line to find the approximate difference between the observed speeds and the predicted speeds.

Answers will vary depending on the line drawn by a student or the equation of the line. For the Top Thrill, the maximum height is about 415 feet and the speed about 100 miles per hour. The line indicated in Problem 2 predicts a speed of 106 miles per hour, so the difference is about 6 miles per hour over the actual speed. For the Kinga Ka, the maximum height is about 424 feet with a speed of 120 miles per hour. The line predicts a speed of about 107 miles per hour, for a difference of 13 miles per hour under the actual speed. (Students can use the graph or the equation to find the predicted speed.)

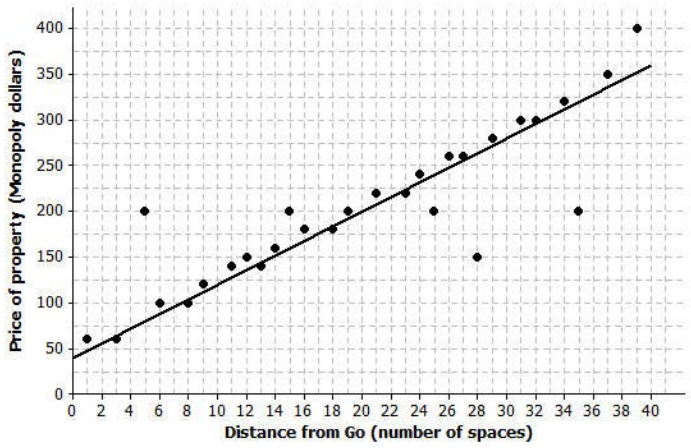
Problem Set Sample Solutions

1. Monopoly is a popular board game in many countries. The scatter plot below shows the distance from "Go" to a property (in number of spaces moving from "Go" in a clockwise direction) and the price of the properties on the Monopoly board. The equation of the line is $P = 8x + 40$, where P represents the price (in Monopoly dollars) and x represents the distance (in number of spaces).

Distance from "Go" (Number of Spaces)	Price of Property (Monopoly Dollars)
1	60
3	60
5	200
6	100
8	100
9	120
11	140
12	150
13	140
14	160
15	200
16	180
18	180
19	200

Distance from "Go" (Number of Spaces)	Price of Property (Monopoly Dollars)
21	220
23	220
24	240
25	200
26	260
27	260
28	150
29	280
31	300
32	300
34	320
35	200
37	350
39	400

Price of Property vs. Distance from "Go" in Monopoly



a. Use the equation to find the difference (observed value − predicted value) for the most expensive property and for the property that is 35 spaces from "Go."

The most expensive property is 39 spaces from "Go" and costs $\$400$. The price predicted by the line would be $8(39) + 40$, or $\$352$. Observed price − predicted price would be $\$400 - \$352 = \$48$. The price predicted for 35 spaces from "Go" would be $8(35) + 40 = \$320$. Observed price − predicted price would be $\$200 - \$320 = -\$120$.

b. Five of the points seem to lie in a horizontal line. What do these points have in common? What is the equation of the line containing those five points?

These points all have the same price. The equation of the horizontal line through those points would be Price $= \$200$.

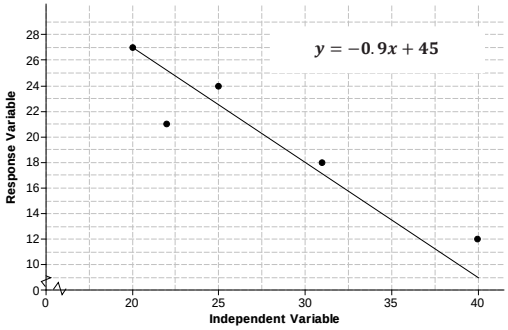
c. Four of the five points described in part (b) are the railroads. If you were fitting a line to predict price with distance from "Go," would you use those four points? Why or why not?

Answers will vary. Because the four points are not part of the overall trend in the price of the properties, I would not use them to determine a line that describes the relationship. I can show this by finding the total error to measure the fit of the line.

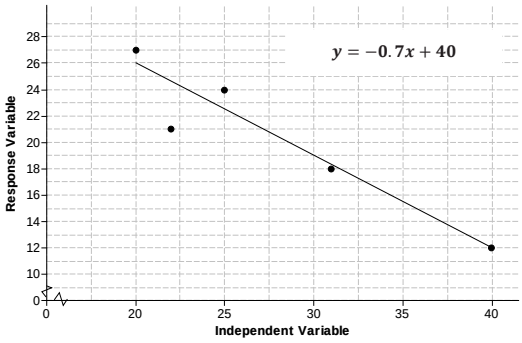
2. The table below gives the coordinates of the five points shown in the scatter plots that follow. The scatter plots show two different lines.

Data Point	Independent Variable	Response Variable
A	20	27
B	22	21
C	25	24
D	31	18
E	40	12

Line 1



Line 2



a. Find the predicted response values for each of the two lines.

Independent	*Observed Response*	*Response Predicted by Line 1*	*Response Predicted by Line 2*
20	27	27	26
22	21	25.2	24.6
25	24	22.5	22.5
31	18	17.1	18.3
40	12	9	12

b. For which data points is the prediction based on Line 1 closer to the actual value than the prediction based on Line 2?

Only for data point A. For data point C, both lines are off by the same amount.

c. Which line (Line 1 or Line 2) would you select as a better fit?

Line 2 because it is closer to more of the data points.

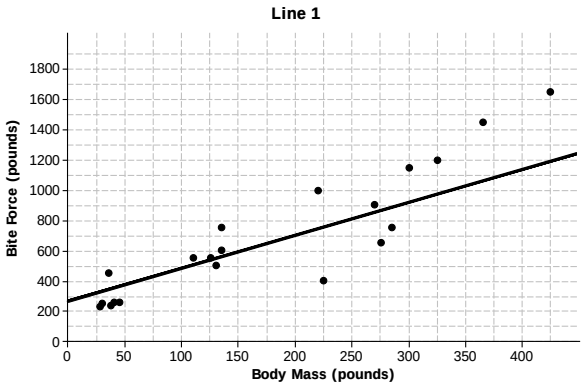
3. The scatter plots below show different lines that students used to model the relationship between body mass (in pounds) and bite force (in pounds) for crocodilian.

   a. Match each graph to one of the equations below and explain your reasoning. Let B represent bite force (in pounds) and W represent body mass (in pounds).

Equation 1	Equation 2	Equation 3
$B = 3.28W + 126$	$B = 3.04W + 351$	$B = 2.16W + 267$

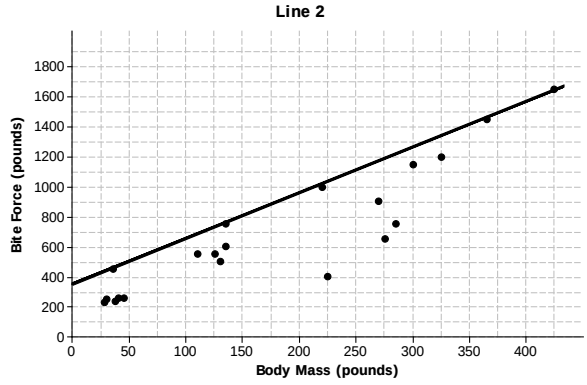
Equation: *3*

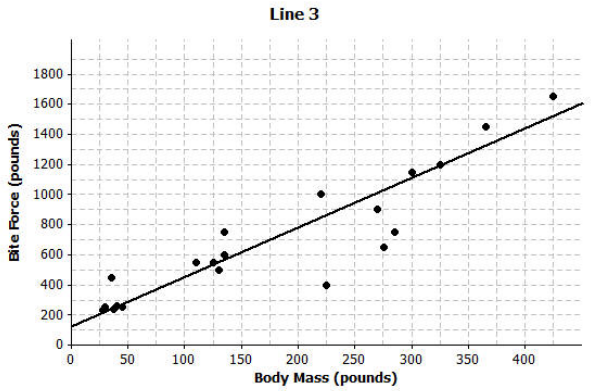
The intercept of 267 appears to match the graph, which has the second largest intercept.



Equation: *2*

The intercept of Equation 2 is larger, so it matches Line 2, which has a y-intercept closer to 400.



Equation: *1* *The intercept of Equation 1 is the smallest, which seems to match the graph.*	

b. **Which of the lines would best fit the trend in the data? Explain your thinking.**

Answers will vary. Line 3 would be better than the other two lines. Line 1 is not a good fit for larger weights, and Line 2 is above nearly all of the points and pretty far away from most of them. It looks like Line 3 would be closer to most of the points.

4. **Comment on the following statements:**

a. **A line modeling a trend in a scatter plot always goes through the origin.**

Some trend lines will go through the origin, but others may not. Often, the value $(0, 0)$ does not make sense for the data.

b. **If the response variable increases as the independent variable decreases, the slope of a line modeling the trend will be negative.**

If the trend is from the upper left to the lower right, the slope for the line will be negative because for each unit increase in the independent variable, the response will decrease.