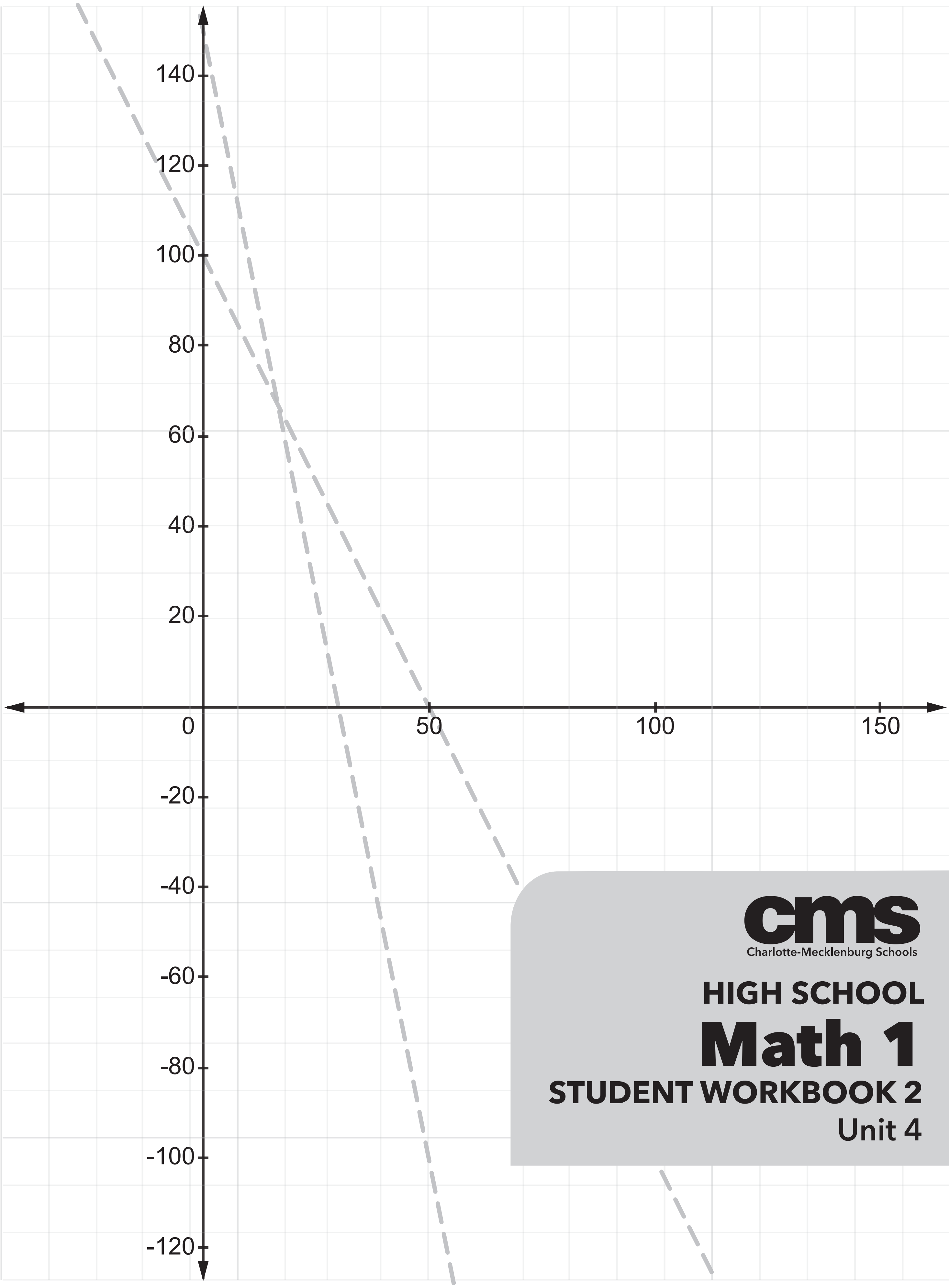


cms
Charlotte-Mecklenburg Schools

HIGH SCHOOL

Math 1

STUDENT WORKBOOK 2

Unit 4

Table of Contents

Unit 4 Linear Equations and Inequalities		
Scatter Plots and Correlation Coefficients	Lesson 1: Describing Scatter Plots	x
	Lesson 2: The Correlation Coefficient	x
	Lesson 3: Linear Models	x
	Lesson 4: Fitting Lines	x
Checkpoint	Lessons 5 & 6: Checkpoint	x
Evaluating Models	Lesson 7: Residuals	x
	Lesson 8: Causal Relationships	x
Mathematical Modeling	Lessons 9 & 10: Mathematical Modeling	x
Post-Test	Lesson 11: Post-Test Activities	x

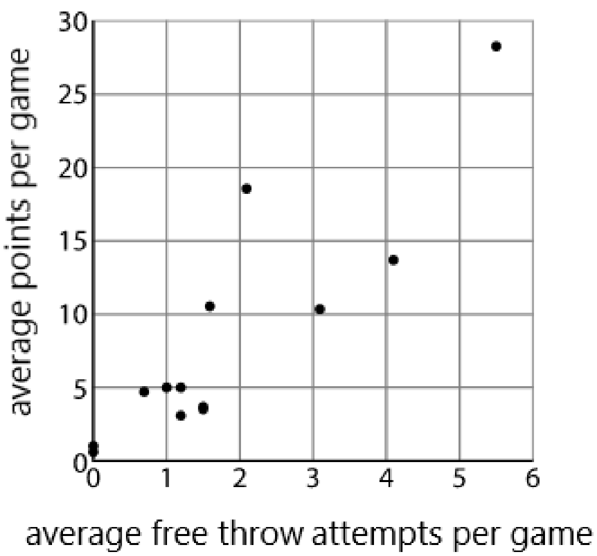
Lesson 1: Describing Scatter Plots

Learning Targets

- I can describe the shape and direction of the relationship between two variables.
- I can use technology to create a scatter plot to show the relationship between two variables.

Bridge

Here is a table and a scatter plot that compares points per game to free throw attempts for a basketball team during a week-long tournament.[1]



Player	Average free throw attempts	Average points per game
A	5.5	28.3
B	2.1	18.6
C	4.1	13.7
D	1.6	10.6
E	3.1	10.4
F	1	5
G	1.2	5
H	0.7	4.7
I	1.5	3.7
J	1.5	3.5
K	1.2	3.1
L	0	1
M	0	0.8
N	0	0.6

1. Circle the point that represents the data for Player E.
2. What does the point $(2.1, 18.6)$ represent?
3. In that same tournament, Player O on another team scored 14.3 points per game with 4.8 free throw attempts per game. Plot a point on the graph that shows this information.

[1] Adapted from IM 6–8 Math https://curriculum.illustrativemathematics.org/MS/index.html, which was originally developed by Open Up Resources and authored by Illustrative Mathematics, and is copyright 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). OUR's 6–8 Math Curriculum is available at https://openupresources.org/math-curriculum/. Adaptations and updates to IM 6–8 Math are copyright 2019 by Illustrative Mathematics, and are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Warm-up: Statistical Questions

1. Select **all** of the statistical questions.
   a. What is the current value of a 2,000-square-foot house in Charlotte, NC?
   b. How does daily attendance in school impact student achievement?
   c. How many counties are there in North Carolina?
   d. Does the neighborhood someone lives in determine their access to healthy foods?
2. Choose one of the statistical questions and describe what data would need to be collected to help answer the question.

Activity 1: Income and Food Access[2]

"Food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life" (World Food Summit, 1996).

Students at a high school in San Antonio, TX decided to explore the access to healthy food items available at grocery stores in different neighborhoods. They decided to collect data on the average household income for the neighborhood (defined by zip code) and the number of organic items available in the local grocery store.

Average household income	Number of organic vegetables offered
$71,186	36
$34,234	4
$71,186	28
$48,760	31
$78,096	78
$40,506	14
$38,166	12
$50,398	18
$49,437	38
$66,073	84
$86,566	61
$78,176	56
$59,154	62
$50,252	44
$48,364	26
$56,274	29
$41,318	15
$125,145	95
$65,911	18

Average household income	Number of organic vegetables offered
$50,252	65
$53,945	50
$59,072	35
$49,437	36
$72,080	28
$108,486	95
$70,530	46
$57,199	29
$78,176	73
$78,288	53
$86,566	86
$84,181	68
$84,181	56
$78,176	85
$84,181	86
$135,547	93
$92,946	82
$77,894	96

[2] Adapted from Skewthescript.org

1. Create a scatter plot to display the (*average household income, number of organic vegetables offered*).

    a. In www.desmos.com/calculator, click on the + icon on the top left of the window. Select the table option.

    b. Copy the data set provided at to the right from bit.ly/U4L1DataSet and paste into the first entry line in Desmos.

    c. As you enter, coordinate pairs will be plotted on the graph. If needed, adjust the graphing window to see the plotted points. Use the magnifying glass located below and to the left of the table to "zoom fit" the graph settings to the data.

Average household income	Number of organic vegetables offered
$71,186	36
$34,234	4
$71,186	28
$48,760	31
$78,096	78
$40,506	14
$38,166	12
$50,398	18
$49,437	38
$66,073	84
$86,566	61
$78,176	56
$59,154	62
$50,252	44
$48,364	26
$56,274	29
$41,318	15
$125,145	95
$65,911	18

Average household income	Number of organic vegetables offered
$50,252	65
$53,945	50
$59,072	35
$49,437	36
$72,080	28
$108,486	95
$70,530	46
$57,199	29
$78,176	73
$78,288	53
$86,566	86
$84,181	68
$84,181	56
$78,176	85
$84,181	86
$135,547	93
$92,946	82
$77,894	96

2. Describe the relationship between the (*average household income, number of organic vegetables offered*).

3. Why are these data important to understand? Do you think a similar trend exists in other cities? Why or why not?

Are You Ready For More?

Students in Charlotte, NC were interested in examining the access in their city. They collected the following data. In this case, they also collected the population within the neighborhood (defined by zip code).

Population	Median household income (2019)	Organic produce available
71048	65963	27
59664	93942	40
49635	59438	43
9280	136333	44
53629	51676	44
37286	91494	44
37309	45808	46
11315	88039	47

Population	Median household income (2019)	Organic produce available
11195	92786	55
43931	52766	55
42263	71914	55
19283	93938	56
28523	90057	57
20317	76022	58
47208	49465	59

1. Create a scatter plot for the (*median household income, organic produce available*) and describe any relationship between the two variables.

2. Compare this relationship to the one you found for San Antonio. What do you think are the reasons for any similarities or differences?

3. Create a scatter plot for the (*population, organic produce available*) and describe any relationship between the two variables.

4. One of the points appears to be an outlier. How does your answer to question 3 change if the outlier is removed?

Activity 2: Shape and Direction

Your teacher will give you a set of cards, each with a different scatter plot. With a partner, complete each of the following:

1. Sort the cards based on the direction of the data displayed in the scatter plot. Record the card(s) for each group.

   - Positive direction: ____
   - Negative direction:____
   - No Relationship: ____

2. Sort the cards based on the shape of the data. Does the scatter plot have a linear or a nonlinear shape? Record the card(s) for each group.

   - Linear shape: ____
   - Nonlinear shape: ____

3. Take turns with your partner. Select one of the cards from the scatter plots. Describe the relationship between the two variables to your partner. For each description your partner shares, listen carefully to their explanation. If you disagree, discuss your thinking and work together to reach an agreement.

Lesson Debrief

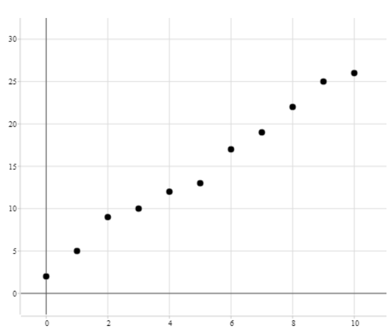
Lesson 1 Summary and Glossary

A scatter plot is used to display two numerical variables. Scatter plots are helpful when determining the relationship between the two variables. This relationship can be described in terms of the shape and the direction.

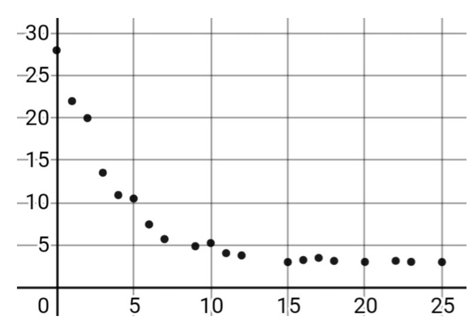
The shape of data in a scatter plot can be described as **linear** or **nonlinear**.

Linear relationship: A relationship between two numerical variables in which the scatter plot displaying the data resembles a line.

Nonlinear relationship: A relationship between two numerical variables in which the scatter plot displaying the data has a pattern other than a line, usually curved in some way.

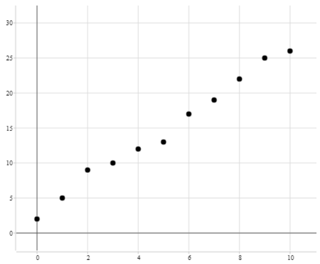


The data points in this scatter plot show a **linear** relationship. The data closely resemble a straight line.

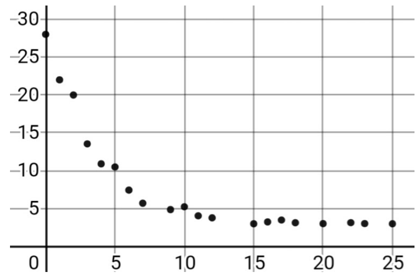


The data points in this scatter plot show a **nonlinear** relationship. The data do not resemble a straight line.

The direction of the data in a scatter plot can be described as **positive** or **negative**.



The data points in this scatter plot show a **positive** relationship.
As the x variable increases, the y variable increases.



The data points in this scatter plot show a **negative** relationship. As the x variable increases, the y variable decreases.

Positive relationship: A relationship between two numerical variables in which one variable tends to be paired with an increase in the other variable.

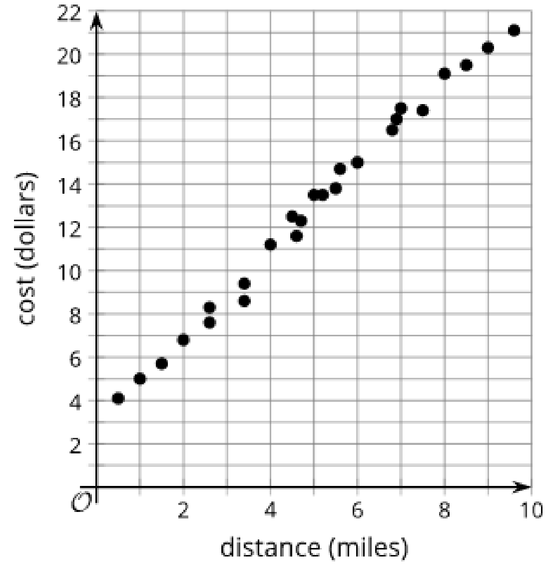
Negative relationship: A relationship between two numerical variables in which one variable tends to be paired with a decrease in the other variable.

The shape and direction of the data as shown in the scatter plot help when describing the relationship.

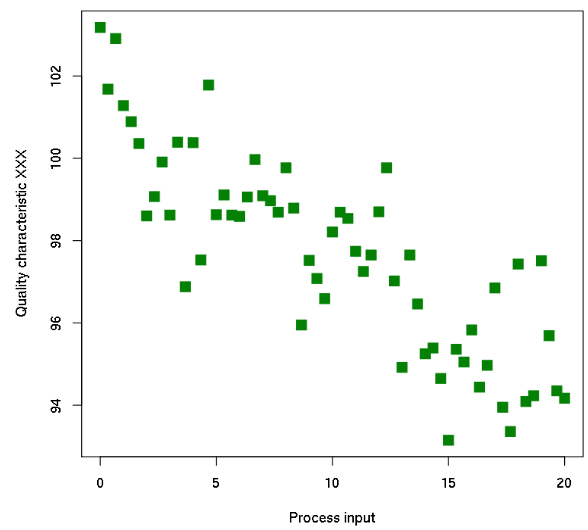
The scatter plot on the left displays the relationship between blood alcohol level and reaction time. There is a **positive, linear** relationship. This means that as the blood alcohol level increases, the reaction time increases.

Unit 4 Lesson 1 Practice Problems

1. The scatter plot shows the relationship between *(distance, cost)*. Describe the shape and direction of the relationship.



2. Here is a scatter plot:



   Select **all** of the following that describe the relationship in the scatter plot:

   a. Linear relationship

   b. Non-linear relationship

   c. Positive relationship

   d. Negative relationship

   e. No relationship

3. (*Technology required*.) The table includes data collected on popcorn sales at a local carnival.

Popcorn price (dollars)	0.50	0.60	0.8	1	1.25	1.50	1.75
Average number of cups sold	150	136	125	120	105	80	75

   a. Use technology to create a scatter plot of the data.

   b. Describe the relationship between the price of the popcorn and the average number of cups sold.

4. (*Technology required.*) The Census Bureau provided the following statistics for the years from 2006–2011.[3]

Year	Median income for all men
2011	$37,653
2010	$38,014
2009	$38,558
2008	$49,134
2007	$41,033
2006	$41,103

Year	Median income for all women
2011	$23,395
2010	$23,657
2009	$24,284
2008	$23,967
2007	$25,005
2006	$24,429

a. Use technology to create one scatter plot for the median income for all men from 2006–2011 and one scatter plot for the median income for all women from 2006–2011. Describe any relationship between the year and the income for each scatter plot.

b. Compare this relationship of median incomes for men and women. What do you think are the reasons for any similarities or differences?

[3] Adapted from Secondary Math 1 Mathematics Vision project http://www.mathematicsvisionproject.org, licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0)

5. (*Technology required*.) Different stores across the country sell the same book for different prices. The table shows the price of a particular book in dollars and the number of books sold at that price.[4]

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. Create a scatter plot using technology

b. Identify the shape and direction of the relationship.

c. Describe the relationship between price charged for the book and the number sold.

d. Is there an outlier? What point?

How is the relationship affected if the outlier is removed?

[4] Adapted from IM 6–8 Math https://curriculum.illustrativemathematics.org/MS/index.html, which was originally developed by Open Up Resources and authored by Illustrative Mathematics, and is copyright 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). OUR's 6–8 Math Curriculum is available at https://openupresources.org/math-curriculum/. Adaptations and updates to IM 6–8 Math are copyright 2019 by Illustrative Mathematics, and are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

6. The robotics team needs to purchase $350 in new equipment. Each student on the team plans to fundraise and contribute equally to the purchase.

   If X represents the total number of students on the team, which expression represents the amount each student needs to contribute?

   Which expression represents the amount each student needs to fundraise?

   a. $\$350 - X$

   b. $\$350 + X$

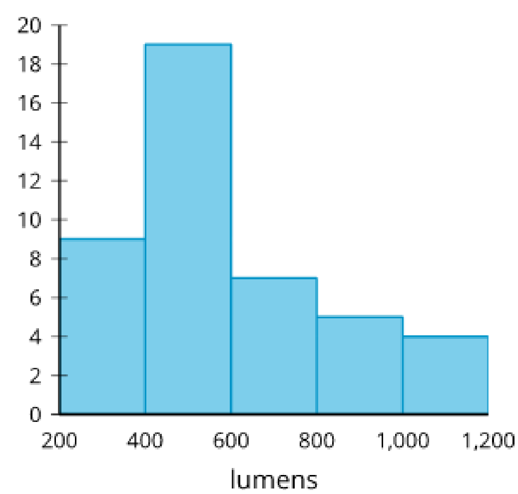
   c. $\frac{\$350}{X}$

   d. $\$350 \cdot X$

Amount each student needs to contribute	Amount each student needs to fundraise

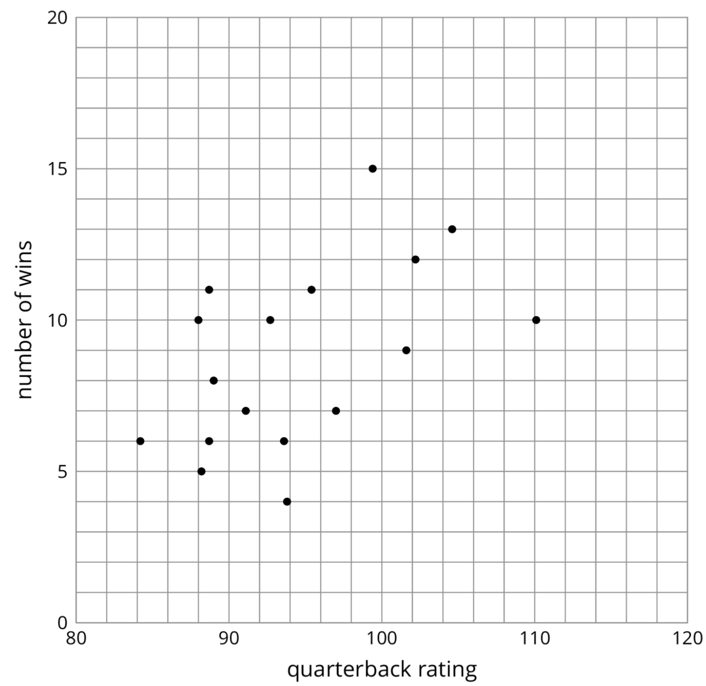
(From Unit 2)

7. Describe the shape of the distribution shown in the histogram that displays the light output, in lumens, of various light sources.



(From Unit 1)

8. Here is a table and scatter plot that show ratings and wins for quarterbacks who started 16 games this season.[5]



a. Circle the point in the scatter plot that represents Player K's data.

b. What does the point $(88, 10)$ represent?

c. Player R is not included in the table because he did not start 16 games this year. He did have a quarterback rating of 99.4, and his team won eight games. On the scatter plot, plot a point that represents Player R's data.

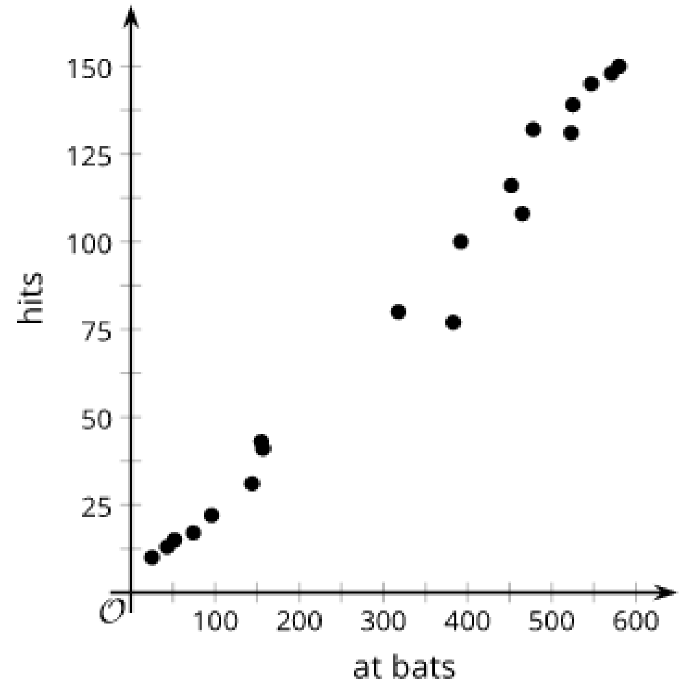
Player	Quarterback rating	Number of wins
A	93.8	4
B	102.2	12
C	93.6	6
D	89	8
E	88.2	5
F	97	7
G	88.7	6
H	91.1	7
I	92.7	10
J	88	10
K	101.6	9
L	104.6	13
M	84.2	6
N	99.4	15
O	110.1	10
P	95.4	11
Q	88.7	11

(Addressing NC.8.SP.1)

[5] Adapted from IM 6–8 Math https://curriculum.illustrativemathematics.org/MS/index.html, which was originally developed by Open Up Resources and authored by Illustrative Mathematics, and is copyright 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). OUR's 6–8 Math Curriculum is available at https://openupresources.org/math-curriculum/. Adaptations and updates to IM 6–8 Math are copyright 2019 by Illustrative Mathematics, and are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

9. The scatter plot shows the number of times a player came to bat and the number of hits they had. The scatter plot includes a point at $(318, 80)$.

   Describe the meaning of this point in this situation.



(Addressing NC.8.SP.1)

Lesson 2: The Correlation Coefficient

Learning Target

- I can describe the strength of a relationship between two variables.

Bridge

Mentally order the numbers from least to greatest.

a. 20.2, 18.2, 19.2

b. -14.6, -16.7, -15.1

c. -0.43, -0.87, -0.66

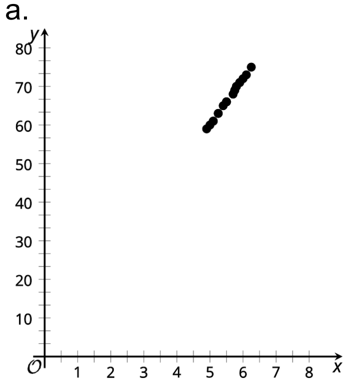
d. 0.50, -0.52, 0.05

Warm-up: Putting the Numbers in Context

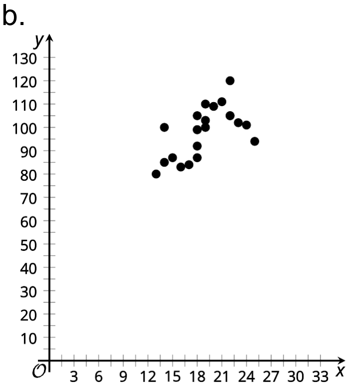
Match the variables to the scatter plot you think they best fit. Be prepared to explain your reasoning.

	x variable	y variable
1.	daily low temperature in Celsius for Denver, CO	boxes of cereal in stock at a grocery in Miami, FL
2.	average number of free throws shot in a season	basketball team score per game
3.	measured student height in feet	measured student height in inches
4.	average number of minutes spent in a waiting room	hospital satisfaction rating

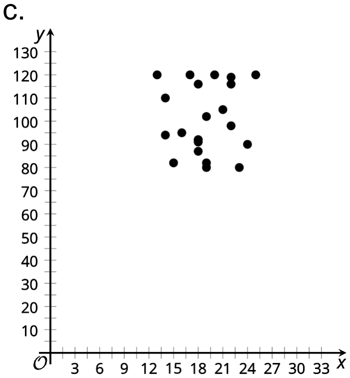
a.



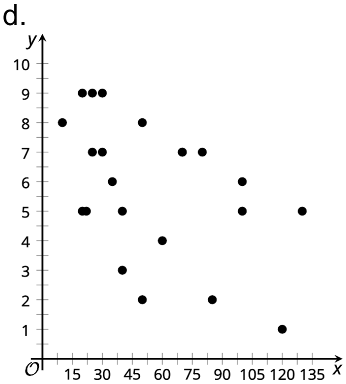
b.



c.



d.



Activity 1: Scatter Plot Fit

Your teacher will give you a set of cards that show scatter plots of data.

1. Sort the scatter plots by direction.

Scatter plots with a negative direction	Scatter plots without direction	Scatter plots with a positive direction

2. Using the diagram below, organize the scatter plots based on the strength of the linear shape.

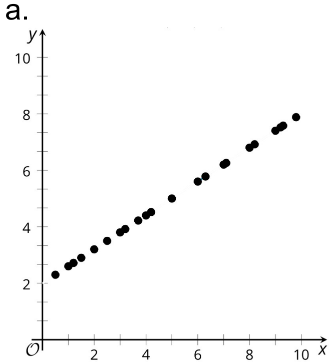
strongest linear shape negative direction	*weakest linear shape without direction*	*strongest linear shape positive direction*

Activity 2: Matching Correlation Coefficients

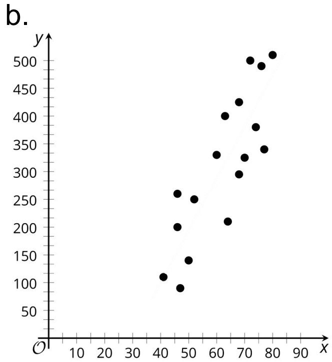
1. Take turns with your partner to match a scatter plot with a correlation coefficient.
2. For each match you find, explain to your partner how you know it's a match.
3. For each match your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.

r = -1	r = -0.95	r = -0.74	r = -0.06	r = 0.48	r = 0.65	r = 0.9	r = 1

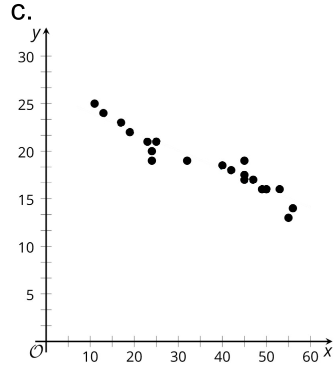
a.



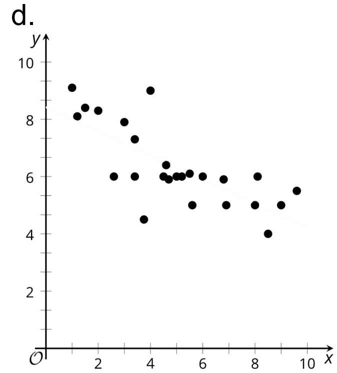
b.



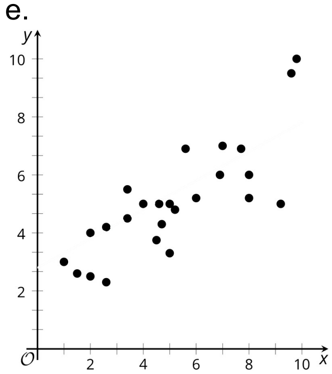
c.



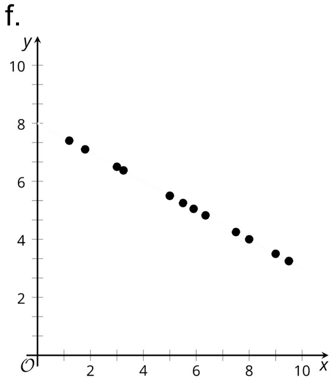
d.



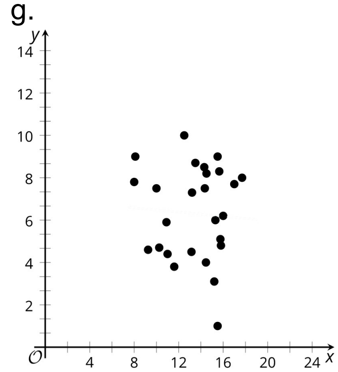
e.



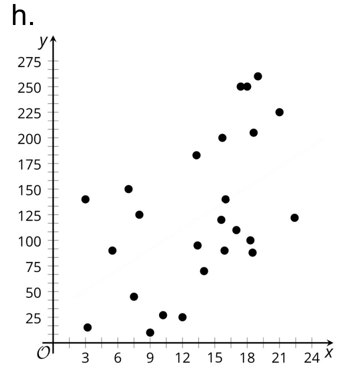
f.



g.



h.



Activity 3: Copious Correlations

Example response:

Cost of a package of light bulbs and number of lightbulbs in the package. $r = 0.96$

Sample Response: The cost of a package of light bulbs and the number of light bulbs in the package has a correlation coefficient value near 1. This means that these variables have a very strong, positive linear relationship. In other words, the price of the package is very closely related to the number of light bulbs in the package (strong relationship) and when one of the variables goes up, the other variable tends to go up, too (positive relationship).

For each situation, describe the relationship between the variables, based on the correlation coefficient. Make sure to mention whether there is a strong relationship or weak relationship as well as whether it is a positive relationship or negative relationship.

1. Number of steps taken per day and number of kilometers walked per day. $r = 0.92$

2. Temperature of a rubber band and distance the rubber band can stretch. $r = 0.84$

3. Car weight and distance traveled using a full tank of gas. $r = -0.86$

4. Average fat intake per citizen of a country and average cancer rate of a country. $r = 0.73$

5. Score on science exam and number of words written on the essay question. $r = 0.28$

6. Average time spent listening to music per day and average time spent watching TV per day. $r = -0.17$

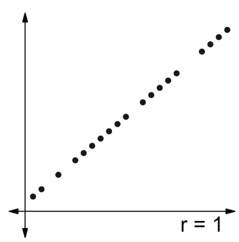
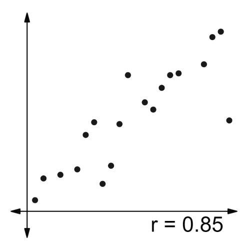
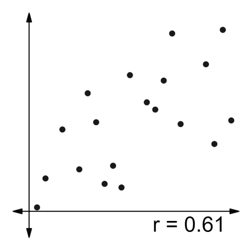
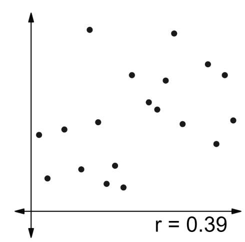
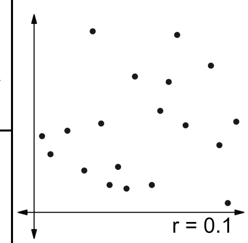
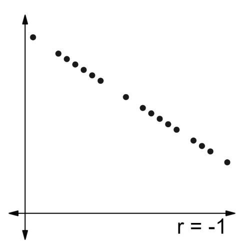
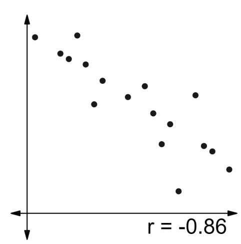
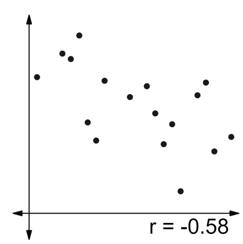
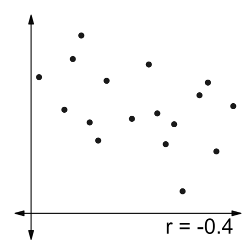
Lesson Debrief

Lesson 2 Summary and Glossary

While scatter plots can be used to identify the shape and direction of a set of data, we need a way to determine the strength of a linear relationship.

The **correlation coefficient** is a number that can be used to describe the strength and direction of a linear relationship. Usually represented by the letter r, the correlation coefficient can take values from -1 to 1. The sign of the correlation coefficient indicates the direction of the relationship. The closer the correlation coefficient is to 1 or -1, the stronger the linear relationship. The closer the correlation coefficient is to 0, the weaker the linear relationship.

Correlation coefficient: A number between -1 and 1 that describes the strength and direction of a linear relationship between two numerical variables. The sign of the correlation coefficient indicates the direction of the relationship.The closer the correlation coefficient is to 1 or -1, the stronger the linear relationship. The closer the correlation coefficient is to 0, the weaker the linear relationship.

	Perfect correlation	Strong correlation	Moderate correlation	Weak correlation	No correlation
Positive	 r = 1	 r = 0.85	 r = 0.61	 r = 0.39	 r = 0.1
Negative	 r = -1	 r = -0.86	 r = -0.58	 r = -0.4	

A correlation coefficient with a value near 1 suggests a strong, positive linear relationship between the variables. This means that most of the data tend to be tightly clustered into the shape of a line, and that when one of the variables increases in value, the other does as well. The number of schools in a community and the population of the community are examples of variables that have a strong, positive correlation. When there is a large population, there is usually a large number of schools, and small communities tend to have fewer schools, so the correlation is positive. These variables are closely tied together, so the correlation is strong.

Similarly, a correlation coefficient near -1 suggests a strong, negative relationship between the variables. Again, most of the data tend to be tightly clustered into the shape of a line, but now, when one value increases, the other decreases. The time since you left home and the distance left to reach school has a strong, negative correlation. As the travel time increases, the distance to school tends to decrease, so this is a negative correlation. The variables are again closely, linearly related, so this is a strong correlation.

Weaker correlations mean there may be other reasons the data are variable other than the connection between the two variables. For example, the number of pets and number of siblings has a weak correlation. There may be some relationship, but there are many other factors that account for the variability in the number of pets other than the number of siblings.

The context of the situation should be considered when determining whether the correlation value is strong or weak. In physics, measuring with precise instruments, a correlation coefficient of 0.8 may not be considered strong. In social sciences, collecting data through surveys, a correlation coefficient of 0.8 may be very strong.

Unit 4 Lesson 2 Practice Problems

1. The correlation coefficient, r, is given for several different data sets. Which value for r indicates the **strongest** linear relationship?

   a. 0.01

   b. -0.34

   c. -0.82

   d. -0.95

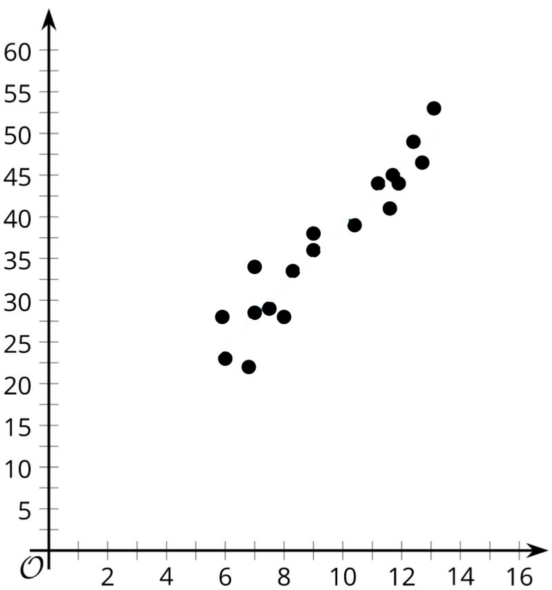
2. Which of the values is the **best** estimate of the correlation coefficient for the data shown in the scatter plot?

   a. -0.9

   b. -0.4

   c. 0.4

   d. 0.9



3. The number of hours worked, x, and the total dollars earned, y, have a strong positive relationship.

   Explain what it means to have a strong positive relationship in this situation.

4. The number of minutes on the phone and the customer satisfaction rating have a weak negative relationship.

   Explain what it means to have a **weak negative** relationship in this context.

5. The correlation coefficient, r, is given for several different data sets. Which value for r indicates the **weakest** linear relationship?

   a. 0.01

   b. 0.5

   c. -0.99

   d. 1

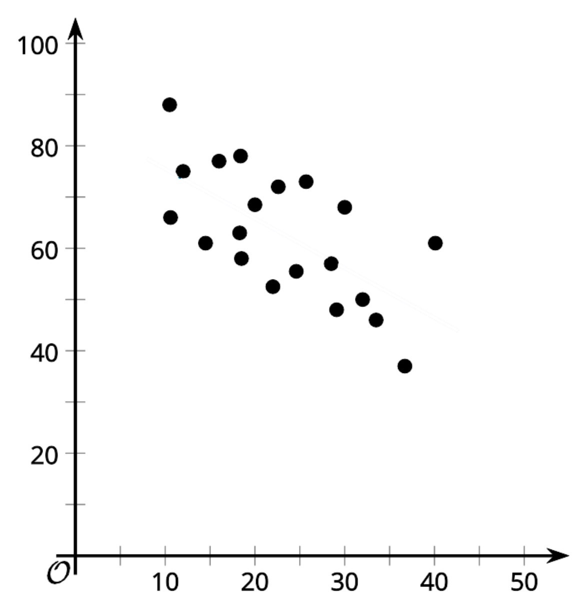
6. Which of the following is the **best** estimate of the correlation coefficient for the relationship shown in the scatter plot?

   a. -0.9

   b. -0.4

   c. 0.4

   d. 0.9



7. (*Technology required.*)

    a. Use the data in the table to make a scatter plot.[1] Data can be accessed through this spreadsheet: https://bit.ly/U4L2DataSet.

    b. Describe the relationship between the variables.

    c. What does the point $(62, 1320)$ represent?

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	1320
Rhesus monkey	7	179
Kangaroo	35	56
Sheep	56	175
Jaguar	100	157
Chimpanzee	52	440
Pig	192	180

(From Unit 4, Lesson 1)

8. For a baseball team fundraiser, Noah washed 16 cars and earned $213 for the team! He charged $18 per car wash.

    a. Write a linear equation in point-slope form representing the amount of money, y, a player could raise for washing x cars.

    b. Write a linear equation in slope-intercept form representing the amount of money, y, a player could raise for washing x cars.

    c. What could the y-intercept of the equation represent?

(From Unit 3)

[1] Adapted from IM 6–8 Math https://curriculum.illustrativemathematics.org/MS/index.html, which was originally developed by Open Up Resources and authored by Illustrative Mathematics, and is copyright 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). OUR's 6–8 Math Curriculum is available at https://openupresources.org/math-curriculum/. Adaptations and updates to IM 6–8 Math are copyright 2019 by Illustrative Mathematics, and are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

9. Solve the equation: $\frac{3}{5}x - 6 = \frac{1}{3}x + 4$

(From Unit 2)

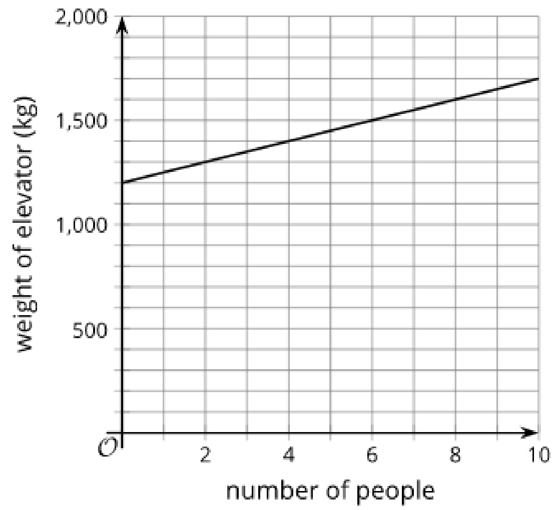
Lesson 3: Linear Models

Learning Targets

- I can assess if a linear model fits the data well and use the linear model to estimate values I want to find.
- I can describe the rate of change and vertical intercept (y -intercept) for a linear model in everyday language.

Bridge

Here is a linear model of the weight of an elevator and the number of people on the elevator.

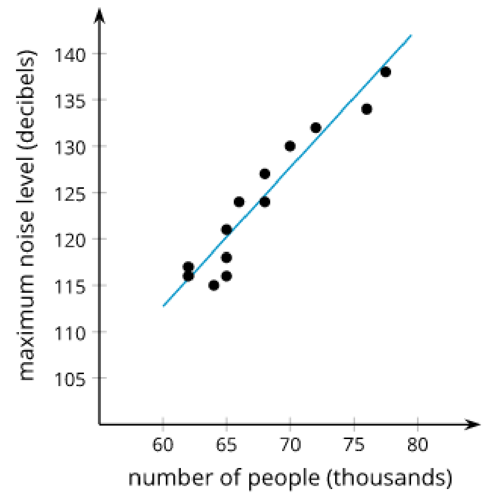


1. Find these values. Explain your reasoning.
    a. the weight of the elevator when 6 people are on it
    b. the number of people on the elevator when it weighs 1,400 kg
    c. the weight of the elevator when no people are on it
    d. the increase in elevator weight for each additional person according to the model
2. Which of your answers corresponds to the slope of the line in the graph?
3. Which of your answers corresponds to the y-intercept of the line in the graph?

Warm-up: Crowd Noise

What do you notice? What do you wonder?

$$y = 1.5x + 22.7$$



Activity 1: Orange You Glad We're Boxing Fruit

1. Watch the video and record the weight for the number of oranges in the box.

Number of oranges	3	4	5	6	7	8	9	10
Weight in kilograms								

2. Use technology to create a scatter plot of the data.

   a. Describe the relationship between the number of oranges and weight in kilograms.

   b. Which of the following is most likely the correlation coefficient of the data? Explain your reasoning.

   -0.92   -0.42   0.1   0.9

3. The linear equation $y = 0.216x + 0.345$ is suggested as a linear model for the data. Use technology to graph the line along with the scatter plot.

    a. How well does the line model the data? Explain your reasoning.

    b. What is the value of the slope, and what does it mean in this situation?

    c. What is the value of the y-intercept and what does it mean in this situation?

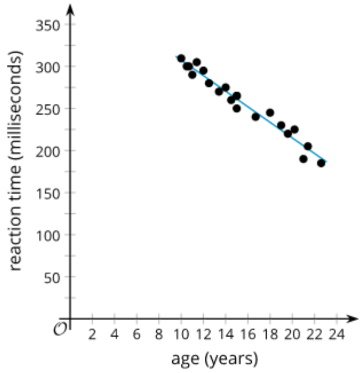
4. Use the line to predict the weight of a box containing 11 oranges. Will this estimate be close to the actual value? Explain your reasoning.

5. Use the line to predict the weight of a box containing 50 oranges. Will this estimate be close to the actual value? Explain your reasoning.

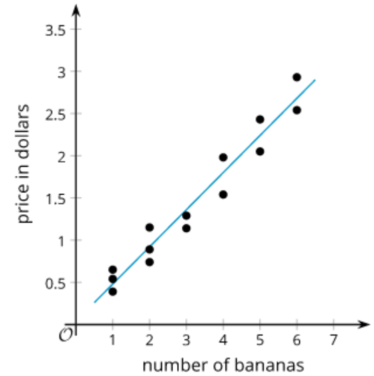
Activity 2: The Slope Is the Thing

Here are several scatter plots. The linear models given here are lines of best fit.

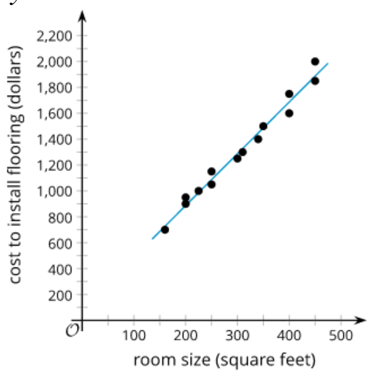
a. $y = -9.25x + 400$



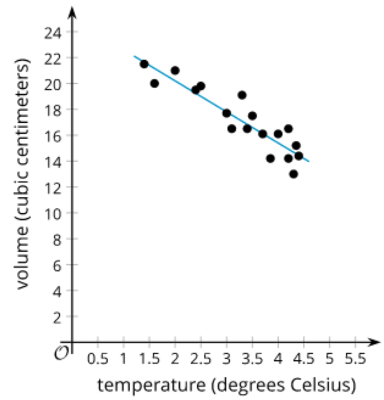
b. $y = 0.44x + 0.04$



c. $y = 4x + 87$



d. $y = -2.4x + 25.0$



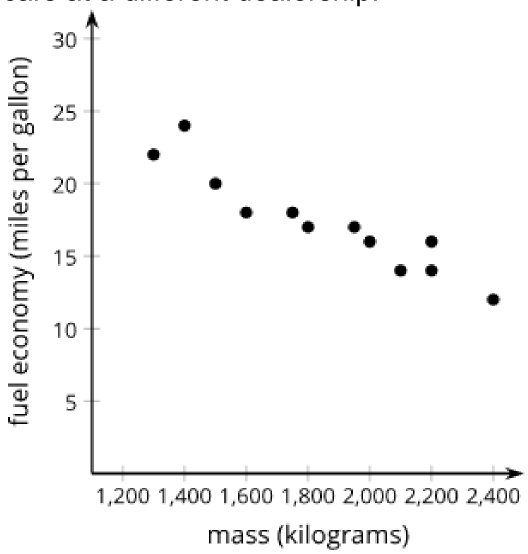
1. Using the horizontal axis for x and the vertical axis for y, interpret the slope of each linear model in the situations shown in the scatter plots.

2. Identify and interpret the y-intercept of each linear model in the situations provided.

Are You Ready For More?

Clare, Diego, and Elena collect data on the mass and fuel economy of cars at different dealerships. Clare finds the line of best fit for data she collected for 12 used cars at a used car dealership. The line of best fit is $y = \frac{-9}{1000}x + 34.3$, where x is the car's mass, in kilograms, and y is the fuel economy, in miles per gallon.

Diego made a scatter plot for the data he collected for 10 new cars at a different dealership.



Elena made a table for data she collected on 11 hybrid cars at another dealership.

Mass (kilograms)	Fuel economy (miles per gallon)
1,100	38
1,200	39
1,250	35
1,300	36
1,400	31
1,600	27
1,650	28
1,700	26
1,800	28
2,000	24
2,050	22

1. Interpret the slope and y-intercept of Clare's line of best fit in this situation.

2. Diego looks at the data for new cars and used cars. He claims that the fuel economy of new cars decreases as the mass increases. He also claims that the fuel economy of used cars increases as the mass increases. Do you agree with Diego's claims? Explain your reasoning.

3. Elena looks at the data for hybrid cars and correctly claims that the fuel economy decreases as the mass increases. How could Elena compare this decrease for hybrid cars to the decrease for new cars? Explain your reasoning.

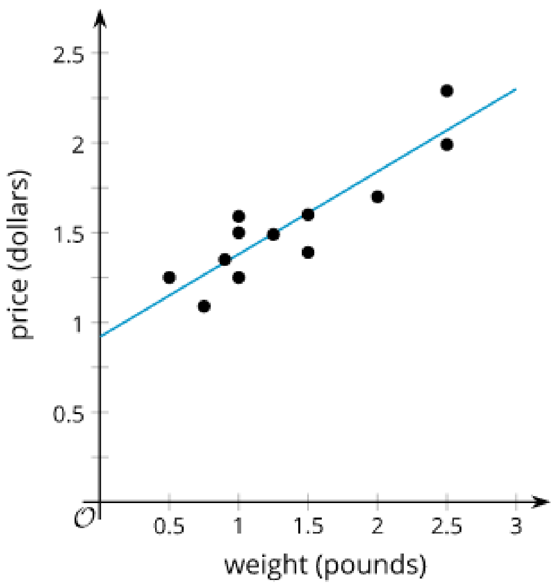
Mass (kilograms)	Fuel economy (miles per gallon)
1,100	38
1,200	39
1,250	35
1,300	36
1,400	31
1,600	27
1,650	28
1,700	26
1,800	28
2,000	24
2,050	22

Lesson Debrief

Lesson 3 Summary and Glossary

While working in math class, it can be easy to forget that reality is somewhat messy. Not all oranges weigh exactly the same amount, beans have different lengths, and even the same person running a race multiple times will probably have different finishing times. We can approximate these messy situations with more precise mathematical tools to better understand what is happening. We can also predict or estimate additional results as long as we continue to keep in mind that reality will vary a little bit from what our mathematical model predicts.

For example, the data in this scatter plot represent the price of a package of broccoli and its weight. The data can be modeled by a line given by the equation $y = 0.46x + 0.92$. The data points do not all fall on the line because there may be factors other than weight that go into the price, such as the quality of the broccoli, the region where the package is sold, and any discounts happening in the store.



We can interpret the y-intercept of the line as the price for the package without any broccoli (which might include the cost of things like preparing the package and shipping costs for getting the vegetable to the store). In many situations, the behavior may not follow the same linear model farther away from the given data, especially as one variable gets close to zero. For this reason, the interpretation of the y-intercept should always be considered in context to determine if it is reasonable to make sense of the value in that way.

We can also interpret the slope as the approximate increase in price of the package for the addition of 1 pound of broccoli to the package.

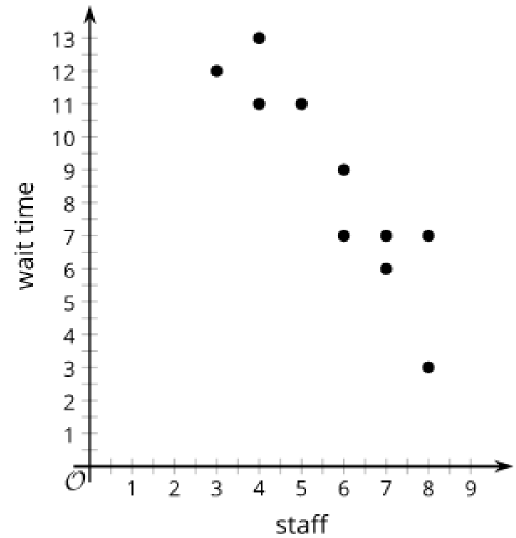
The equation also allows us to predict additional values for the price of a package of broccoli for packages that have weights near the weights observed in the data set. For example, even though the data does not include the price of a package that contains 1.7 pounds of broccoli, we can predict the price to be about \$1.70 based on the equation of the line, since $0.46 \cdot 1.7 + 0.92 \approx 1.70$.

On the other hand, it does not make sense to predict the price of 1,000 pounds of broccoli with these data points, because there may be many more factors that will influence the pricing of packages with weights so much larger than the data points presented here.

Unit 4 Lesson 3 Practice Problems

1. The scatter plot shows the number of minutes people had to wait for service at a restaurant and the number of staff available at the time.

   A line that models the data is given by the equation $y = -1.62x + 18$, where y represents the wait time, and x represents the number of staff available.

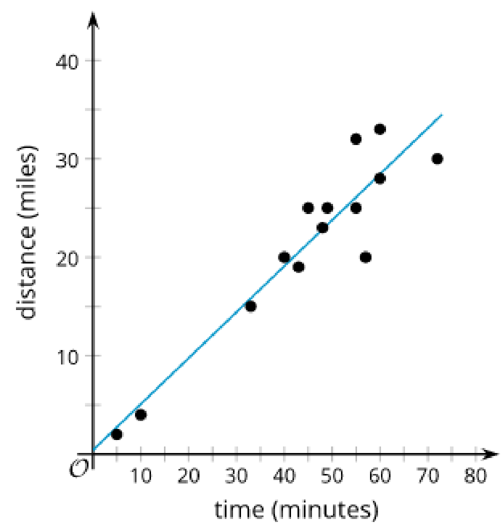


   a. The slope of the line is -1.62. What does this mean in this situation? Is it realistic?

   b. The y-intercept is $(0, 18)$. What does this mean in this situation? Is it realistic?

2. A taxi driver records the time required to complete various trips and the distance for each trip.

   The best fit line is given by the equation $y = 0.467x + 0.417$, where y represents the distance in miles, and x represents the time for the trip in minutes.



   a. Use the best fit line to predict the distance for a trip that takes 20 minutes. Show your reasoning.

   b. Use the best fit line to predict the time for a trip that is 6 miles long. Show your reasoning.

3. (*Technology required.*) The table below represents the height of students, in inches, in Ms. Maas' Math 2 class with their corresponding scores on the last test.

Height (inches)	58	61	61	63	64	67	68	68	72	72	73	74
Test score (out of 100)	80	94	72	75	88	96	89	90	73	84	100	87

a. Use technology to make a scatter plot.

b. Is there a strong or weak relationship between the students' heights and test scores?

c. Is there a positive or negative relationship between the students' heights and test scores?

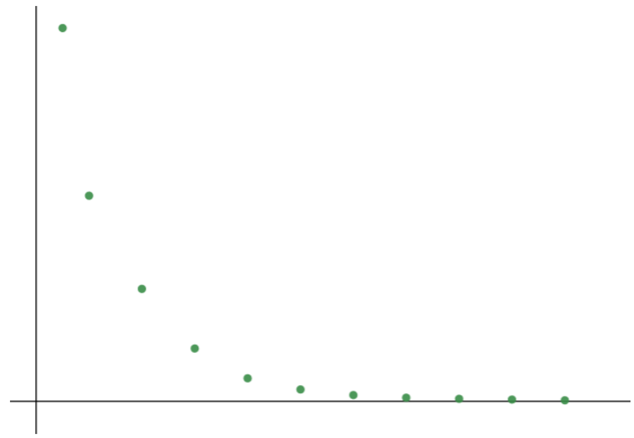
d. Did you expect a different answer? Why or why not?

(From Unit 4, Lesson 2)

4. Does the scatter plot represent:

   a. A linear or nonlinear relationship?

   b. A positive or negative relationship? Or neither?



(From Unit 4, Lesson 1)

5. Parallelogram $ABCD$ has vertices at $A\,(-3, 6.5)$, $B\,(-1, -1.5)$, and $C\,(4.5, 0)$. What are the coordinates of Point D?

(From Unit 3)

6. In a video game, players form teams and work together to earn as many points as possible for their team. Each team can have between two and four players. Each player can score up to 20 points in each round of the game. Han and three of his friends decided to form a team and play a round.

   Write an expression, an equation, or an inequality for each quantity described here. If you use a variable, specify what it represents.

   a. the allowable number of players on a team

   b. the number of points Han's team earns in one round if every player earns a perfect score

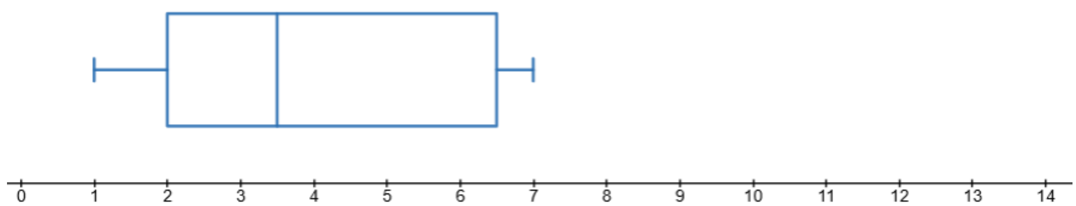
   c. the number of points Han's team earns in one round if no players earn a perfect score

   d. the number of players in a game with six teams of different sizes: two teams have four players each, and the rest have three players each

   e. the possible number of players in a game with eight teams

(From Unit 2)

7. The box plot below represents a data set with a maximum value of 7. If the maximum value were adjusted:

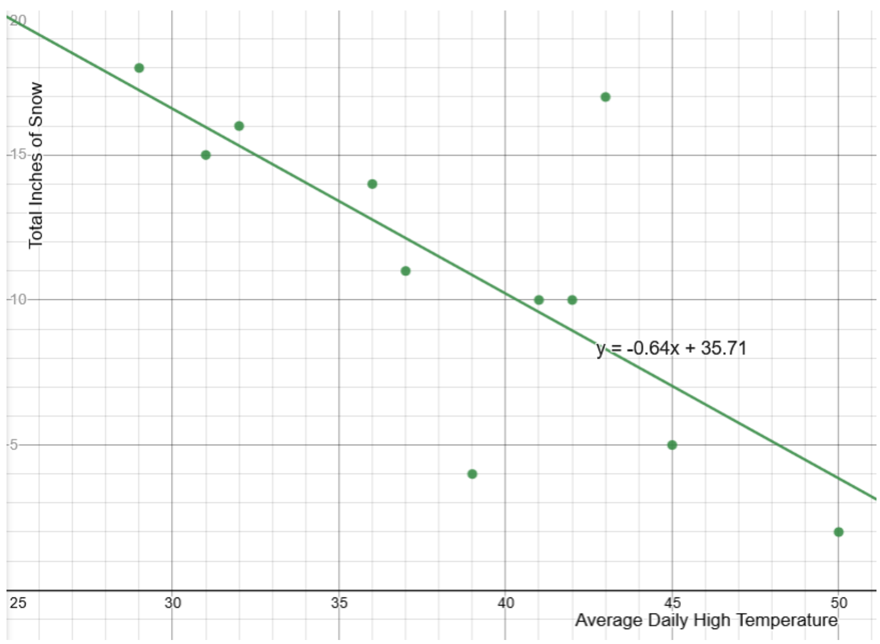


a. What is the smallest whole number it could be so that the whisker to the right would be longer than the box to the right?

b. How would adjusting the maximum value affect the median? How do you know?

(From Unit 1)

8. The scatter plot and line of best fit below represent the average daily high temperature in winter (x) and total snowfall in inches (y) for several cities one winter. In Charlotte, the average high temperature one winter was 51^0 F. Based on the linear relationship, how much snow would you expect Charlotte to get that winter?



(Addressing NC.8.SP.3)

Lesson 4: Fitting Lines

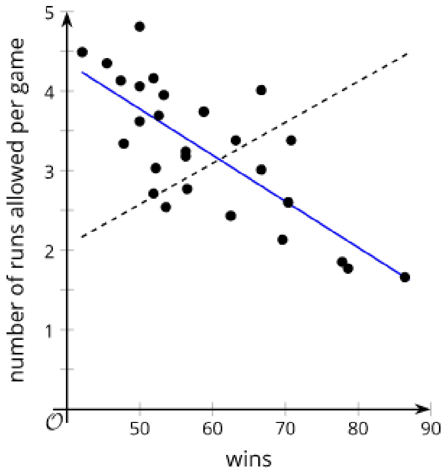
Learning Targets

- I can use technology to find the line of best fit.
- I can use technology to calculate the correlation coefficient.

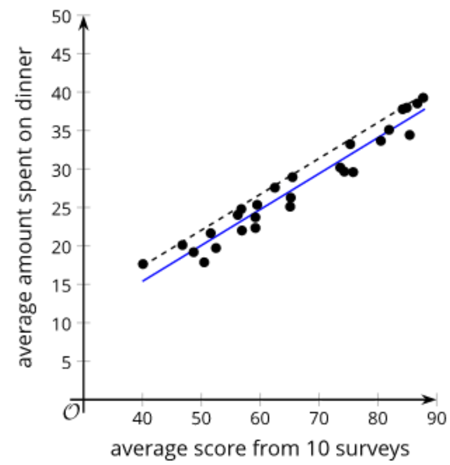
Warm-up: Selecting the Best Line

Which of the lines is the best fit for the data in each scatter plot? Explain your reasoning.

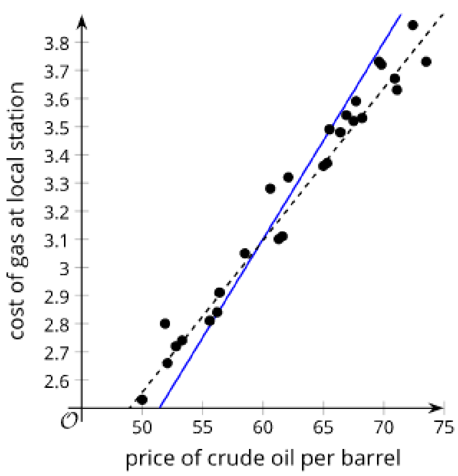
1.



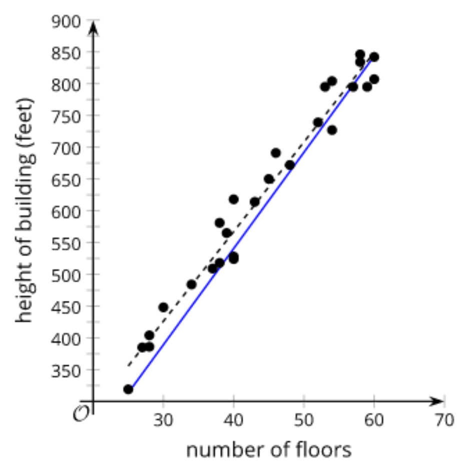
2.



3.



4.

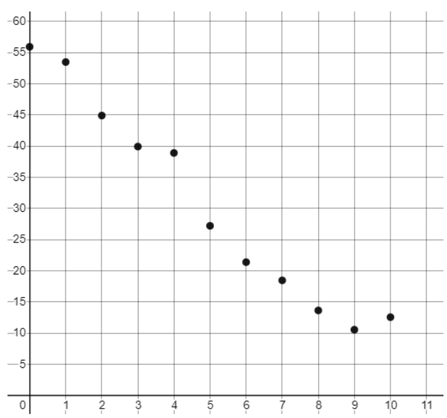


Activity 1: Fitting Lines with Technology

Desmos demonstration to calculate the regression equation, display the line of best fit with the scatter plot, and recognize the correlation coefficient.

Weeks after January 3, 2021	NC new COVID cases (in thousands)
0	55.918
1	53.471
2	44.89
3	39.911
4	38.894
5	27.203
6	21.372
7	18.452
8	13.628
9	10.539
10	12.548

- Access the data on the NC New Covid Cases during the 10 weeks after January 3, 2021 from the Activity 1 tab: https://bit.ly/U4L4DataSet.
- Access the Desmos graphing calculator (www.desmos.com/calculator), click the + icon in the top left corner, and select "table."
- Enter the data into the table. The points will be graphed, creating the scatter plot. Adjust the graph settings manually or by using the "zoom fit" feature.



- To calculate the equation for the line of best fit (the regression equation), go to the next line. Type "y1~mx1+b". This will appear as y_1~mx_1+b, as shown.
- The following will be displayed:
  - the statistics: in which the correlation coefficient, r , can be found
  - the parameters of m (the slope) and b (the y-intercept)
  - the graph of the equation displayed with the scatter plot

$y_1 \sim mx_1 + b$

STATISTICS

$r^2 = 0.962$

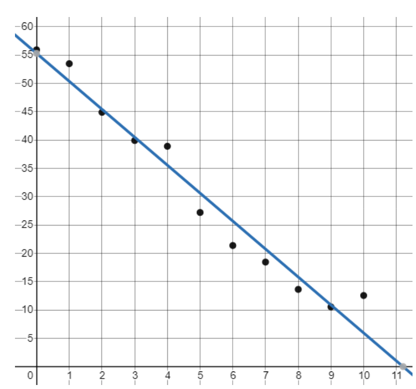
$r = -0.9808$

RESIDUALS

e_1 plot

PARAMETERS

$m = -4.93458$ $b = 55.2935$



For each data table, calculate the regression equation and the correlation coefficient. Round values to the nearest hundredth (two decimal places).

1.

x	y
0	3
1	7
2	9
3	10
4	15
5	19
6	21
7	28

2.

x	y
0	28
1	21
2	20
3	19
4	13
5	15
6	11
7	10

Activity 2: Ice Cream Sales

The average temperature outside in degrees Fahrenheit and the weight of ice cream sold in pounds at a local grocery store are recorded in the table.

Daily high temperature outside in °F (x)	66	69	71	75	77	80	84	90
Weight of ice cream sold in pounds (y)	37	44	43	49	52	61	65	70

1. Use technology to create a scatter plot of this data.
2. Use technology to compute the regression equation.
   a. Record the regression equation. Round values to the nearest hundredth (two decimal places).
   b. Record the correlation coefficient.
3. What are the values for the slope and y-intercept? What do these values mean in this situation?

4. Use the line of best fit to predict:

    a. the number of pounds of ice cream sold if the daily high temperature outside is 70 °F.

    b. the number of pounds of ice cream sold if the daily high temperature outside is 100 °F.

    c. the daily high temperature outside if there were 50 pounds of ice cream sold.

Are You Ready For More?

(*Technology required.*) Priya uses several different ride services to get around her city. The table shows the distance, in miles, she traveled during her last 10 trips and the price of each trip, in dollars.

1. Priya creates a scatter plot of the data using the distance, x, and the price, y. She determines that a linear model is appropriate to use with the data. Use technology to find the regression equation.

2. Interpret the slope and the y-intercept of the line of best fit in this situation.

Distance (miles)	Price (dollars)
3.1	12.5
4.2	14.75
5	16
3.5	13.25
2.5	12
1	9
0.8	8.75
1.6	9.75
4.3	12
3.3	14

3. Use the line of best fit to estimate the cost of a 3.6-mile trip. Will this estimate be close to the actual value? Explain your reasoning.

4. On her next trip, Priya tries a new ride service and travels 3.6 miles, but she pays only $4.00 because she receives a discount. Include this trip in the table and calculate the regression equation for the 11 trips. Did the slope of the line of best fit increase, decrease, or stay the same? Why? Explain your reasoning.

5. Priya uses the new ride service for her 12th trip. She travels 4.1 miles and is charged $24.75. How do you think the slope of the regression equation will change when this 12th trip is added to the table?

Lesson Debrief

Lesson 4 Summary and Glossary

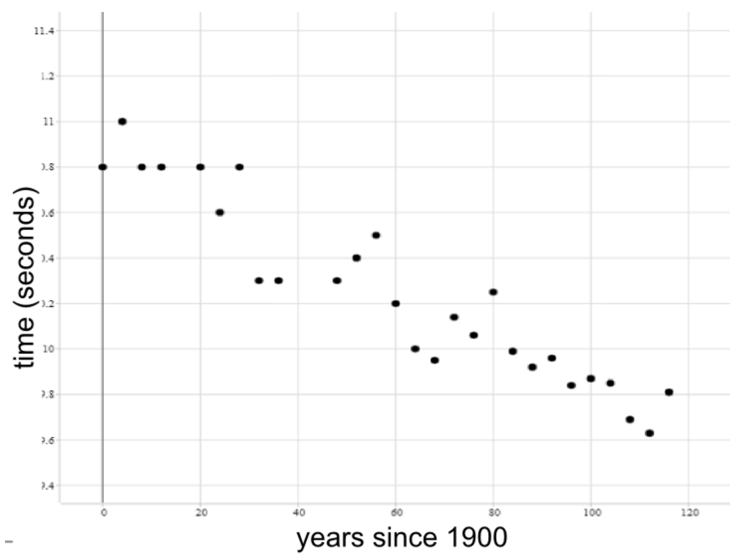
For data sets that appear to have a linear relationship, calculating an equation for the line of best fit can help you describe the relationship between the variables and make predictions. The equation for the line of best fit is called a regression equation.

Take for example the data of winning times for the Men's 100-Meter Run in the Olympics.

Years since 1900	Time (sec)
0	10.8
4	11
8	10.8
12	10.8
20	10.8
24	10.6
28	10.8
32	10.3
36	10.3
48	10.3
52	10.4
56	10.5
60	10.2
64	10

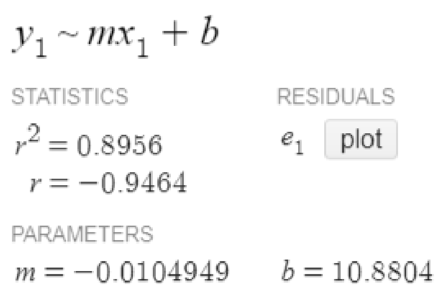
Years since 1900	Time (sec)
68	9.95
72	10.14
76	10.06
80	10.25
84	9.99
88	9.92
92	9.96
96	9.84
100	9.87
104	9.85
108	9.69
112	9.63
116	9.81

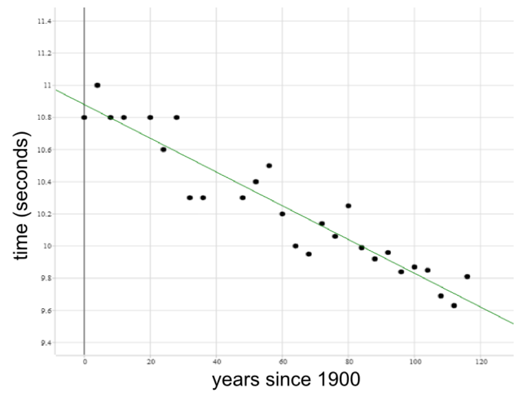
The scatter plot of this data reveals a strong, negative linear relationship.



We can use technology to graph the line of best fit and find the regression equation that describes it. To calculate a line of best fit in Desmos, type "y1~m*x1+b" on the entry line below the table of values. Under "parameters" you will see the slope (m) and y-intercept (b) of the line of best fit. Use these to write the regression equation.

Using this process, we find that the line of best fit has the equation $y = -0.01x + 10.88$. This line is shown below. We can also identify the correlation coefficient as $r = -0.9464$ which indicates a strong, negative relationship between the variables.





The *y*-intercept of this line is 10.88; this is the *predicted* winning time for the Men's 100-Meter Run for the year 1900. This predicted value is very close to the actual winning time of 10.8 seconds. The slope of the best fit line is -0.01. In other words, our linear model predicts that the Men's 100-Meter Run winning time decreases by 0.01 seconds for every year that goes by.

There were no Summer Olympics held in 1940 due to World War II. However, using our model we could predict that the winning time in 1940 would have been:

$$y = -0.01(40) + 10.88$$

$y = 10.48$ seconds

This is an **interpolated** value and is a reasonable prediction.

Our model can also predict the winning time for the 2024 Olympics:

$$y = -0.01(124) + 10.88$$

$y = 9.64$ seconds

However, this is an **extrapolated** value. This prediction falls outside the boundaries of our data set and should be interpreted with caution. Why? Imagine extending the line downward and to the right. Eventually, the line will cross the *x*-axis, predicting *negative* times for the Men's 100-Meter Run, which does not make sense. This means that the linear pattern cannot continue forever. Since we don't know at what point the line will stop being a good model, our prediction for the year 2024 is less reliable.

Interpolation: When a predicted value comes from an *x*-value inside the boundaries of the given data set.

Extrapolation: When a predicted value comes from an *x*-value that falls outside the boundaries of the given data set.

Unit 4 Lesson 4 Practice Problems

1. (*Technology required.*)

x	y
83	102
87	115
91	107
93	122
97	125
97	127
101	120
104	127

   a. Use graphing technology to create a scatter plot and find the regression equation for the line of best ft.

   b. What does the best fit line estimate for the y value when x is 100?

2. (*Technology required.*)

x	y
2.3	6.2
2.8	5.7
3.1	4.7
3	3.2
3.5	3
3.8	2.8

   a. What is the equation of the line of best fit? Round values to the nearest hundredth (two decimal places).

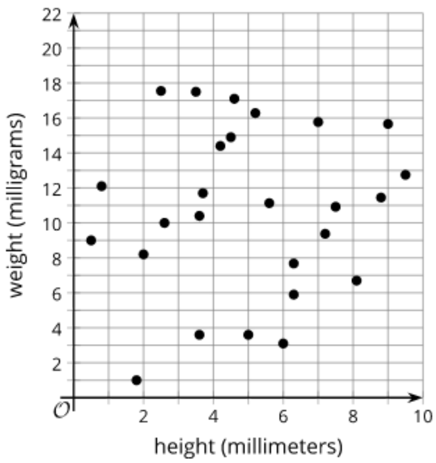
   b. What does the equation predict for y when x is 2.3? Round values to the nearest thousandth (three decimal places).

   c. How does the predicted value compare to the actual value from the table when x is 2.3?

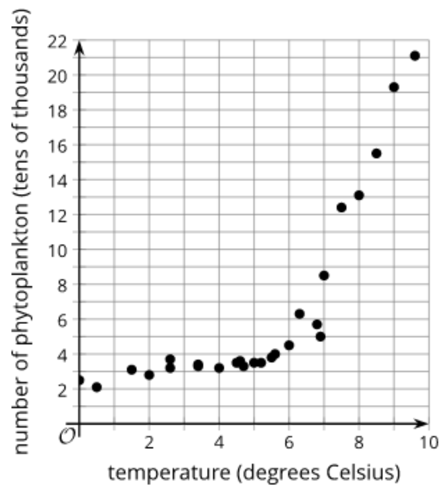
   d. How does the predicted value compare to the actual value from the table when x is 3?

3. Which of these scatter plots shows data that would best be modeled with a linear function? Explain your reasoning.

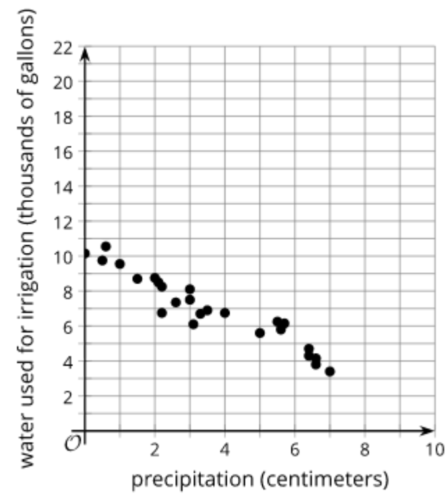
a.



b.



c.



4. A seed is planted in a glass pot, and its height is measured in centimeters every day.

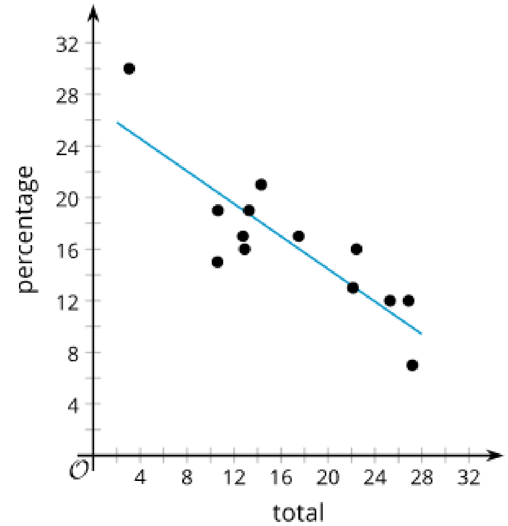
   The best fit line is given by the equation $y = 0.404x - 5.18$, where y represents the height of the plant above ground level and x represents the number of days since it was planted.

   a. What is the slope of the best fit line? What does the slope of the line mean in this situation? Is it reasonable?

   b. What is the y-intercept of the best fit line? What does the y-intercept of the line mean in this situation? Is it reasonable?

(From Unit 4, Lesson 3)

5. At a restaurant, the total bill and the percentage of the bill left as a tip are represented in the scatter plot.



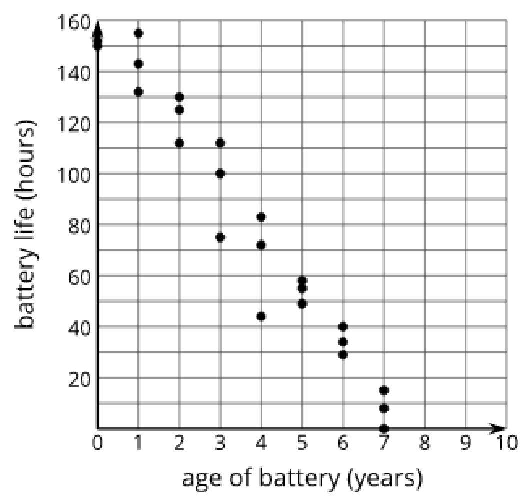
The best fit line is represented by the equation $y = -0.632x + 27.1$, where x represents the total bill in dollars and y represents the percentage of the bill left as a tip.

a. What does the best fit line predict for the percentage of the bill left as a tip when the bill is \$15? Is this reasonable?

b. What does the best fit line predict for the percentage of the bill left as a tip when the bill is \$50? Is this reasonable?

(From Unit 4, Lesson 3)

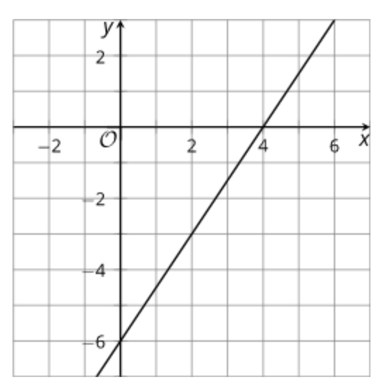
6. A recent study investigated the amount of battery life remaining in alkaline batteries of different ages. The scatter plot shows this relationship between the different alkaline batteries tested.



The scatter plot includes a point at $(7, 15)$. Describe the meaning of this point in this situation.

(From Unit 4, Lesson 1)

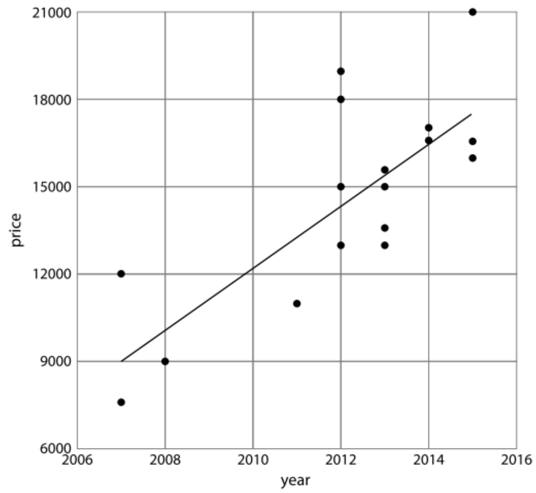
7. Here is the graph of the equation $3x - 2y = 12$. Select **all** the coordinate pairs that represent a solution to the equation.



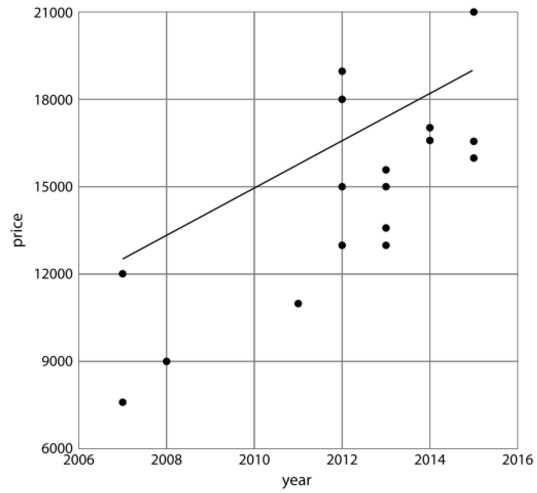
a. $(2, -3)$
b. $(4, 0)$
c. $(5, -1)$
d. $(0, -6)$
e. $(2, 3)$

(Unit 3)

8. The scatter plots below both show the year and price for the same 17 used cars. However, each scatter plot shows a different model for the relationship between year and price.[1]



A



B

a. Look at diagram A.

- For how many cars does the model in diagram A make a good prediction of its price?
- For how many cars does the model underestimate the price?
- For how many cars does it overestimate the price?

b. Look at diagram B.

- For how many cars does the model in diagram B make a good prediction of its price?
- For how many cars does the model underestimate the price?
- For how many cars does it overestimate the price?

c. For how many cars does the prediction made by the model in diagram A differ by more than $3,000?

What about the model in diagram B?

d. Which model does a better job of predicting the price of a used car from its year?

(Addressing NC.8.SP.2)

[1] Adapted from IM 6–8 Math https://curriculum.illustrativemathematics.org/MS/index.html, which was originally developed by Open Up Resources and authored by Illustrative Mathematics, and is copyright 2017–2019 by Open Up Resources. It is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). OUR's 6–8 Math Curriculum is available at https://openupresources.org/math-curriculum/. Adaptations and updates to IM 6–8 Math are copyright 2019 by Illustrative Mathematics, and are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

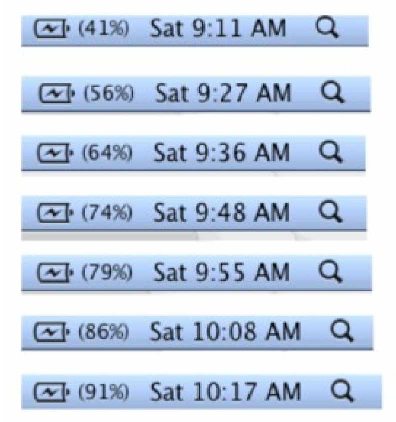
Lessons 5 & 6: Checkpoint

Learning Targets

- I can continue to grow as a mathematician and challenge myself.
- I can share what I know mathematically.

Station C: Laptop Battery Charge[1]

Elena forgot to plug in her laptop before she went to bed. She wants to take the laptop to her friend's house with a full battery. The pictures below show screenshots of the battery charge indicator after she plugs in the computer at 9:11 a.m.



1. Use the screenshots to consider the questions below.

   a. The screenshots suggest a relationship between two variables. What are the two variables in this situation?

   b. Use technology to make a scatter plot of the data.

   c. Use technology to draw a line that fits the data and find the equation of the line in slope-intercept form.

   d. What is the slope of the line? Explain the meaning of the slope in the context of the problem.

   e. What is the y-intercept of the equation of the line? Explain the meaning of the y-intercept in the context of the problem.

   f. At what time does your model predict the battery was 50% charged?

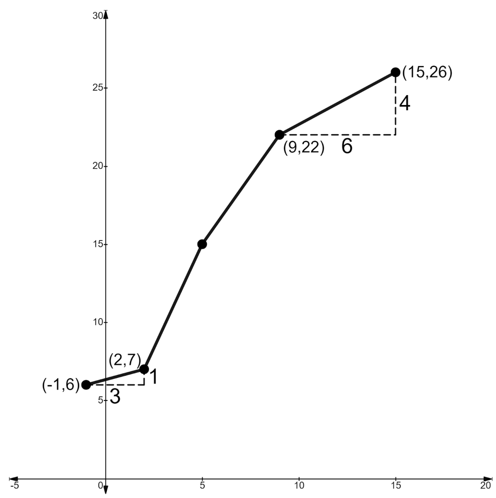
   g. When can Elena expect to have a fully charged battery?

[1] Adapted from https://tasks.illustrativemathematics.org/

2. At 9:27 AM Elena makes a quick calculation: *The battery seems to be charging at a rate of 1 percentage point per minute. So the battery should be fully charged at 10:11 AM.*

   Explain Elena's calculation. Is her estimate most likely an under- or over-estimate? How does it compare to your prediction?

3. Compare the **average rate of change** of the battery charging function on the first given time interval and on the last given time interval. What does this tell you about how the battery is charging?

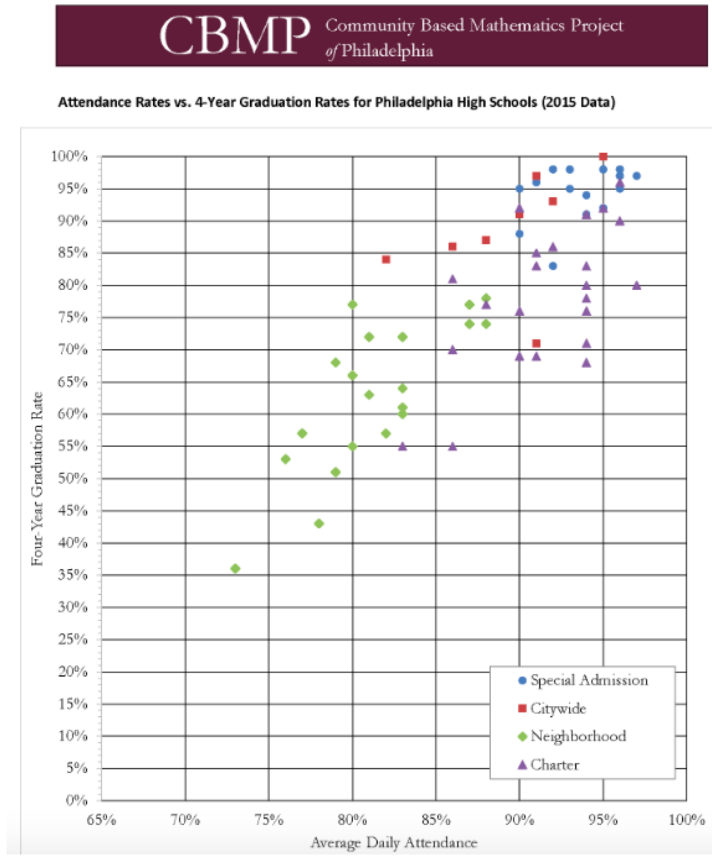
Learn More	
	**Average rate of change:** Recall that rate of change is how one quantity changes in relation to another quantity. Linear functions will have a constant rate of change, while non-linear functions' rates of change vary based on the interval of the graph you are looking at. In the non-linear graph to the left, for example, the rate of change between $(-1, 6)$ and $(2, 7)$ is $\frac{7-6}{2-(-1)}$, or $\frac{1}{3}$. The rate of change between $(9, 22)$ and $(15, 26)$ is $\frac{26-22}{15-9}$, or $\frac{4}{6}$, which is $\frac{2}{3}$. A sample comparison of the average rate of change of the first and last intervals might look like: The first interval has an average rate of change of $\frac{1}{3}$, while the last interval has an average rate of change of $\frac{2}{3}$. The rate of change of the last interval is greater than the rate of change of the first interval meaning the graph is steeper in the last interval compared to the first.

4. How long would it take for the battery to charge if it started out completely empty?

Station D: School Data

1. In Philadelphia, PA there are four types of high schools: neighborhood schools, citywide schools, charter schools, and special admission schools. The graph below compares the average daily attendance rates at these different schools with the rate of students who graduate from those schools in four years.

   What do you notice? What do you wonder?



2. Consider the questions below.

   a. Choose 10–20 high schools in your district or state and research two characteristics you are interested in. Here are some examples:

      - four-year graduation rate compared to percentage of chronic absenteeism
      - class size for Math 1 compared to four-year graduation rate
      - percentage of students participating in an AP assessment compared to percentage of students attending college after graduation
      - percentage of economically disadvantaged students compared to number of suspensions or expulsions per 1000 students

   b. Do you predict any trends?

   c. Create a scatter plot of the data and determine the line of best fit. Is there a relationship between the two characteristics you researched? What trends do you see? What follow-up questions or relationships does this make you want to research? What might be some possible reasons for these trends?

Station F: A Midpoint Miracle[2]

Let's investigate quadrilaterals further by examining midpoint quadrilaterals!

1. Construct a quadrilateral and its midpoint quadrilateral using the following steps:

   Step 1: Go to desmos.com/geometry.

   Step 2: In the "Construct" panel on the left, select "Polygon." Click on one point on the grid and then click on another point, creating the first side of your quadrilateral. Click on two more grid points to make the other two vertices of the quadrilateral. End by connecting the 4th side to the very first point you clicked to close your quadrilateral.

   Step 3: Create a midpoint quadrilateral:

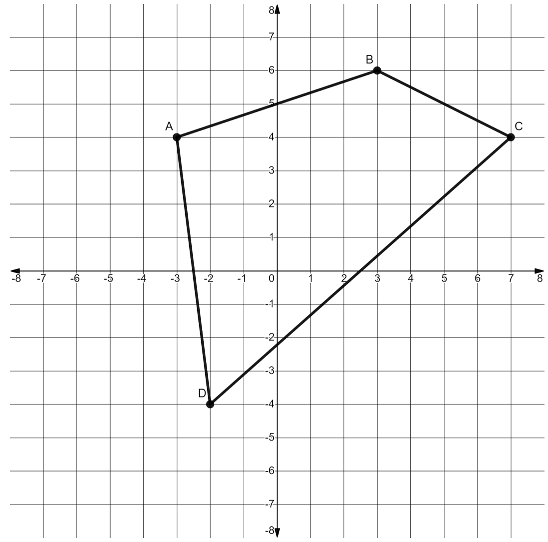
   - Click on "Midpoint" under "More Tools" in the "Construct" panel. Then, click on one side of the quadrilateral. Desmos will mark the midpoint of that side. Repeat this step until midpoints of each side of the quadrilateral have been located.
   - Connect the midpoints to form a midpoint quadrilateral by selecting the "Polygon" tool in the "Construct" panel and clicking on each midpoint, going clockwise, ending by clicking on the midpoint you started with.

2. What do you notice about the midpoint quadrilateral?

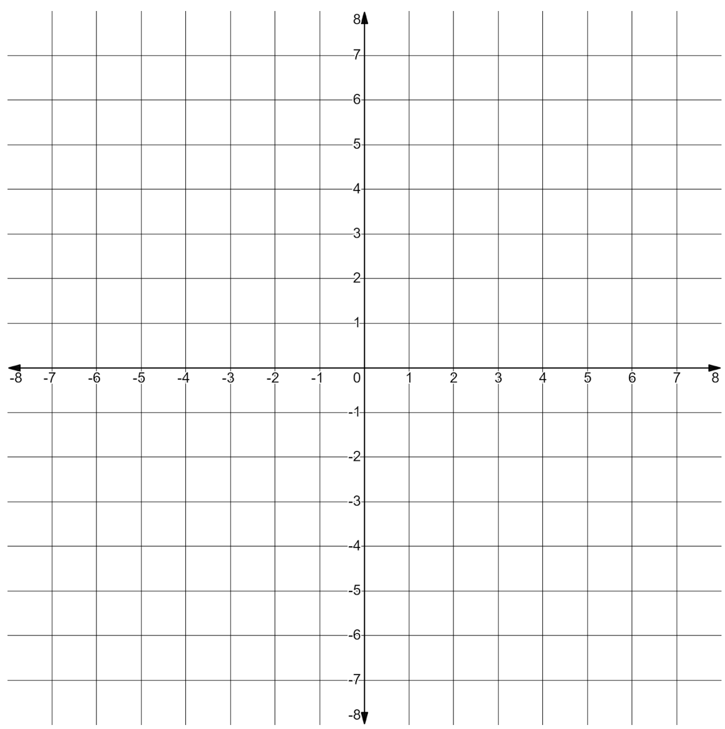
3. Create different versions of the original quadrilateral by dragging its vertices around using the "Select" tool in the "Construct" panel. As you drag points, the midpoint quadrilateral will adjust accordingly. Does your observation in Question 2 seem to true for all of the midpoint quadrilaterals? If you have a different observation now, what is it?

[2] Adapted from https://tasks.illustrativemathematics.org/

4. In this diagram, $ABCD$ is the original quadrilateral. Find the midpoint of each side, and label the midpoints *P*, *Q*, *R*, and *S*. Draw the midpoint quadrilateral $PQRS$. Now prove that your observation from Questions 2 or 3 is also true for quadrilateral $PQRS$.



5. If time allows, and you are ready for more, create your own quadrilateral $ABCD$ (and corresponding midpoint quadrilateral $PQRS$) on the coordinate plane below to further test your claim.



Station G: Cat and Dog Food

Andre works at Pet Palace grooming and kennel. He is in charge of purchasing the food and other supplies for the cats and dogs to use during their stay. Andre's manager asked for the cost for dog food, cat food, leashes, and brushes, but Andre only has information for the total amount he spent. Use the information below to calculate the costs for each of these items.[3]

1. One week, Andre bought three bags of *Tabitha Tidbits* and four bags of *Figaro Flakes* for $43.00. The next week, he bought three bags of *Tabitha Tidbits* and six bags of *Figaro Flakes* for $54.00. Based on this information, figure out the price of one bag of each type of cat food. Explain your reasoning.

2. One week, Andre bought two bags of *Brutus Bites* and three bags of *Milo Munchies* for $42.50. The next week, he bought five bags of *Brutus Bites* and six bags of *Milo Munchies* for $94.25. Based on this information, figure out the price of one bag of each type of dog food. Explain your reasoning.

3. Andre purchased six dog leashes and six cat brushes for $45.00 for Elena to use while pampering the pets. Later in the summer, he purchased three additional dog leashes and two cat brushes for $19.00. Based on this information, figure out the price of each item. Explain your reasoning.

[3] Adapted from Secondary Math 1 Mathematics Vision project http://www.mathematicsvisionproject.org, licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0)

4. One week Andre purchased two boxes of cat treats and three boxes of dog treats for $18.50. The next week, he bought two boxes of cat treats and two boxes of dog treats for $14.00. The third week, he bought five boxes of both cat and dog treats for $35.00. Based on this information, figure out the price of each item. Explain your reasoning.

5. Andre has noticed that because his purchases have been somewhat similar, it has been easy to figure out the cost of each item. However, his last set of receipts has him puzzled. One week, he tried out cheaper brands of cat and dog food. On Monday, he purchased three small bags of cat food and five small bags of dog food for $22.75. Because he went through the small bags quite quickly, he had to return to the store on Thursday to buy two more small bags of cat food and three more small bags of dog food, which cost him $14.25. Based on this information, figure out the price of each bag of the cheaper dog food. Explain your reasoning.

Lesson 7: Residuals

Learning Target

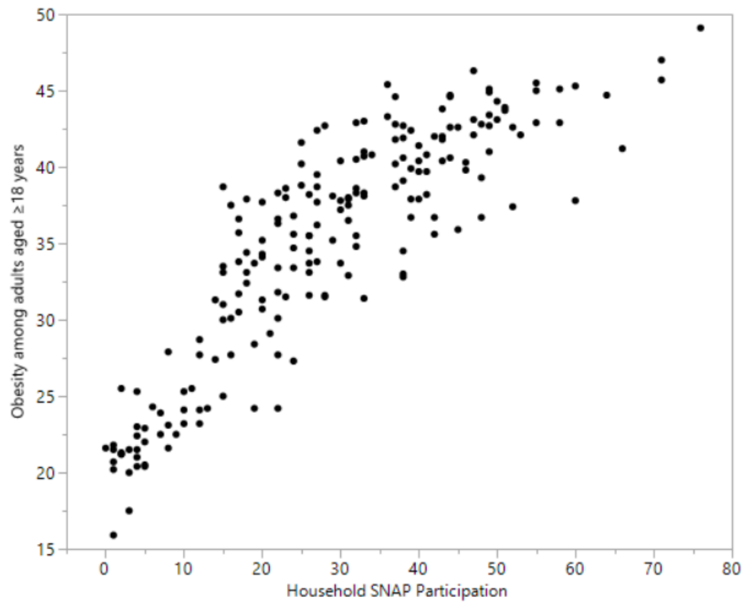
- I can plot and calculate residuals for a data set and use the information to judge whether a linear model is a good fit.

Bridge

The following scatter plot represents the frequency of obesity based on household participation in SNAP (food stamps) by census tract in Baltimore, MD.[1]

a. Draw a linear model for the data, paying attention to the closeness of the data points on either side of the line.

b. Estimate the equation of the best fit line using the slope and y-intercept of your linear model.



(Image source[2])

c. Lin wanted to read more about this and learned that there are many factors that contribute to this relationship that include affordability of healthy food, access to quality healthcare, and access to safe and affordable places to exercise. In what way could access to grocery stores with healthy options also play a role in this relationship?

[1] North Carolina Department of Public Instruction. (2018). *Instructional Support Tools for Achieving New Standards: 8th Grade.* https://files.nc.gov/dpi/documents/curriculum/mathematics/scos/current/8th-unpacking.pdf

[2] Borders, S. (2018). *Barriers to and Facilitators of, Providing an Effective Approach to Preventing Obesity in Women Ages 18 to 39 A Geospatial Analysis of Baltimore.* https://www.researchgate.net/publication/328489797_Barriers_to_and_Facilitators_of_Providing_an_Effective_Approach_to_Preventing_Obesity_in_Women_Ages_18_to_39_A_Geospatial_Analysis_of_Baltimore_Barriers_to_and_Facilitators_of_Providing_an_Effective_A

Warm-up: Differences in Expectations

Mentally calculate how close the predicted value is to the actual value using the difference:
$actual\ value - predicted\ value$.

1. Actual value: 24.8 grams. Predicted value: 19.6 grams

2. Actual value: $112.11. Predicted value: $109.30

3. Actual value: 41.5 centimeters. Predicted value: 45.90 centimeters

4. Actual value: -1.34 degrees Celsius. Predicted value: -2.45 degrees Celsius

Activity 1: Orange Return

1. Use technology to make a scatter plot of orange weights and find the line of best fit.

    a. In desmos.com/calculator, click on the + icon to add a table. In the table, enter in the following values, which can also be found on the Activity 1 tab: https://bit.ly/U4L7DataSet.

    b. Find the line of best fit for this scatter plot. Under the table, type the following in a new entry line: $y_1 \sim m \cdot x_1 + b$

Number of oranges	Weight in kilograms
3	1.027
4	1.162
5	1.502
6	1.617
7	1.761
8	2.115
9	2.233
10	2.569

Use the following table for questions 2–5.

Number of oranges	**Weight in kilograms**	**Linear prediction weight in kilograms**
3	1.027	
4	1.162	
5	1.502	
6	1.617	
7	1.761	
8	2.115	
9	2.233	
10	2.569	

2. What does the linear model predict for the weight of the box of oranges for each of the number of oranges? Complete the “linear prediction weight in kilograms” column. (To show the y-value predicted by the linear model at a given x-value, click and hold on the best fit line at the x-value you're considering.)

3. Compare the actual weight of the box with three oranges in it to the predicted weight of the box with three oranges in it. Explain or show your reasoning.

4. How many oranges are in the box when the linear model predicts the weight best? Explain or show your reasoning.

5. How many oranges are in the box when the linear model predicts the weight least well? Explain or show your reasoning.

6. The difference between the actual value and the value predicted by a linear model is called the residual. If the actual value is greater than the predicted value, the residual is positive. If the actual value is less than the predicted value, the residual is negative.

   For the orange weight data set, what is the residual for the best fit line when there are 3 oranges? In Desmos, on the same axes as the scatter plot, plot this residual point using the format (3, residual value) on the next entry line.

7. In Desmos, it is possible to graph the residuals all at once by clicking the button under residuals labeled “plot.” The residual values are also automatically added to your scatterplot table. Check out the graph of the residuals.

   What is the residual value for eight oranges and what are the coordinates of the point on the residual plot?

8. Which point on the scatter plot has the residual closest to zero?

   What does this mean about the weight of the box with that many oranges in it?

9. How can you use the residuals to decide how well a line fits the data?

Activity 2: Best Residuals

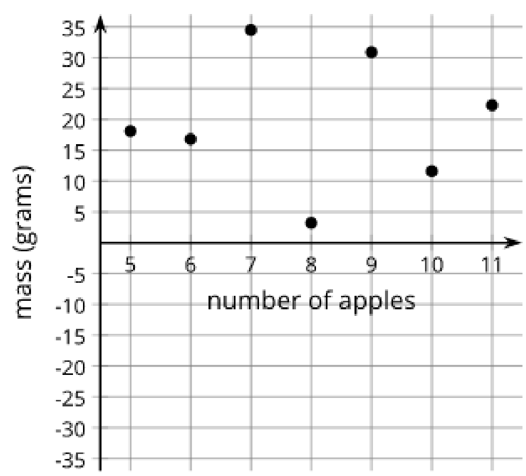
1. Match the scatter plots and given linear models to the graph of the residuals.

2. Turn the scatter plots over so that only the residuals are visible.

   Based on the residuals, which line would produce the most accurate predictions? Which line fits its data worst?

Are You Ready For More?

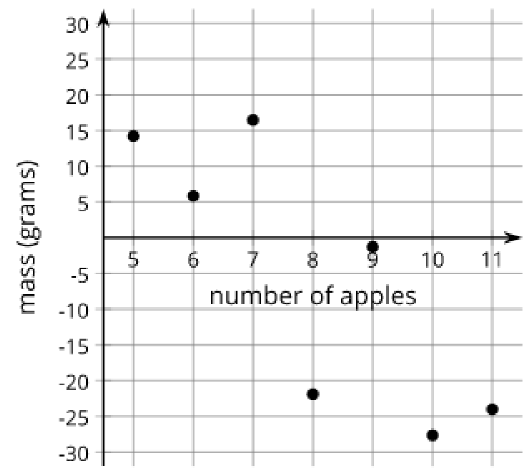
1. Tyler estimates a line of best fit for some linear data about the mass, in grams, of different numbers of apples. Here is the graph of the residuals.



   a. How do the points on the scatter plot compare to Tyler's best fit line?

   b. How well does Tyler's line of best fit model the data? Explain your reasoning.

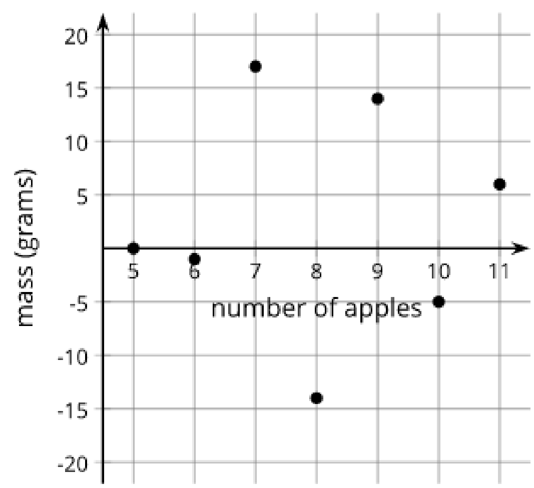
2. Lin estimates a line of best fit for the same data. The graph shows the residuals.



   a. How do the points on the scatter plot compare to Lin's best fit line?

   b. How well does Lin's line of best fit model the data? Explain your reasoning.

3. Kiran also estimates a line of best fit for the same data. The graph shows the residuals.



   a. How do the points on the scatter plot compare to Kiran's best fit line?

   b. How well does Kiran's line of best fit model the data? Explain your reasoning.

4. Who has the best estimate of the line of best fit—Tyler, Lin, or Kiran? Explain your reasoning.

Lesson Debrief

Lesson 7 Summary and Glossary

When fitting a linear model to data, it can be useful to look at the residuals. **Residuals** are the difference between the y-value for a point in a scatter plot and the value predicted by the linear model for that x-value.

Residual: The difference between the y-value for a point in a scatter plot and the value predicted by the linear model for the associated x-value.

In the scatter plot showing the length of fish and the age of fish, the point (2,100) represents a fish who is 2 years old and 100 mm long. We can compare this to what the best fit line predicts for the length of a two-year-old-fish:

$$y = 34.08 \cdot 2 + 23.78$$

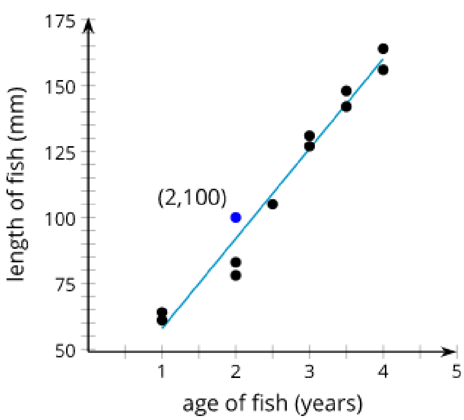
$$y = 91.94\ mm$$

To find the residual, we take the actual length of the two-year-old fish and subtract the length that our model predicts:

$$100 - 91.94 = 8.06\ mm$$

The residual of $8.06\ mm$ means the actual fish is about 8 millimeters longer than the linear model predicts for a fish of that same age.

$$y = 34.08x + 23.78$$



When the point on the scatter plot is above the line, it has a positive residual. When the point on the scatter plot is below the line, the residual is a negative value. A line that has smaller residuals would be more likely to produce predictions that are close to the actual value.

Unit 4 Lesson 7 Practice Problems

1. Han creates a scatter plot that displays the relationship between the number of items sold, x, and the total revenue, y, in dollars. Han creates a line of best fit and finds that the residual for the point $(12, 1000)$ is 75. The point $(13, 930)$ has a residual of -40. Interpret the meaning of -40 in the context of the problem.

2. The line of best fit for a data set is $y = 1.1x + 3.4$. Find the residual for each of the coordinate pairs, (x, y).

   a. $(5, 8.8)$

   b. $(2.5, 5.95)$

   c. $(0, 3.72)$

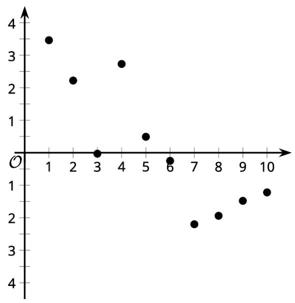
   d. $(1.5, 5.05)$

   e. $(-3, 0)$

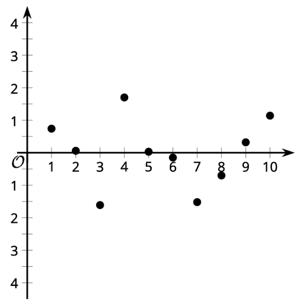
   f. $(-5, -4.86)$

3. Plots of the residuals for four different models of the same data set are displayed. Which of the following plots fit the data **best**?

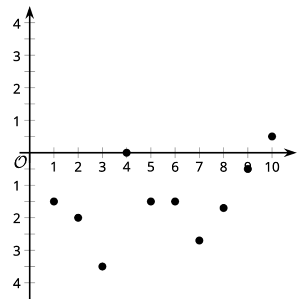
a.



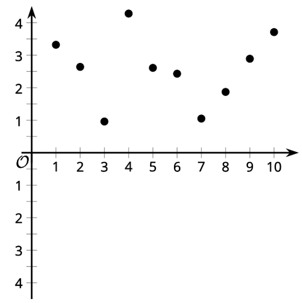
b.



c.



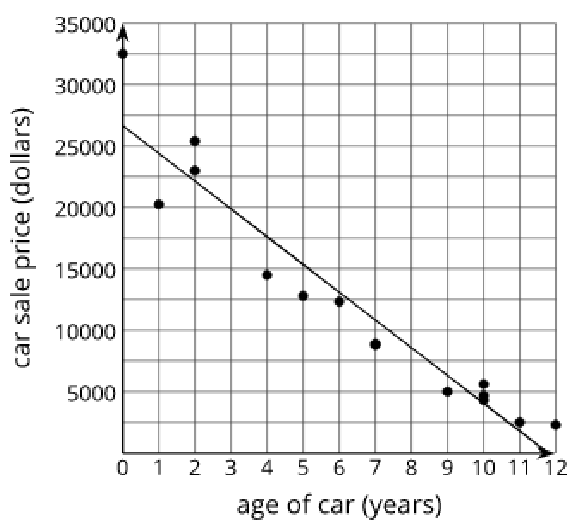
d.



4. A local car salesperson created a scatter plot to display the relationship between a car's sale price in dollars, y, and the age of the car in years, x. The scatter plot and the line of best fit are displayed in the graph.

   The car salesperson looks at the residuals for the car sales.

   a. For a car that is four years old, does the salesperson sell above or below her average selling price? Explain your reasoning.



   b. For a car that is 12 years old, does the salesperson sell above or below her average selling price? Explain your reasoning.

5. (*Technology required.*) Data about the outside temperature and gas used for heating a building are given in the table.

   Use technology to create the line of best fit for the data. The data can be found on the Practice Problem tab: https://bit.ly/U4L7DataSet.

Temperature (deg F) x	Gas usage (therms) y
58	5,686
62	7,373
64	5,805
67	5,636
70	3,782
73	3,976
74	3,351
74	3,396
75	2,936
73	3,078
65	4,549
59	7,022
58	6,106
62	4,566
64	4,608
67	5,790
70	6,501
73	3,843

   a. What is the equation of the line of best fit for this data? Round numbers to the nearest whole number.

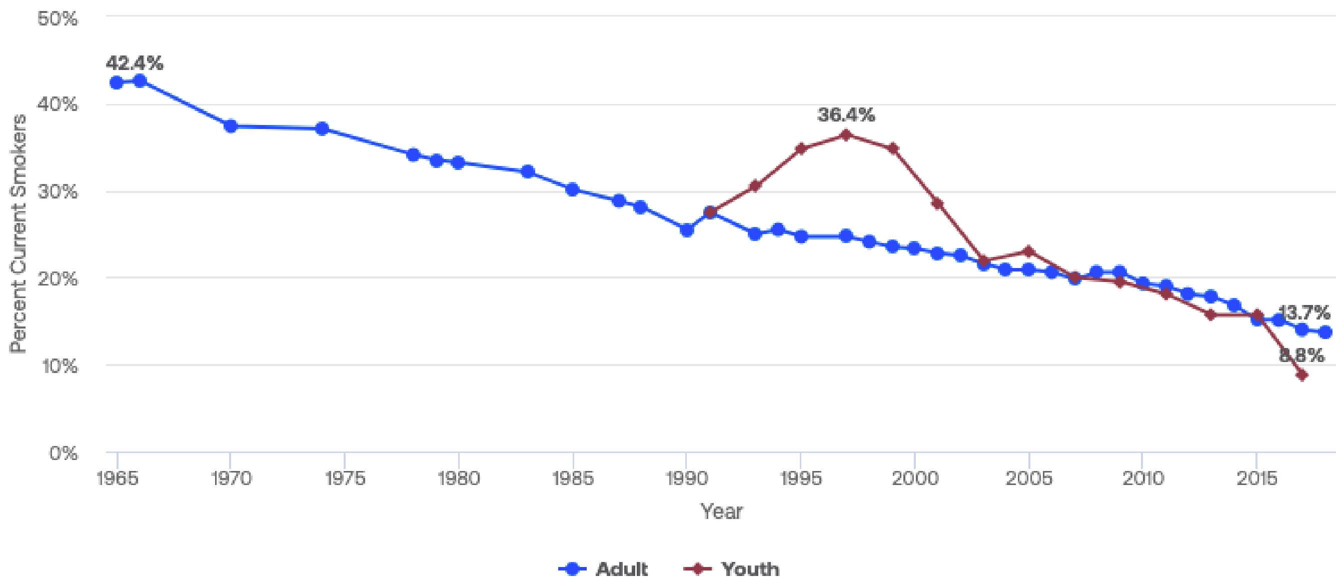
   b. What is the slope of the line of best fit? What does it mean in this situation?

   c. What does the line of best fit predict for gas usage when the outside temperature is 59 degrees Fahrenheit?

   d. How does the actual gas usage compare to the predicted gas usage when the outside temperature is 59 degrees Fahrenheit?

(From Unit 4, Lesson 4)

6. The graph below, from analysis of Center for Disease Control data by the American Lung Association, shows trends in cigarette smoking from 1965 to 2018.



(Image source[3])

Based on the graph:

a. Describe the pattern of cigarette smoking in adults and youth over the past 55 years.

b. Would a linear model be a good fit for the cigarette smoking rates for adults? For youth? Explain your answer.

(From Unit 4, Lesson 1)

[3] American Lung Association. *Overall Tobacco Trends*. https://www.lung.org/research/trends-in-lung-disease/tobacco-trends-brief/overall-tobacco-trends

7. One equation in a system of linear equations is $2x + 3y = -10$. If the system has no solutions, what could be the other equation?

(From Unit 3)

8. Solve for x:

   a. $3.5x + 9 = 5x - 18$

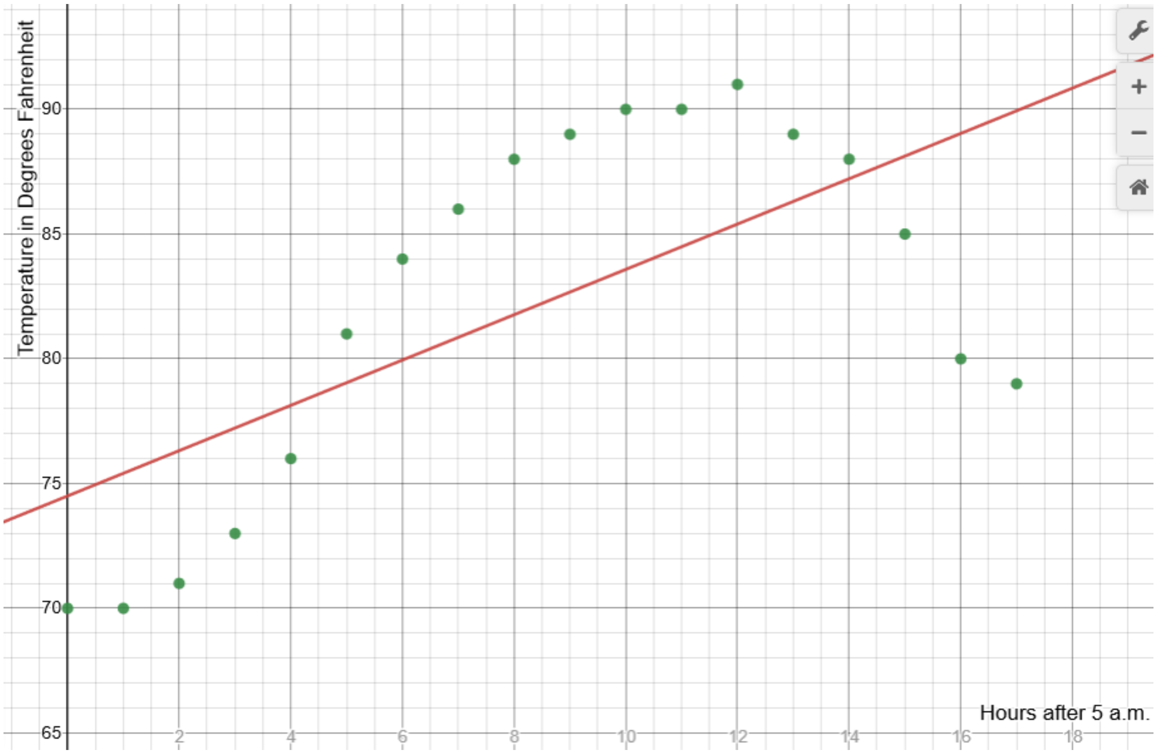
   b. $3.5x + 9 > 5x - 18$

   c. For questions a and b, is 18 a solution to a, b, both, or neither? Why?

   d. For questions a and b, is 20 a solution to a, b, both, or neither? Why?

(From Unit 2)

9. The scatterplot below represents the temperature every hour on June 21 in Charlotte, NC, starting at 5:00 a.m. Andre draws a line of best fit to try to predict the temperature the next morning.



Is Andre’s line a good model to predict the temperature on the morning of June 22? Explain your reasoning.

(Addressing 8.SP.2)

Lesson 8: Causal Relationships

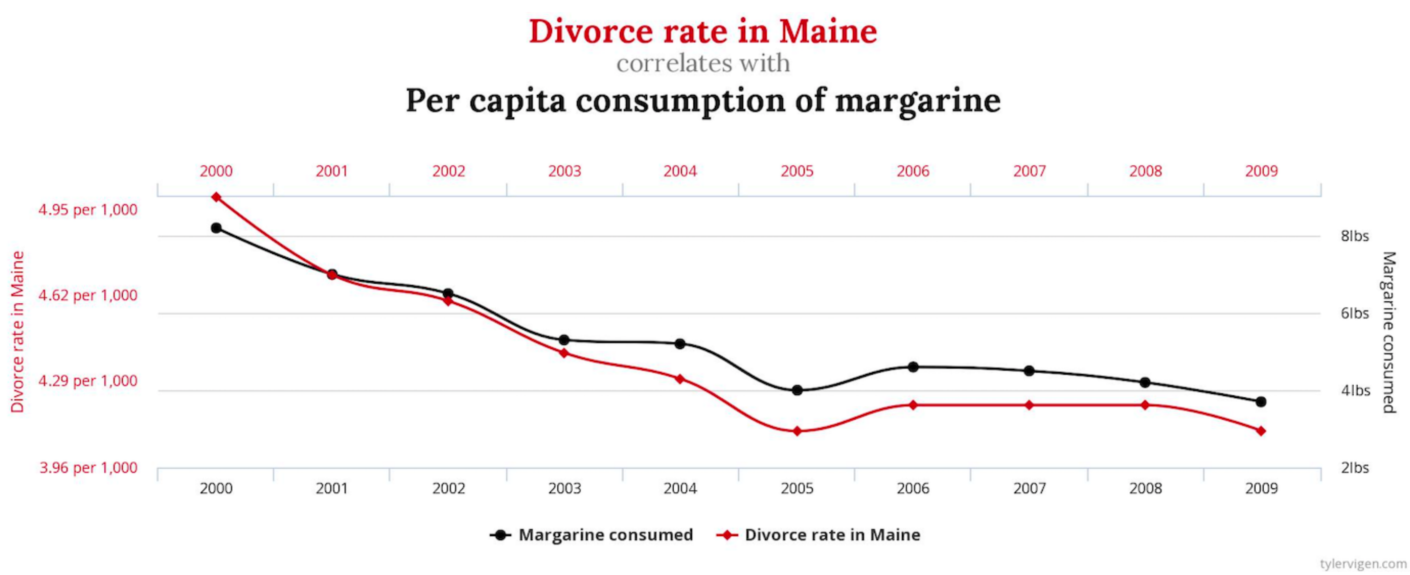
Learning Target

- I can look for connections between two variables to analyze whether or not there is a causal relationship.

Bridge

The following graph[1] displays the divorce rate in Maine and the number of pounds of margarine consumed per person each year. Study the graph.

What do you notice? What do you wonder?



[1] Adapted from Tyler Vigen

Warm-up: Used Car Relationships

Describe the strength and direction of the relationship you expect for each pair of variables. Explain your reasoning.

1. Used car price and original sale price of the car

2. Used car price and number of cup holders in the car

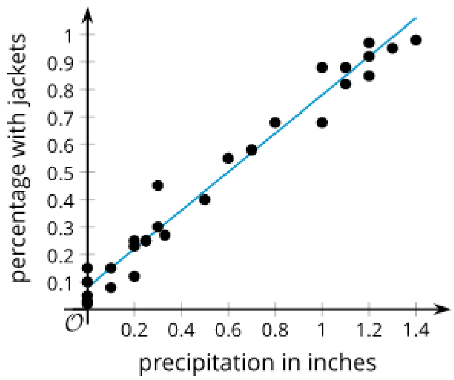
3. Used car price and number of oil changes the car has had

4. Used car price and number of miles the car has been driven

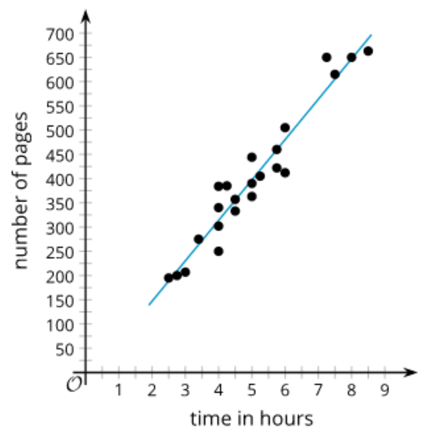
Activity 1: Causation or Association?

Each of the scatter plots shows a strong relationship. Write a sentence or two describing how you think the variables are related.

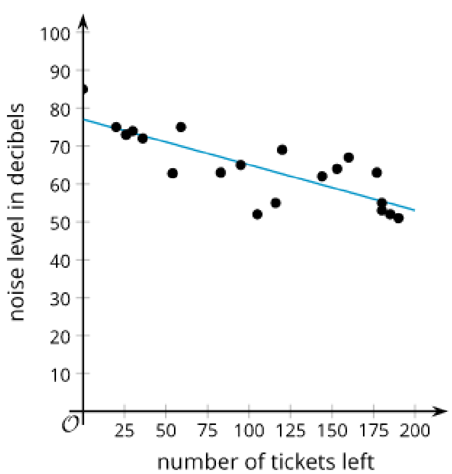
1. During the month of April, Elena keeps track of the number of inches of rain recorded for the day and the percentage of people who come to school with rain jackets.



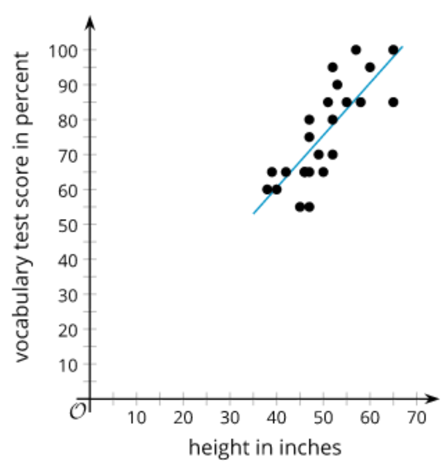
2. A school book club has a list of 100 books for its members to read. They keep track of the number of pages in the books the members read from the list and the amount of time it took to read the book.



3. Number of tickets left for holiday parties at a venue and noise level at the party.



4. The height and score on a test of vocabulary for several children aged 6 to 13.



Activity 2: Find Your Cause

1. Your teacher will give you a set of cards. Individually, categorize each relationship as causation or association without causation. Once finished, share your categorization with your partner.

   For each category, explain to your partner how you know each card belongs in that category. Carefully listen to your partner as they explain their selections. If you disagree, discuss your thinking and work to reach an agreement. After coming to an agreement, record the examples in the agreed-upon category.

Causation	Association without causation

2. If time allows, brainstorm with your partner an example of a relationship that has association but not causation.

Are You Ready For More?

1. Look through news articles or advertisements for claims of causation or association. Find two or three claims and read or watch the articles or the advertisement. Answer these questions for each of the claims.
   a. What is the claim?
   b. What evidence is provided for the claim?
   c. Does there appear to be evidence for causation or correlation? Explain your thinking.
2. Choose the claim with the least or no evidence. Describe an experiment or other way that you could collect data to show correlation or causation.

Activity 3: Fossil Puzzle

An anthropologist finds a fossilized humerus bone of an ancient human ancestor. The humerus is an arm bone running from the shoulder to the elbow. It is 24 centimeters in length.

a. How tall do you think this ancient human was?

b. In order to estimate the height of this ancient human, Kiran decided to measure the humerus of some of his friends and family and see if that would give him an idea.

The table below shows the length of the humerus (in centimeters) and the height (in inches) of 20 of Kiran's friends and family. This data can be found on the Activity 3 tab: https://bit.ly/U4L8DataSets.

Humerus (cm)	33	37	35.5	31.5	30.3	30.5	30.5	28	35	36	28	29	30.5	25.4	33.3	23	26	33	27.8	28.4
Height (in)	64	73	70	67	61.8	67	66	63	66	72	59	62	73	64	68	65	67	75	56	60

c. Use the data Kiran collected to estimate the height of the ancient human.

d. How confident are you in your answer? What information helped you determine your confidence?

e. How did you use mathematics to estimate the height of the ancient human?

Lesson Debrief

Lesson 8 Summary and Glossary

Humans are wired to look for connections and then use those connections to learn about the world around them. One way to notice connections is by looking for a pair of variables with a relationship. For example, if we notice that people who tend to eat many calories also have a higher chance of having a heart attack, we might wonder if lowering our calorie intake would improve our health.

One common mistake people tend to make using statistics is to think that all relationships between variables are causal. Scatter plots can only show a relationship between the two variables. To determine if change in one of the variables actually causes a change in the other variable, or has a causal relationship, the context must be better understood and other options ruled out.

For example, we might expect to see a strong, positive relationship between the number of snowboard rentals and sales of hot chocolate during the months of September through January. This does not mean that an increase in snowboard rentals causes people to purchase more hot chocolate. Nor does it mean that increased sales of hot chocolate cause people to rent snowboards more. More likely there is a third variable, such as colder weather, that might be causing both variables to increase at the same time. In this instance, there is an **association** between snowboard rentals and hot chocolate sales but not necessarily a causation.

Association: A relationship between two or more variables.

On the other hand, sometimes there is a **causal relationship**. A strong, positive relationship between hot chocolate sales and small marshmallow sales may be linked because people buying hot chocolate may want to add small marshmallows to the drink, so an increase in the sales of hot chocolate actually causes the marshmallow sale increase.

Causal relationship: A relationship in which a change in one of the variables causes a change in the other variable.

Unit 4 Lesson 8 Practice Problems

1. Priya creates a scatter plot showing the relationship between the number of steps she takes and her heart rate. The correlation coefficient is 0.88.

    a. Describe the association between Priya’s steps and her heart rate.

    b. Do either of the variables cause the other to change? Explain your reasoning.

2. Kiran creates a scatter plot showing the relationship between the number of students attending drama club and the number of students attending poetry club each week. The correlation coefficient is -0.36.

    a. Describe the association between the variables.

    b. Do either of the variables cause the other to change? Explain your reasoning.

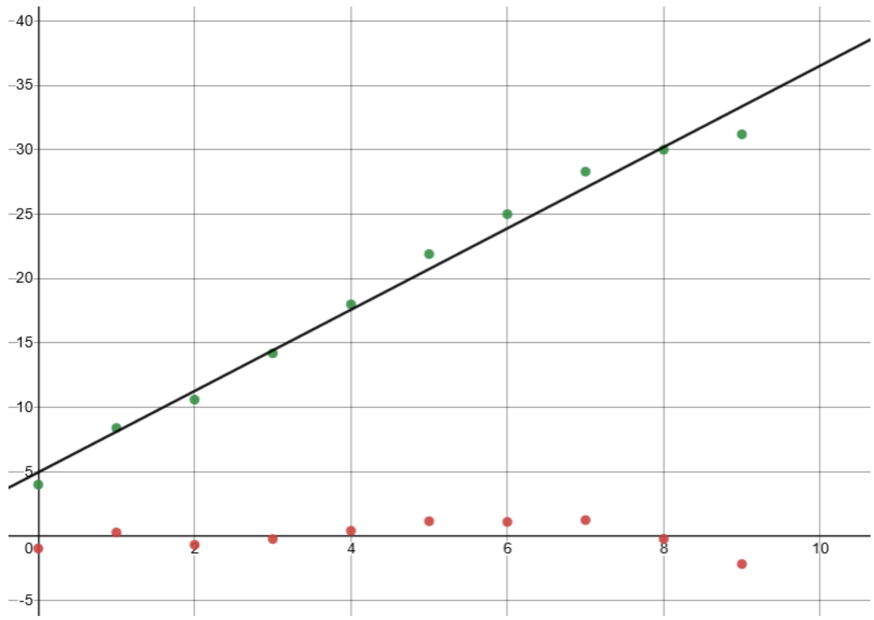
3. A news website shows a scatter plot with a negative relationship between the amount of sugar eaten and happiness levels. The headline reads, “Eating sugar causes happiness to decrease!”

   a. What is wrong with this claim?

   b. What is a better headline for this information?

4. The graph below shows a scatter plot, the line of best fit, and the residual plot.

   Add a point to the scatter plot at $x = 10$, and plot the corresponding residual. How did you decide where to plot your point and the residual?



(From Unit 4, Lesson 7)

5. (*Technology required.*) The following table shows the average yearly price for one gallon of gas from the years 2000–2010, according to the U.S. Energy Information Administration. Data can be found on the right and on the practice problem tab: https://bit.ly/U4L8DataSets.

Year	Price per gallon
2000	\$1.52
2001	\$1.46
2002	\$1.39
2003	\$1.60
2004	\$1.90
2005	\$2.31
2006	\$2.62
2007	\$2.84
2008	\$3.30
2009	\$2.41
2010	\$2.84

   a. Determine the line of best fit and correlation coefficient for the gas prices. Assume the year 2000 is where $x = 0$.

   b. Explain the meaning, in context, of the slope and vertical intercept of the line of best fit.

   c. Is the line of best fit a good model for the data? Explain your reasoning.

(From Unit 4, Lessons 2–4)

6. The number of miles driven, x, and the number of gallons remaining in the gas tank, y, have a strong negative relationship.

   Explain what it means to have a **strong negative** relationship in this context.

(From Unit 4, Lesson 2)

7. (*Technology required.*) Data can be found on the right and on the practice problem tab: https://bit.ly/U4L8DataSets.

x	y
10.2	31
10.4	27
10.5	29
10.5	30
10.5	31
10.6	26
10.8	25
10.8	26
10.9	27
11	24
11.2	22

   a. What is an equation of the line of best fit?

   b. What is the value of the correlation coefficient?

(From Unit 4, Lesson 2)

8. Here are the heights in inches of the players on the boys basketball team at Fanwood High School:

   73, 71, 74, 83, 72, 72, 68, 71

   a. Calculate the five-number summary for the heights of the players on the team.

   b. Which height is a potential outlier? Describe what that means in relation to the other players.

   c. How would the standard deviation be affected if the outlier was removed? How do you know?

(From Unit 1)

Lessons 9 & 10: Mathematical Modeling[1]

Learning Targets

- I can use mathematics to model real-world situations.
- I can test and improve mathematical models for accuracy in representing and predicting real things.

Advice on Modeling

These are some steps that successful modelers often take and questions that they ask themselves. You don't necessarily have to do all of these steps, or do them in order. Only do the parts that you think will help you make progress.

Understand the Question
Think about what the question means before you start making a strategy to answer it. Are there words you want to look up? Does the scenario make sense? Is there anything you want to get clearer on before you start? Ask your classmates or teacher if you need to.

Refine the Question
If necessary, rewrite the question you are trying to answer so that it is more specific.

Estimate a Reasonable Answer
If you don't have enough information to decide what's reasonable, try to come up with an answer that would be too low, and an answer that would be too high.

Identify Unknowns
- What are the meaningful quantities in this situation? Write them down.
- What information would be useful to know? In order to get that information, you could: look it up, take a measurement, or make an assumption.

Gather Information
Write down any of the unknown information that you find. Organize your information in a way that makes sense to you.

Experiment!
Try different ideas to make progress toward answering your question. If you are stuck, think about:
- Helpful ways to organize the information you have or organize your work
- Questions you *can* answer using the information you have
- Ways to represent mathematical relationships or sets of data (tables, equations, graphs, statistical plots)
- Tools that are available for representing mathematics, both digital and analog

Check Your Reasoning
Do you have a first answer to your question? Great! See if it's reasonable.
- Make sure you can explain what the answer means in terms of the original problem.
- Check your precision: Is your answer overly precise? Not precise enough?

Use and Improve Your Model
- Did you make assumptions or measurements? How can you express your model more generally, so that it would work for a range of numbers instead of the specific numbers you used?
- What are the limitations of your model? That is, what are some ways it is not realistic? Does it only work for certain inputs but not others? Are there any meaningful inputs affecting the outcome that are not accounted for? If possible, improve your model to take these into account.
- What are the implications of your model? That is, what should people or organizations do differently or smarter as a result of what your model shows? What would be effective ways to communicate with them?
- What are the areas for further research? That is, what new things are you wondering about that could be investigated, by you or someone else?

[1] Adapted from IM 9–12 Math Algebra 1 Modeling Prompts https://curriculum.illustrativemathematics.org/HS/index.html,

Modeling Rubric

<table>
<tr><th rowspan="2">Skill</th><th colspan="3">Score</th><th rowspan="2">Notes or Comments</th></tr>
<tr><th>Proficient</th><th>Developing</th><th>Needs Revisiting</th></tr>
<tr><td rowspan="2">1. Decide What to Model</td><td>● Assumptions made are clearly identified and justified. Resulting limitations are stated when appropriate.
● Variables of interest are clearly identified and chosen wisely, and appropriate units of measure are used.</td><td>● Assumptions are noted but lacking in justification or difficult to find.
● Variables of interest are noted, but may lack justification, be difficult to find, or not be measured with appropriate units.</td><td>● No assumptions are stated.
● No variables are defined.</td><td></td></tr>
<tr><td colspan="4">To improve at this skill, you could:
● Ask questions about the situation to understand it better
● Check the assumptions you're making to see if they're reasonable (Try asking a friend, or imagining that you're a person involved in the scenario. Would those assumptions make sense to you?)
● Double-check the variables you've identified: Are there other quantities in the situation that could vary? Is there something you've identified as a variable that is actually fixed or determined? (Remember that more abstract things like time and speed are also quantities.)</td></tr>
<tr><td rowspan="2">2. Formulate a Mathematical Model</td><td>● An appropriate model is chosen and represented clearly.
● Diagrams, graphs, etc. are clear and appropriately labeled.</td><td>Parts of the model are unclear, incomplete, or contain mistakes.</td><td>No model is presented, or the presentation contains significant errors.</td><td></td></tr>
<tr><td colspan="4">To improve at this skill, you could:
● Check your model more carefully to make sure it really fits well
● Consider a wider variety of possible models, to find one that fits the situation better
● Think about the situation more deeply before trying to find a model
● Convince a skeptic: Pretend that you think your model is inadequate, or ask a friend to pretend to be skeptical of it. What would a skeptic find wrong with your model? Try to fix those things, or explain why they're not actually problems.</td></tr>
</table>

<table>
<tr><th rowspan="2">Skill</th><th colspan="3">Score</th><th rowspan="2">Notes or Comments</th></tr>
<tr><th>Proficient</th><th>Developing</th><th>Needs Revisiting</th></tr>
<tr><td rowspan="2">3. Use Your Model to Reach a Conclusion</td><td>● Solution is relevant to the original problem.
● Reader can easily understand the reasoning leading to the solution.
● Relevant details are included like units of measure.</td><td>Solution is not well-aligned to the original problem, or aspects of the solution are difficult to understand or incomplete.</td><td>No solution is provided.</td><td></td></tr>
<tr><td colspan="4">To improve at this skill, you could:
● Double-check your calculations: Show them to someone else to see if they agree, or take a break and look at your calculations again later
● Make sure your calculations are justified by your model: Ask yourself how you decided what to calculate, and see if your reasoning matches up with your model
● Think more deeply about what your conclusions mean in the original scenario: Imagine you're a person involved in the scenario, or explain your conclusions to someone else and see if they have questions</td></tr>
<tr><td rowspan="2">4. Refine and Share Your Model</td><td>● The model’s implications are clearly stated.
● The limitations of the model and solution are addressed.</td><td>The limitations of the model and solution are addressed but lacking in depth or ignoring key components.</td><td>No interpretation of model and solution is provided.</td><td></td></tr>
<tr><td colspan="4">To improve at this skill, you could:
● Think more creatively about what your conclusions mean: Ask yourself "If I was involved in this situation, what would I understand better because of these conclusions? What would I want to do next?"
● Be skeptical of your model: What don't you like about it, and what can you do to fix those things?
● Explain your model to someone else: Tell them how it works and why it's good. If you're not sure how it works or why it's good, you might need to change it.</td></tr>
</table>

Workspace for Modeling Prompt # _____

Workspace for Modeling Prompt # _____

Modeling Prompt # ____ Reflection

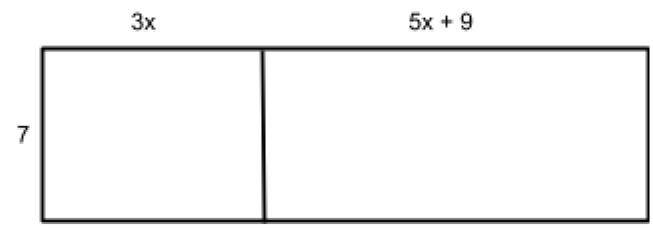
Lesson 11: Post-Test Activities

Learning Targets

- I understand the reasoning for and will strive to meet the expectations communicated by my teacher.
- I can add and subtract linear expressions.

Activity 2: Area Model

Rectangle A



1. Write an expression for the area of the entire rectangle A.

   Describe the steps you took and show your work.

2. Double the vertical length of rectangle A to form a new rectangle and call it rectangle B.

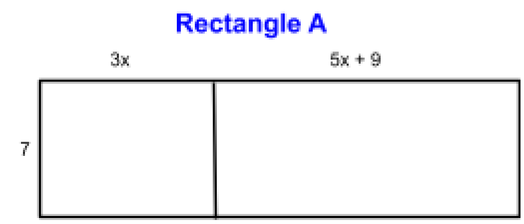
   Write an expression that represents the area of rectangle B. Describe the steps you took and show your work.

3. Triple the vertical length of rectangle A to form a new rectangle and call it rectangle C.

   Write an expression that represents the area of rectangle C. Describe the steps you took and show your work.

4. Multiply the vertical length of rectangle A times a factor x to form a new rectangle and call it rectangle D.

   Write an expression that represents the area of rectangle D. Describe the steps you took and show your work.



5. What patterns do you notice in the areas of rectangles A-D? What do you wonder?

6. Add 15 units to each section of rectangle A to form a new rectangle and call it rectangle L.

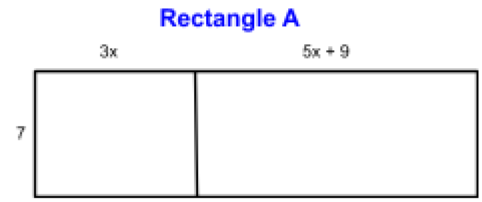
   Write an expression that represents the area of rectangle L. Describe the steps you took and show your work.

7. Add 50 units to each section of rectangle A to form a new rectangle and call it rectangle M.

   Write an expression that represents the area of rectangle M. Describe the steps you took and show your work.

8. Add x units to each section of rectangle A to form a new rectangle and call it rectangle N.

   Write an expression that represents the area of rectangle N. Describe the steps you took and show your work.



9. What patterns do you notice in the areas of rectangles A, L, M, and N? Describe what you know now.

10. If time is available, choose a few tasks from below to solve:

    a. Write at least two equivalent expressions for the circumference of the circle with a radius of $6x - 3$ meters.

    b. Simplify:

    $$(5x - 8.3) + (2.1x + 3.9) + (4.7 - 3.3x) - (3.2x + 2)$$

    c. Find the perimeter of this trapezoid: