Lesson 8: Analyzing Bivariate Data

Goals

- Create a scatter plot and draw a line to fit bivariate data, and identify (orally and in writing) outliers that appear in the data.
- Interpret (orally and in writing) features of a scatter plot with a line of fit, including outliers, slope of the line, and clustering.

Learning Targets

- I can analyze a set of data to determine associations between two variables.

Lesson Narrative

In this lesson, students bring everything they have studied in the unit so far to analyze and interpret bivariate data in context (MP4). They create a scatter plot, identify outliers, fit a line, and determine and interpret the slope of the line. They compare actual and predicted values. They reflect on what they have learned about modeling bivariate data.

Alignments

Addressing

- 8.SP.A.1: Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
- 8.SP.A.2: Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
- 8.SP.A.3: Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.

Building Towards

- 8.SP.A.3: Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.

Instructional Routines

- MLR1: Stronger and Clearer Each Time

- MLR8: Discussion Supports
- Think Pair Share

Required Materials

Dried linguine pasta
We specified linguine since it is flatter and less likely to roll around than spaghetti.

Straightedges
A rigid edge that can be used for drawing line segments. Sometimes a ruler is okay to use as a straightedge, but sometimes it is preferable to use an unruled straightedge, like a blank index card.

Required Preparation

Class data (or data from another group if it fits the classroom culture better) from the second lesson in the unit on arm span and height.

Student Learning Goals

Let's analyze data like a pro.

8.1 Speed vs. Step Length

Warm Up: 5 minutes
The purpose of this warm-up is for students to interpret the rate of change of a function to describe a trend (MP2). It is used to help students recall the role of slope in showing associations for data that can be fit with a linear model in anticipation of looking for associations in this lesson.

Building Towards

- 8.SP.A.3

Launch

Give students 2 minutes of quiet work time followed by a whole-class discussion.

Student Task Statement

A researcher found an association between a dog's stride length and its speed: the longer a dog's steps, the faster it goes. The predicted speed in meters per second, s, as a function of step length in meters, l, is

$$s = 4l - 1.6$$

What does the rate of change of the function tell you about the association between stride length and speed?

Student Response

For every one meter increase in a dog's step length, its speed increases by 4 meters per second.

Activity Synthesis

Ask students to share what the rate of change of this function tells them about the trend. Record and display their responses for all to see. To include more students in the conversation, consider asking some of the following questions:

- “Does anyone agree or disagree with this reasoning? Why?”
- “Did anyone reason about the rate of change in a different way?”
- “Did anyone reason about the rate of change in the same way but would describe the trend differently?”
- “Does anyone want to add on to ____'s reasoning?”

8.2 Animal Brains

15 minutes (there is a digital version of this activity)
All of the information from this section about scatter plots comes into play as students analyze data about animal body and brain weights. Students begin with a table of data and create a scatter plot. After seeing the scatter plot, students pick out any outliers and fit a line to the scatter plot. Finally, the slope of the line is estimated and its meaning interpreted in context (MP2).

Addressing

- 8.SP.A.1
- 8.SP.A.2
- 8.SP.A.3

Instructional Routines

- MLR8: Discussion Supports
- Think Pair Share

Launch

Arrange students in groups of 2. Give students 5 minutes of quiet work time followed by 5 minutes of partner discussion and 5 minutes of whole-class discussion.

If using the digital activity, students can still work independently to analyze the scatter plot and answer the prompts. Then students can discuss their thinking in groups of 2 followed by whole-class discussion.

Anticipated Misconceptions

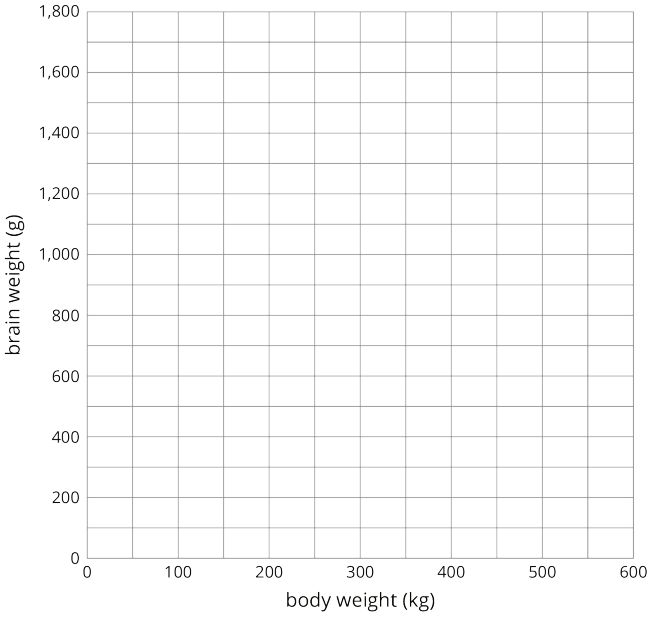
When estimating slope, some students won’t use the scales of the axes correctly, so the slope is reported incorrectly. Some students may not notice the different units of weight used on each axis.

Student Task Statement

Is there an association between the weight of an animal's body and the weight of the animal's brain?

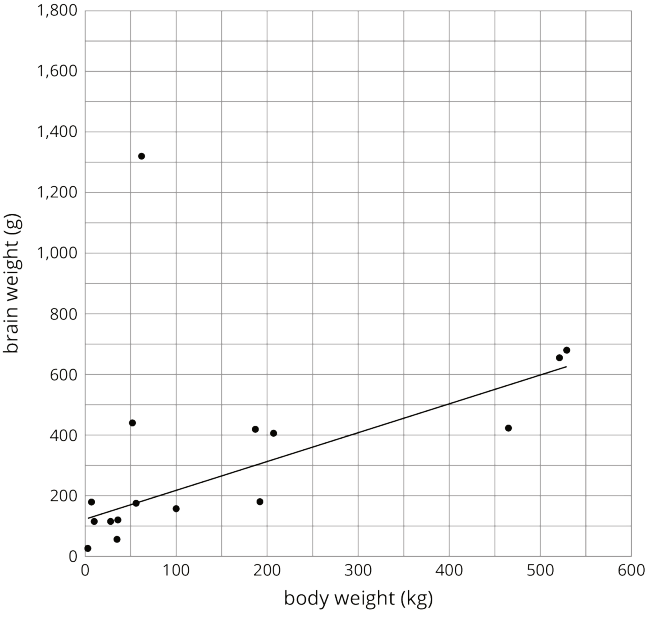
Use the data in the table to make a scatter plot. Are there any outliers?

animal	body weight (kg)	brain weight (g)
cow	465	423
grey wolf	36	120
goat	28	115
donkey	187	419
horse	521	655
potar monkey	10	115
cat	3	26
giraffe	529	680
gorilla	207	406
human	62	1,320
rhesus monkey	7	179
kangaroo	35	56
sheep	56	175
jaguar	100	157
chimpanzee	52	440
pig	192	180



1. After removing the outliers, does there appear to be an association between body weight and brain weight? Describe the association in a sentence.
2. Using a piece of pasta and a straightedge, fit a line to your scatter plot, and estimate its slope. What does this slope mean in the context of brain and body weight?
3. Does the fitted line help you identify more outliers?

Student Response



The human data is an outlier.

1. There seems to be a positive, linear association between body and brain weights.

2. The slope is about 1. This means that for every increase of 1 kilogram of body weight, the model predicts an increase of 1 gram of brain weight.
3. The chimpanzee data may also be an outlier.

Are You Ready for More?

Use one of the suggestions or find another set of data that interested you to look for associations between the variables.

- Number of wins vs number of points per game for your favorite sports team in different seasons
- Amount of money grossed vs critic rating for your favorite movies
- Price of a ticket vs stadium capacity for popular bands on tour

After you have collected the data,

1. Create a scatter plot for the data.
2. Are any of the points very far away from the rest of the data?
3. Would a linear model fit the data in your scatter plot? If so, draw it. If not, explain why a line would be a bad fit.
4. Is there an association between the two variables? Explain your reasoning.

Student Response

Answers vary.

Activity Synthesis

The goal of this discussion is to ensure students can make sense of the data given all the tools from this unit.

Consider asking some of the following questions:

- "Which data did you consider outliers?" (human and chimpanzee)
- "How did you determine your fitted line?"
- "Let's assume the trend you found continues past the end of the scatter plot. A Tyrannosaurus Rex is a dinosaur that is estimated to have a body weight of about 8,000 kg. What do you expect its brain weight to be?" (About 8,000 g or 8 kg)

Access for English Language Learners

Speaking: MLR8 Discussion Supports. Use this routine to support whole-class discussion. For each response or observation that is shared, ask students to restate and/or revoice what they heard using mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped to clarify the original statement.
Design Principle(s): Support sense-making

8.3 Equal Body Dimensions

15 minutes (there is a digital version of this activity)
In this activity students create another scatter plot to analyze the data they collected about their classmates in a previous lesson (MP4). A suggested linear model is compared to the data and a particular point is identified in both the scatter plot and data table.

Although the scatter plots are left to students to organize, the only linear model considered is $y = x$ which is symmetric when switching which variable is represented on each axis. If possible, identify any groups who have axes switched to bring up in the discussion.

Note: Some students may be sensitive about their body measurements and providing alternate data allows the class to work with actual values without making students uncomfortable. Depending on your class, consider providing a similar data set to the one collected in the earlier lesson (measurements from the staff, a different class, or invented data that is similar to the data collected).

Addressing

- 8.SP.A.1
- 8.SP.A.2

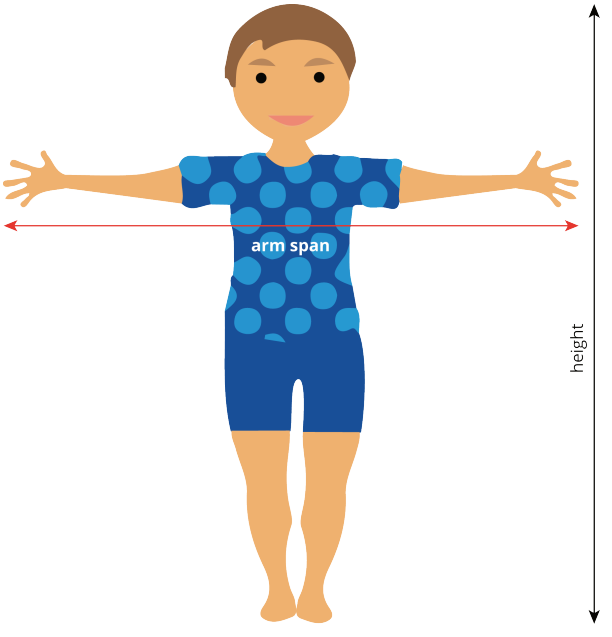
Instructional Routines

- MLR1: Stronger and Clearer Each Time

Launch

Give students 10 minutes of quiet work time followed by 5 minutes of whole-class discussion.

Display the class data collected earlier in the unit. Display this image to remind students what their data represents.



If using the digital activity, students will create their scatter plots using technology.

Access for Students with Disabilities

Action and Expression: Provide Access for Physical Action. Provide access to the digital version of the task or graphing software. If needed, provide students with a completed table of student data of arm span and height collected in an earlier lesson.
Supports accessibility for: Visual-spatial processing; Conceptual processing; Organization

Access for English Language Learners

Writing, Speaking, Listening: MLR1 Stronger and Clearer Each Time. Use this routine to give students a structured opportunity to revise and refine their response to: "Is the line $y = x$ a good fit for the data? If so, explain why. If not, find the equation of a better line." Ask each student to meet with 2–3 other partners in a row for feedback. Provide students with prompts for feedback that will help them strengthen their ideas and clarify their language (e.g., "Why do you think....?", "Can you give an example?", "How did you determine a better line?", etc.). Students can borrow ideas and language from each partner to strengthen their final version.
Design Principle(s): Optimize output (for generalization)

Student Task Statement

Earlier, your class gathered data on height and arm span.

1. Sometimes a person's arm span is the same as their height. Is this true for anyone in the class?

2. Make a scatter plot for the arm span and height data, and describe any association.

3. Is the line $y = x$ a good fit for the data? If so, explain why. If not, find the equation of a line that fits the data better.

4. Examine the scatter plot. Which person in your class has the *largest* ratio between their arm span and their height? Explain or show your reasoning.

Student Response

Answers vary based on class data.

1. Any data points on the line $y = x$ represent students who have equal arm spans and heights.
2. Graphs should have one axis labeled with height (cm) and one with arm span (cm). There is a positive, linear association between height and arm span, because as height increases, arm span tends to increase.
3. The line $y = x$ may be a good fit, because in general, human arm span and height are about the same.
4. The person with the largest ratio will have a data point that is farther from the line $y = x$ than the data points of the other students.

Activity Synthesis

The goal of this activity is for students to use the methods they have learned in this unit to explore data they have collected. To highlight some of the main points, select 2–3 students to respond to each question:

- “What is the slope of the line that best fit the data?” (Answers vary based on class data, but should be close to 1.)
- If any groups had axes switched, select these groups to show their scatter plots. "What would a point on the line $y = x$ represent in each graph?" (A person whose arm span and height are identical. It does not matter which way the axes are drawn for this line.)
- “What does it mean about a person whose point is above the line $y = x$? What does it mean about a person whose point is below the line?” (Depending on the axes, a person whose point is off of the line has a longer arm span than expected for their height or is taller than expected for their arm span.)
- “Suppose you measure height and arm span for people in you neighborhood. When you make a scatter plot, you notice two clusters of data: one group of data in the lower left of the scatter plot and another group in the upper right. What does this mean? Why do you think there may be these clusters?” (It may mean there are a lot of small children and taller adults.)

Lesson Synthesis

The goal of this discussion is to help students reflect on all of the things they have learned about bivariate data in this unit. Consider asking some of the following questions:

- “What does a point in a scatter plot tell you?” (Two measurements about an individual in a population).
- “What is an association between variables?” (A trend that suggests that as one variable increases, the other variable tends to increase if it is a positive association or decrease if it is a negative association.)
- “What does a fitted line tell you about the data?” (It represents a model that can be used to make predictions about the dependent variable based on the value of the independent variable.)
- “What does the slope of a fitted line tell you about the data?” (The amount the dependent variable will increase (or decrease) for a one-unit increase in the independent variable.)

8.4 Drawing a Line

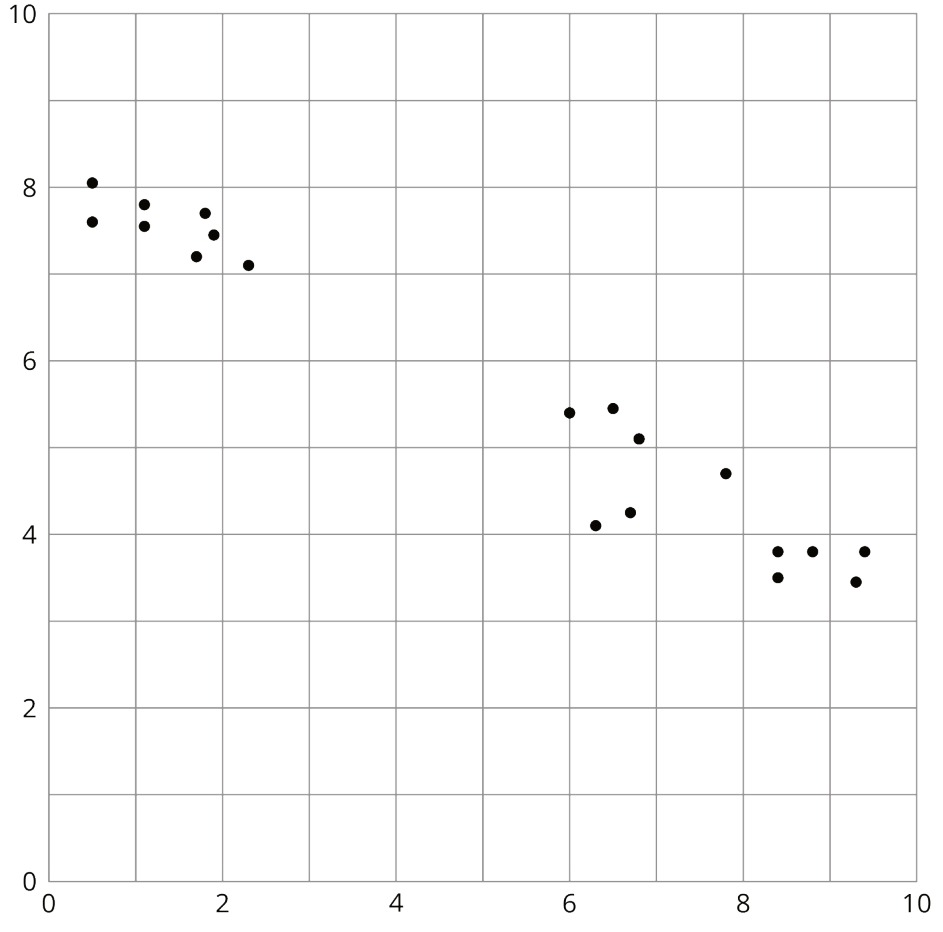
Cool Down: 5 minutes
Students use a given scatter plot to propose a good linear model for the data. They then use that model to predict the value of a new data point within the data.

Addressing

- 8.SP.A.2
- 8.SP.A.3

Student Task Statement

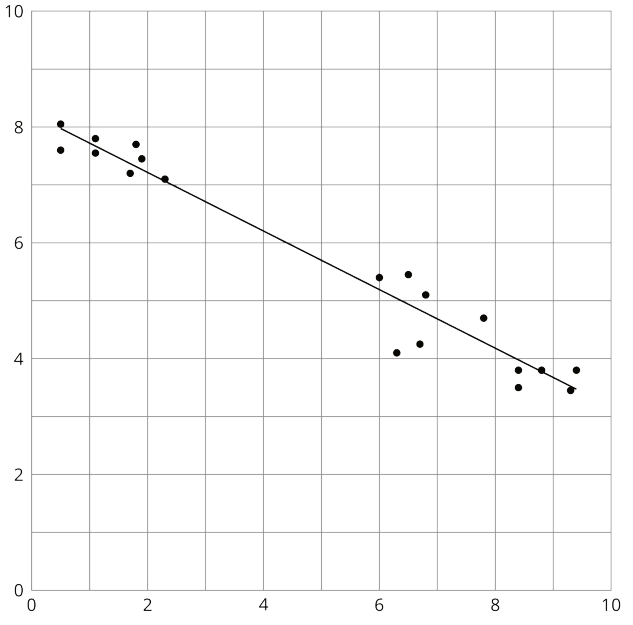
1. Draw a line on the scatter plot that fits the data well.



2. A new point will be added to the scatter plot with $x = 4$. What do you predict for the y-value of this point?

Student Response

1. Answers vary. Sample response:



2. Answers vary. Sample responses: 6, 6.1, 5.9

Student Lesson Summary

People often collect data in two variables to investigate possible associations between two numerical variables and use the connections that they find to predict more values of the variables. Data analysis usually follows these steps:

1. Collect data.
2. Organize and represent the data, and look for an association.
3. Identify any outliers and try to explain why these data points are exceptions to the trend that describes the association.
4. Find a function that fits the data well.

Although computational systems can help with data analysis by graphing the data, finding a function that might fit the data, and using that function to make predictions, it is important to understand the process and think about what is happening. A computational system may find a function that does not make sense or use a line when the situation suggests that a different model would be more appropriate.

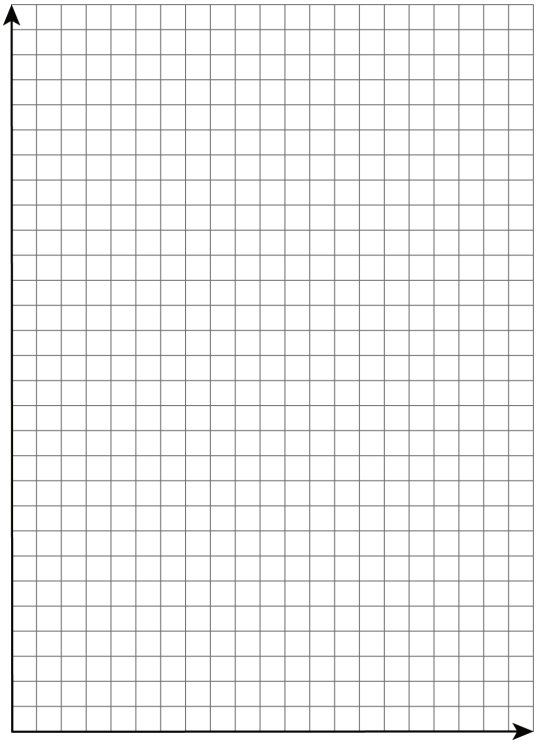
Lesson 8 Practice Problems

Problem 1

Statement

Different stores across the country sell a book for different prices. The table shows the price of the book in dollars and the number of books sold at that price.

price in dollars	number sold
11.25	53
10.50	60
12.10	30
8.45	81
9.25	70
9.75	80
7.25	120
12	37
9.99	130
7.99	100
8.75	90



a. Draw a scatter plot of this data. Label the axes.

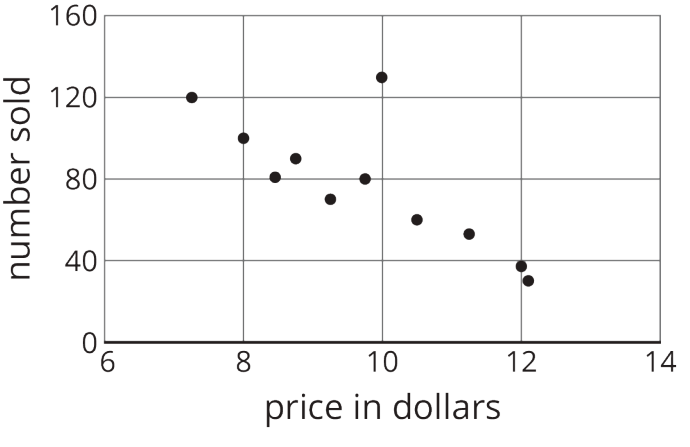
b. Are there any outliers? Explain your reasoning.

c. If there is a relationship between the variables, explain what it is.

d. Remove any outliers, and draw a line that you think is a good fit for the data.

Solution

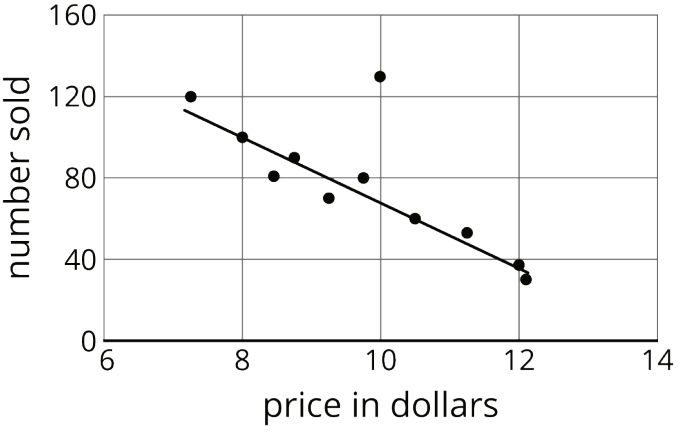
a.



b. Yes, at $(9.99, 130)$. This point is much higher than expected on the scatter plot.

c. There is a negative linear relationship between the variables. When the price increases, the number of books sold decreases.

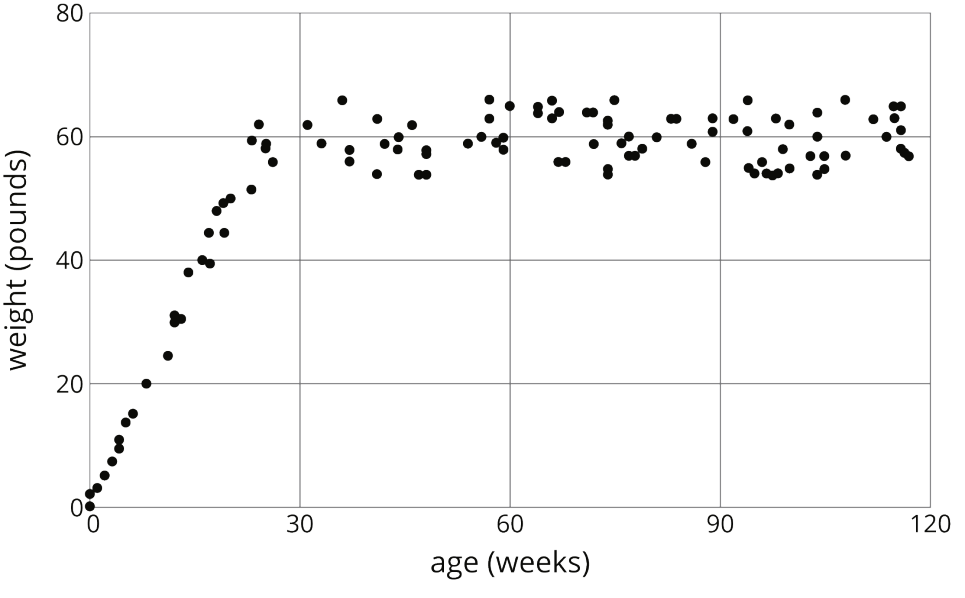
d. Answers vary. Sample response:



Problem 2

Statement

Here is a scatter plot:



Select **all** the following that describe the association in the scatter plot:

A. Linear association

B. Non-linear association

C. Positive association

D. Negative association

E. No association

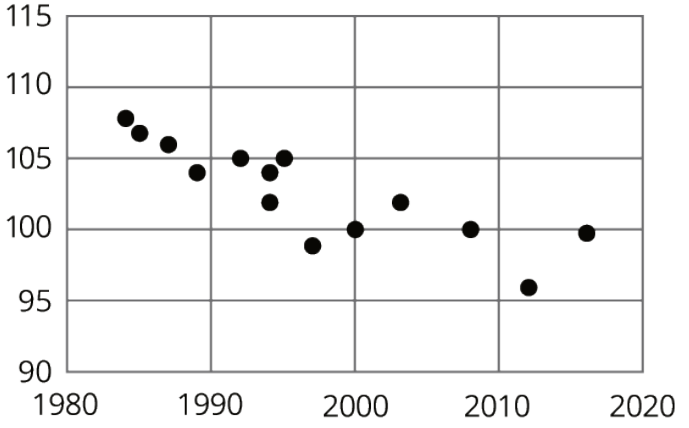
Solution

["B", "C"]
(From Unit 6, Lesson 7.)

Problem 3

Statement

Using the data in the scatter plot, what can you tell about the slope of a good model?



A. The slope is positive.

B. The slope is zero.

C. The slope is negative.

D. There is no association.

Solution

C
(From Unit 6, Lesson 6.)