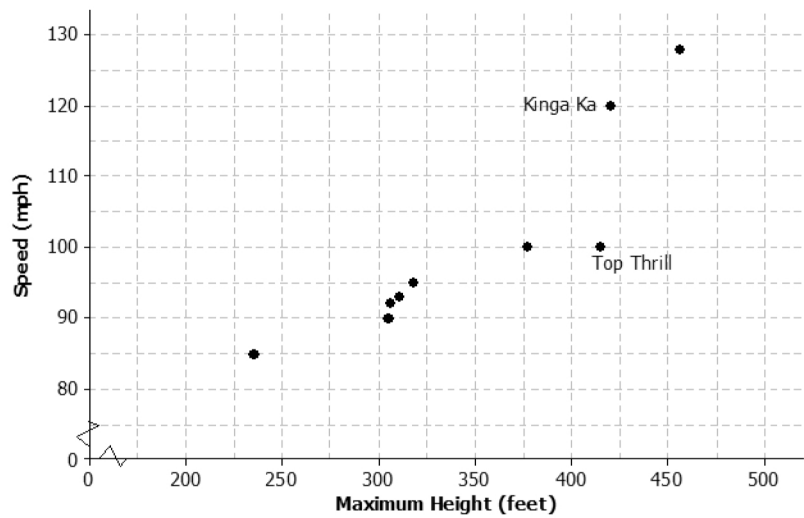Name _____   Date_____

Determining the Equation of a Line Fit to Data

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.
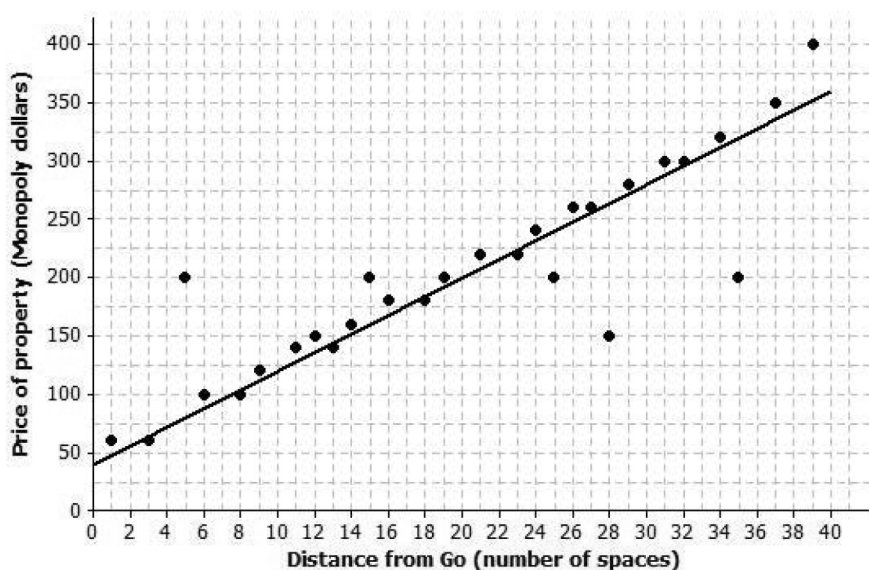

2. Find the equation of your line. Show your steps.

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.

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)	Distance from "Go" (number of spaces)	Price of Property (Monopoly dollars)
1	60	21	220
3	60	23	220
5	200	24	240
6	100	25	200
8	100	26	260
9	120	27	260
11	140	28	150
12	150	29	280
13	140	31	300
14	160	32	300
15	200	34	320
16	180	35	200
18	180	37	350
19	200	39	400

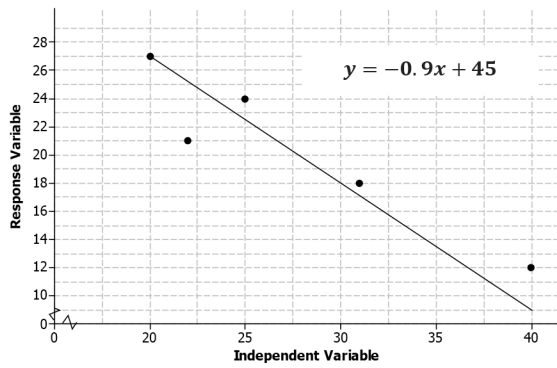**Price of Property vs. Distance from "Go" in Monopoly**

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."

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?

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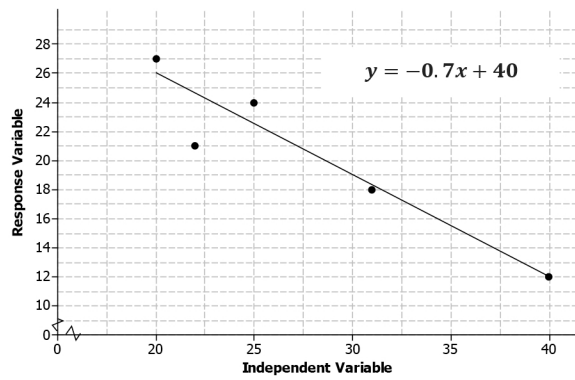
2. The table below gives the coordinates of the five points shown in the scatter plots that follow. The scatter plots show two different lines.

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

Line 1

$y = -0.9x + 45$

Line 2

$y = -0.7x + 40$

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

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

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

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).
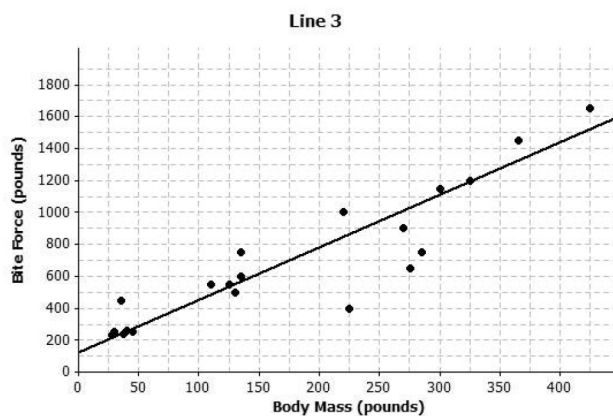
Equation 1
$B = 3.28W + 126$

Equation 2
$B = 3.04W + 351$

Equation 3
$B = 2.16W + 267$

Equation:	
Equation:	

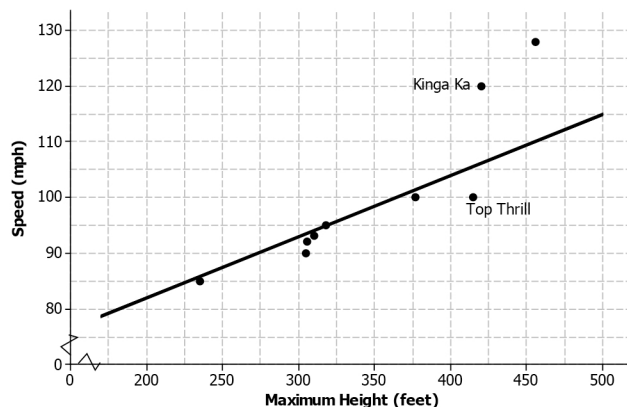| Equation:	

Line 3
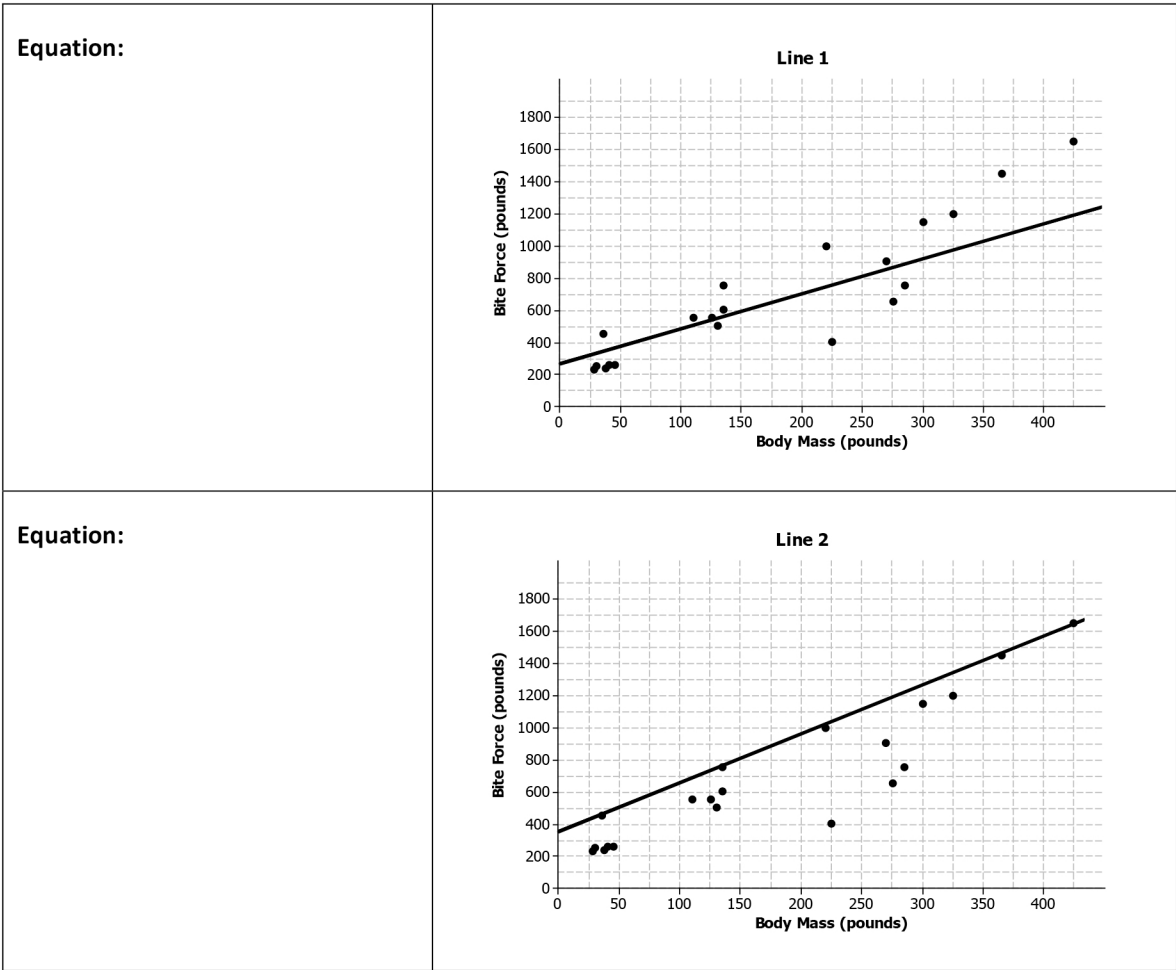
b. Which of the lines would best fit the trend in the data? Explain your thinking.

4. Comment on the following statements:

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

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

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.**
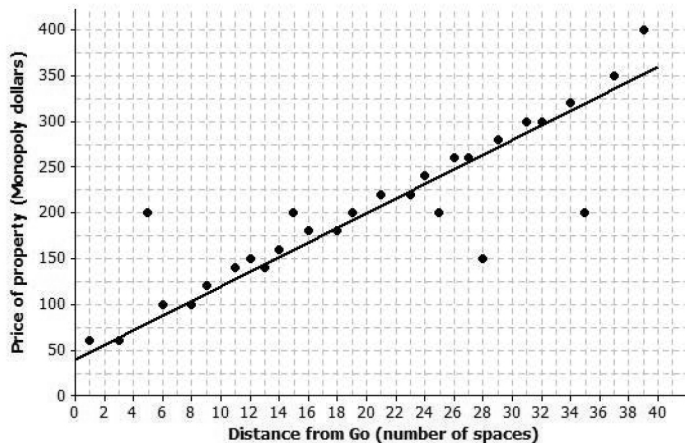
   Answers will vary depending on the line drawn by a student or the equation of the line. For the Top Thrill, the maximum height is about 415 feet and the speed about 100 miles per hour. The line indicated in Problem 2 predicts a speed of 106 miles per hour, so the difference is about 6 miles per hour 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.)

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)	Distance from "Go" (Number of Spaces)	Price of Property (Monopoly Dollars)
1	60	21	220
3	60	23	220
5	200	24	240
6	100	25	200
8	100	26	260
9	120	27	260
11	140	28	150
12	150	29	280
13	140	31	300
14	160	32	300
15	200	34	320
16	180	35	200
18	180	37	350
19	200	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.
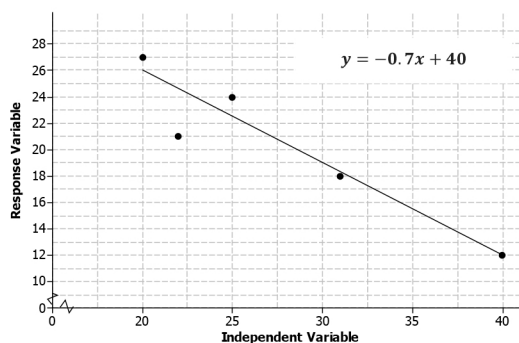
2.   The table below gives the coordinates of the five points shown in the scatter plots that follow. The scatter plots show two different lines.

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

Line 1

$$y = -0.9x + 45$$

Line 2

$$y = -0.7x + 40$$
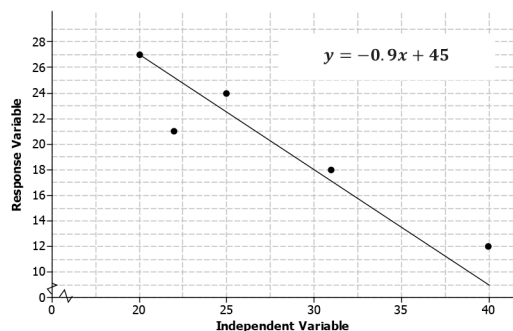
a.   Find the predicted response values for each of the two lines.

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).

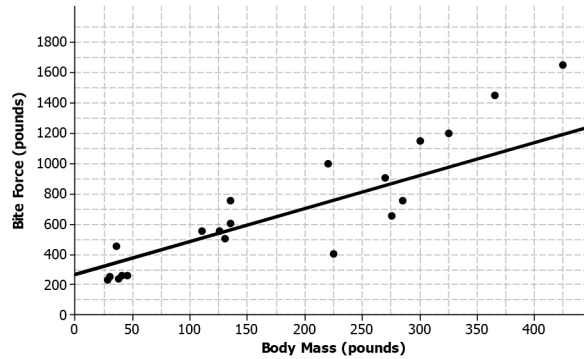| Equation 1	Equation 2	Equation 3
$B = 3.28W + 126$	$B = 3.04W + 351$	$B = 2.16W + 267$

Equation: *3*

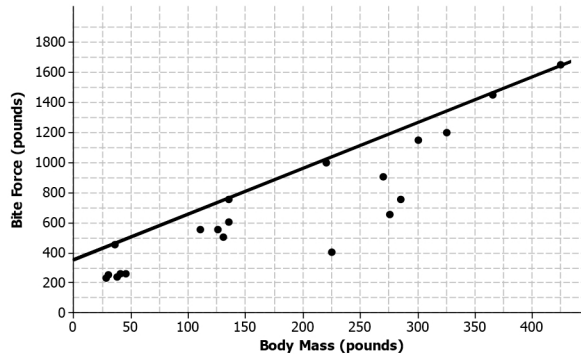*The intercept of 267 appears to match the graph, which has the second largest intercept.*

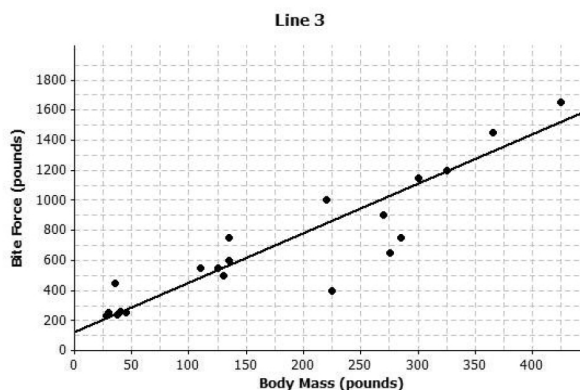Line 1

Equation: *2*

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

Line 2

Equation: *1*

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

Line 3

    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.