Lesson 4

Fitting a Line to Data

Goals

- Compare and contrast (orally) values in a data set with predictions made using a given line.
- Comprehend that a model of data, such as a line of fit, can be used to predict values that are not given in the data.
- Identify (orally) obvious outliers on a scatter plot.

Learning Targets

- I can pick out outliers on a scatter plot.
- I can use a model to predict values for data.

Lesson Narrative

In this lesson, student focus becomes more holistic and they begin to see a set of data points as a single thing that can be analyzed, not just as a bunch of disconnected points. For the first time students see that sometimes the relationship between two variables can be modeled with a line, although they continue to analyze the connections between the scatter plot and the line by comparing individual points. As they zoom out and see these connections, students begin to see the structure of the scatter plot and to use that structure to reason abstractly and quantitatively to connect the points to real situations.

Student Learning Goal

Let's look at the scatter plots as a whole.

Access for Students with Diverse Abilities

- Action and Expression (Activity 1)

Access for Multilingual Learners

- MLR8: Discussion Supports (Activity 1, Activity 2)

Instructional Routines

- Notice and Wonder

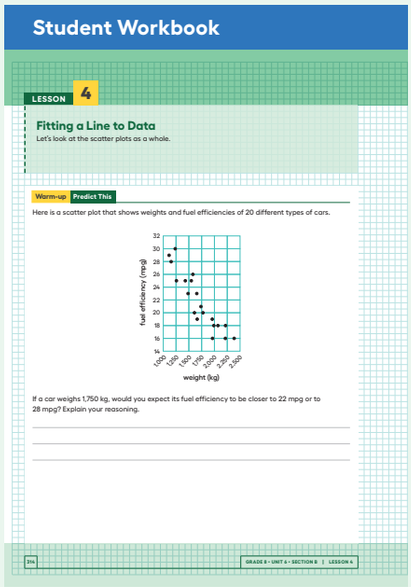
Required Preparation

Activity 2:
For the digital version of the activity, acquire devices that can run the applet.

Activity 3:
For the digital version of the activity, acquire devices that can run the applet.

Lesson Timeline

Warm-up	Activity 1	Activity 2	Lesson Synthesis	Assessment: Cool-down
5 min	15 min	15 min	10 min	5 min


Student Workbook

LESSON 4

Fitting a Line to Data

Let's look at the scatter plots as a whole.

Warm-up Predict This

Here is a scatter plot that shows weights and fuel efficiencies of 20 different types of cars.

If a car weighs 1,750 kg, would you expect its fuel efficiency to be closer to 22 mpg or to 28 mpg? Explain your reasoning.

Warm-up

Predict This

5 min

Activity Narrative

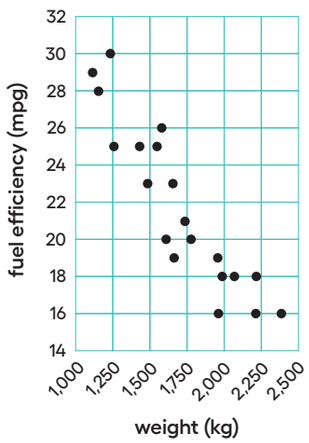
In this *Warm-up*, students are asked to think about a point in the context of the entire data set. They compare two different possible predictions for the dependent variable, given a value for the independent variable. This sets them up to understand and interpret a linear model for the relationship between independent and dependent variables. They use the potential location of a point not included in the scatter plot to answer a question about the context, based on the visual structure of the scatter plot.

Launch

Arrange students in groups of 2. Give 1 minute of quiet work time followed by 1 minute to check their solution with their partner. Follow with a whole-class discussion.

Student Task Statement

Here is a scatter plot that shows weights and fuel efficiencies of 20 different types of cars.



If a car weighs 1,750 kg, would you expect its fuel efficiency to be closer to 22 mpg or to 28 mpg? Explain your reasoning.

We expect the fuel efficiency to be closer to 22 mpg. There are several cars close to that weight, and their fuel efficiency is between 18 and 22 mpg. The cars that have a fuel efficiency close to 28 mpg have a weight less than 1,250 kg, although there is one car that has a weight greater than 1,500 kg and a fuel efficiency of 26 mpg. But that one has a higher fuel efficiency than any other car with a weight between 1,500 and 1,750 kg.

Activity Synthesis

Display the graph for all to see. Poll the class to see if they think the fuel efficiency is closer to 22 mpg or 28 mpg. If they are all in agreement that the answer is closer to 22 mpg, ask a few students to share their reasoning. If there is disagreement, ask students to share their reasoning and come to an agreement. If it does not come up in the discussion, ask students to look at cars whose fuel efficiency is close to 28 mpg and note that their weights are quite a bit less. Then look at cars with a weight close to 1,750 kg, and note that their fuel efficiency is between 18 and 22 mpg. As a whole class, decide where to plot both potential points, and point out that 1 is close to the other nearby values and 1 is very far away.

Activity 1

Battery Life

15 min

Activity Narrative

There is a digital version of this activity.

In this lesson, the meaning of the words "model" and "modeling" are explained in terms of a linear model. Students are not expected to define the words, but should be comfortable understanding and using them. A model is used to predict future battery life of a device not included in the data, as well as compare existing data points to the model.

In the digital version of the activity, students use an applet to analyze a scatter plot with a linear model. The applet allows students to interact with the linear model and data to answer questions.

Launch

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

To introduce the context, tell students that batteries in rechargeable devices will lose capacity over time and the device will not work for as long. In the experiment from this activity, when the devices were first bought, they were left on until they turned off automatically from not having enough battery to continue running. Then the devices were used regularly and the experiment was run again 1 year later.

Ask students,

- *"Do you think battery life 1 year later will be longer, about the same, or shorter than when the device is new?"*

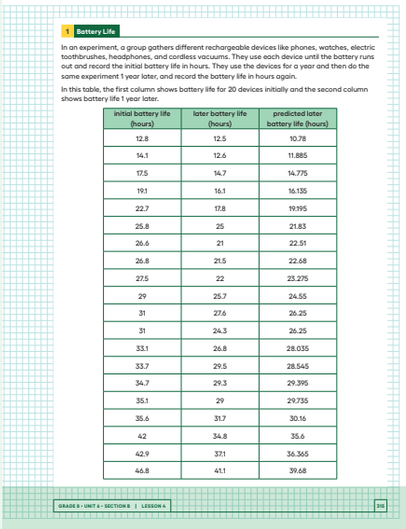
  It will be shorter because the battery wears out through use.

Access for Students with Diverse Abilities (Activity 1, Student Task)

Action and Expression: Develop Expression and Communication. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, "It looks like ..." and "We are trying to ..."

Supports accessibility for: Language, Organization

Student Workbook



1 Battery Life

In an experiment, a group gathers different rechargeable devices like phones, watches, electric toothbrushes, headphones, and cordless vacuums. They use each device until the battery runs out and record the initial battery life in hours. They use the devices for a year and then do the same experiment 1 year later, and record the battery life in hours again.

In this table, the first column shows battery life for 20 devices initially and the second column shows battery life 1 year later.

initial battery life (hours)	later battery life (hours)	predicted later battery life (hours)
12.8	12.5	10.78
14.1	12.6	11.885
17.5	14.7	14.775
19.1	16.1	16.135
22.7	17.8	19.195
25.8	25	21.83
26.6	21	22.51
26.8	21.5	22.68
27.5	22	23.275
29	25.7	24.55
31	27.6	26.25
31	24.3	26.25
33.1	26.8	28.035
33.7	29.5	28.545
34.7	29.3	29.395
35.1	29	29.735
35.6	31.7	30.16
42	34.8	35.6
42.9	37.1	36.365
46.8	41.1	39.68

GRADE 8 • UNIT 6 • SECTION B | LESSON 4 316

"What are some things that you think affect how the battery life changes from 1 year of use?"

the materials used in the battery, how much the device is used during the year, the type of device

Tell students that they will look at some data for battery life in rechargeable devices. Previously, they have used mathematics to analyze real-world situations, identifying variables in a situation and describing their relationships mathematically. This process is called "modeling," and the mathematical description is called a "model." Sometimes students make assumptions about the situation or ignore some features so the model is simpler.

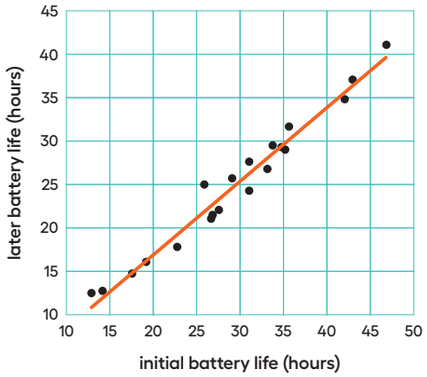
Student Task Statement

In an experiment, a group gathers different rechargeable devices like phones, watches, electric toothbrushes, headphones, and cordless vacuums. They use each device until the battery runs out and record the initial battery life in hours. They use the devices for a year and then do the same experiment 1 year later, and record the battery life in hours again.

In this table, the first column shows battery life for 20 devices initially and the second column shows battery life 1 year later.

initial battery life (hours)	later battery life (hours)	predicted later battery life (hours)
12.8	12.5	10.78
14.1	12.6	11.885
17.5	14.7	14.775
19.1	16.1	16.135
22.7	17.8	19.195
25.8	25	21.83
26.6	21	22.51
26.8	21.5	22.68
27.5	22	23.275
29	25.7	24.55
31	27.6	26.25
31	24.3	26.25
33.1	26.8	28.035
33.7	29.5	28.545
34.7	29.3	29.395
35.1	29	29.735
35.6	31.7	30.16
42	34.8	35.6
42.9	37.1	36.365
46.8	41.1	39.68

The scatter plot shows the battery life measurements for 20 devices together with the graph of $y = 0.85x - 0.1$.



The function described by the equation $y = 0.85x - 0.1$ is a *model* of the relationship between the initial and later battery life for a device.

This model *predicts* the later battery life from the device's initial battery life. These predicted battery lives are shown in the third column of the table.

1. Two devices that both have an initial battery life of 31 hours have different later battery life. What are their later battery lives? How can you see this in the table? How can you see this in the graph?

   27.6 hours and 24.3 hours

   You can see this in the 2 rows in the table that show 31 in the first column. You can see this in the graph by finding the 2 points with $x = 31$.

2. The model predicts that when the initial battery life is 31 hours, the later battery life will be 26.25. How can you see this in the graph? How can you see this using the equation?

   The point $(31, 26.25)$ is on the graph of the line. We can substitute $x = 31$ into the equation to get $y = 0.85 \cdot 31 - 0.1$ which gives $y = 26.25$

3. One of the devices has an initial battery life of 29 hours. What does the model predict for its later battery life? How does that compare to the actual later battery life?

   The model predicts that the later battery life will be 24.55 hours, because $y = 0.85 \cdot 29 - 0.1$, which gives $y = 24.55$. The actual later battery life is 25.7 hours. The actual later battery life is 1.15 hours greater than the prediction, because $25.7 - 24.55 = 1.15$.

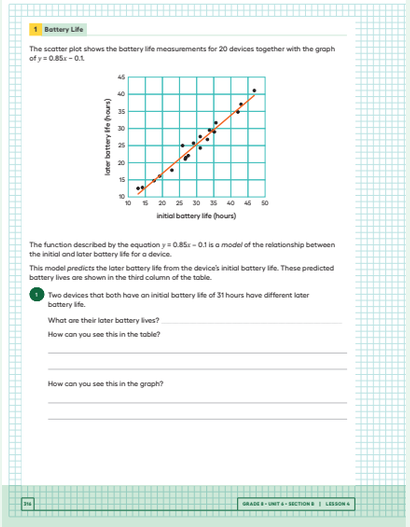
4. Find a device for which the model makes a very good prediction of the actual later battery life. How can you see this in the table? In the graph?

   Sample response: The best prediction is made for the device with an initial battery life of 19.1 hours where the actual later battery life is only 0.035 less than the predicted later battery life. This can be seen in the table in the row with 19.1 in the first column. The corresponding point in the scatter plot is very close to the line.

5. Find a device for which the model's prediction is not very close to the actual later battery life. How can you see this in the table? In the graph?

   Sample response: The prediction for the device that has an initial battery life of 25.8 hours is off by more than 3 hours. This can be seen in the sixth row with a 25.8 in the first column. The corresponding point in the scatter plot is relatively far above the line.

Student Workbook


1 Battery Life

The scatter plot shows the battery life measurements for 20 devices together with the graph of $y = 0.85x - 0.1$.

later battery life (hours)
initial battery life (hours)

The function described by the equation $y = 0.85x - 0.1$ is a *model* of the relationship between the initial and later battery life for a device.

This model predicts the later battery life from the device's initial battery life. These predicted battery lives are shown in the third column of the table.

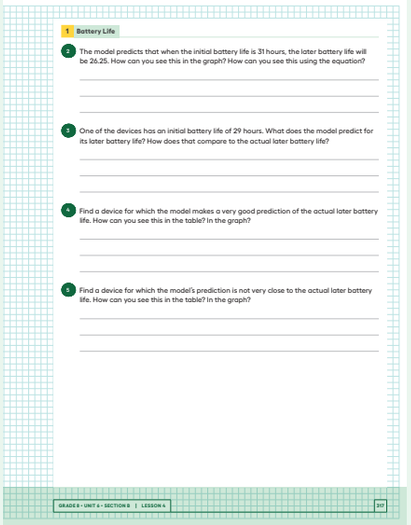
1. Two devices that both have an initial battery life of 31 hours have different later battery life.

What are their later battery lives?

How can you see this in the table?

How can you see this in the graph?

Student Workbook


1 Battery Life

2. The model predicts that when the initial battery life is 31 hours, the later battery life will be 26.25. How can you see this in the graph? How can you see this using the equation?

3. One of the devices has an initial battery life of 29 hours. What does the model predict for its later battery life? How does that compare to the actual later battery life?

4. Find a device for which the model makes a very good prediction of the actual later battery life. How can you see this in the table? In the graph?

5. Find a device for which the model's prediction is not very close to the actual later battery life. How can you see this in the table? In the graph?

Access for Multilingual Learners (Activity 1, Synthesis)

MLR8: Discussion Supports. Display sentence frames to support students as they work with their partner and explain their reasoning: "I think ____ because _____." or "I (agree/disagree) because _____." *Advances: Speaking, Conversing*

Instructional Routines

Notice and Wonder

ilclass.com/r/10694948
Please log in to the site before using the QR code or URL.

Activity Synthesis

The goal of this discussion is to help students understand the relationship between the data and a linear model of the data.

To highlight the relationship, ask:

- *"What does a point in the scatter plot represent?"*

  the actual initial and later battery life of a device

- *"What does the line represent?"*

  the predicted later battery life based on the initial battery life of a device

- *"What does it mean when a point is close to the line? When it is far away from the line?"*

  When it is close, the model predicts the later battery life well. The farther away in the vertical direction, the worse the prediction.

- *"How can you use the graph to predict the later battery life of a device that has an initial battery life of 20 hours? How can you use the equation?"*

  Find the point on the line that lines up with $x = 20$ or substitute $x = 20$ into the model equation.

Activity 2

The Agony of the Feet

15 min

Activity Narrative

There is a digital version of this activity.

A scatter plot is shown and the points interpreted in context. Later, a linear model is graphed with the scatter plot to help students see an obvious outlier. In the discussion, the term **outlier** is introduced.

In the digital version of the activity, students use an applet to analyze a scatter plot with a linear model. The applet allows students to more easily examine the data and linear model. Use the digital version if available to allow students to focus on analyzing the scatter plot without worrying about construction.

Launch

Keep students in groups of 2. Display the image of the feet.

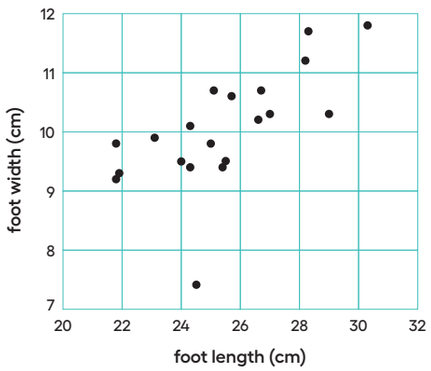
Ask students,

"*What do you notice? What do you wonder?*"

Ideally, they will notice that these are pictures of a foot of 3 different people. All 3 feet are approximately the same length, but different widths. Tell students that all of these feet are a size 8. However, they wouldn't all necessarily find the same shoe equally comfortable, because of the varying widths. (Besides the numerical size, some shoes also come in different widths.) Students should understand that human feet can vary in both length and width.

Student Task Statement

Here is a scatter plot that shows lengths and widths of 20 different left feet.



1. Estimate the widths of the longest foot and the shortest foot.

   The longest foot has a width of about 11.9 cm. The shortest foot has a width of about 9.2 cm.

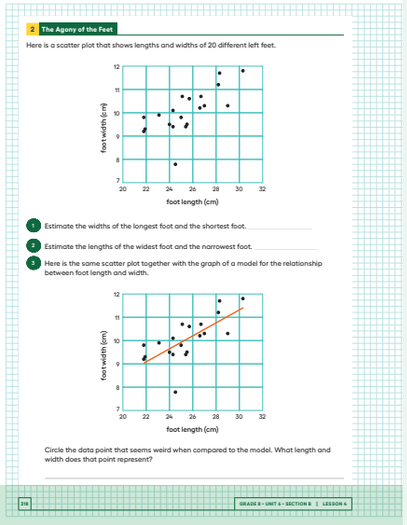
2. Estimate the lengths of the widest foot and the narrowest foot.

   The widest foot has a length of about 30.4 cm. The narrowest foot has a length of about 24.5 cm.

Access for Multilingual Learners (Activity 2, Launch)

MLR8: Discussion Supports.
Use multimodal examples to show the meaning of width and length. Use verbal descriptions along with gestures, drawings, or concrete objects to show their meaning on the image.
Advances: Listening, Representing

Student Workbook


2 The Agony of the Feet

Here is a scatter plot that shows lengths and widths of 20 different left feet.

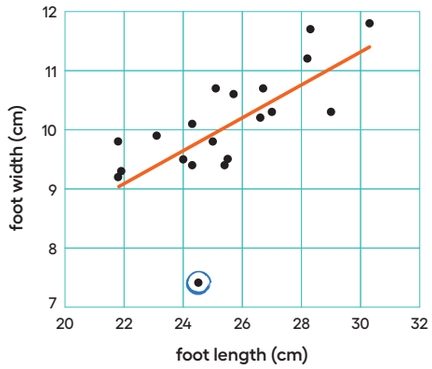
foot width (cm)

foot length (cm)

1. Estimate the widths of the longest foot and the shortest foot.
2. Estimate the lengths of the widest foot and the narrowest foot.
3. Here is the same scatter plot together with the graph of a model for the relationship between foot length and width.

Circle the data point that seems weird when compared to the model. What length and width does that point represent?

3. Here is the same scatter plot together with the graph of a model for the relationship between foot length and width.



Circle the data point that seems weird when compared to the model. What length and width does that point represent?

The point that seems weird represents a length of about 24.5 cm and a width of about 7.4 cm.

Activity Synthesis

Introduce the term **outlier**. An outlier is a point that is separated from the rest of the data. Sometimes data sets have outliers. There could be an outlier because there really is a data point that is very different than the others. Or, there could be an error in collecting or entering the data. When there are outliers, one has to make a judgment about whether to include it in the analysis or not.

Lesson Synthesis

To help students see the connection between a scatter plot and a linear model for a data set, ask:

- *"What kind of* model *for a data set did we investigate today?"*

  A linear model

- *"What does this kind of model help us do?"*

  see the trend in the data more clearly and make predictions

- *"What does it mean when a data point is closer to the line in the vertical direction that represents a linear model? What does it mean when a data point is farther from the line in the vertical direction?"*

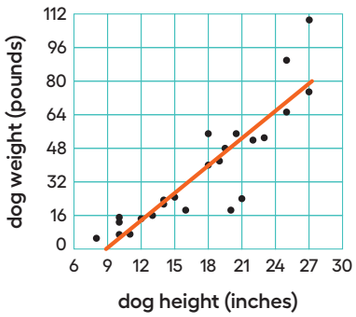
  When the data point is closer to the line, it represents data that fits the prediction well. When the point is farther from the line, the data does not fit the prediction well.

- *"In your own words, how can you identify an outlier in a scatter plot?"*

  A point that is far from the other points in the scatter plot represents an outlier.

Lesson Summary

Sometimes, we can use a linear function as a model of the relationship between two variables. For example, here is a scatter plot that shows heights and weights of 25 dogs together with the graph of a linear function which is a model for the relationship between a dog's height and its weight.

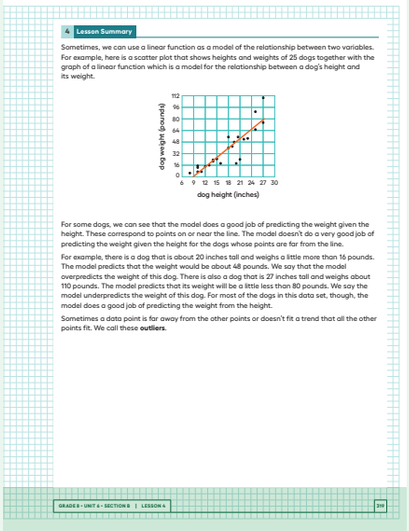


For some dogs, we can see that the model does a good job of predicting the weight given the height. These correspond to points on or near the line. The model doesn't do a very good job of predicting the weight given the height for the dogs whose points are far from the line.

For example, there is a dog that is about 20 inches tall and weighs a little more than 16 pounds. The model predicts that the weight would be about 48 pounds. We say that the model overpredicts the weight of this dog. There is also a dog that is 27 inches tall and weighs about 110 pounds. The model predicts that its weight will be a little less than 80 pounds. We say the model underpredicts the weight of this dog. For most of the dogs in this data set, though, the model does a good job of predicting the weight from the height.

Sometimes a data point is far away from the other points or doesn't fit a trend that all the other points fit. We call these **outliers**.

Student Workbook


4 Lesson Summary

Sometimes, we can use a linear function as a model of the relationship between two variables. For example, here is a scatter plot that shows heights and weights of 25 dogs together with the graph of a linear function which is a model for the relationship between a dog's height and its weight.

dog weight (pounds)
112 96 80 64 48 32 16 0
6 9 12 15 18 21 24 27 30
dog height (inches)

For some dogs, we can see that the model does a good job of predicting the weight given the height. These correspond to points on or near the line. The model doesn't do a very good job of predicting the weight given the height for the dogs whose points are far from the line.

For example, there is a dog that is about 20 inches tall and weighs a little more than 16 pounds. The model predicts that the weight would be about 48 pounds. We say that the model overpredicts the weight of this dog. There is also a dog that is 27 inches tall and weighs about 110 pounds. The model predicts that its weight will be a little less than 80 pounds. We say the model underpredicts the weight of this dog. For most of the dogs in this data set, though, the model does a good job of predicting the weight from the height.

Sometimes a data point is far away from the other points or doesn't fit a trend that all the other points fit. We call these **outliers**.

GRADE 8 • UNIT 6 • SECTION B | LESSON 4 219

Responding To Student Thinking

More Chances
Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Cool-down

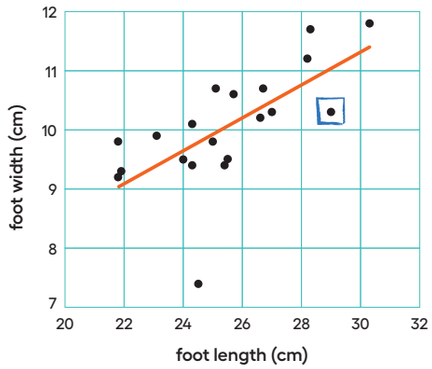
A 1 Foot Foot

5 min

Given a scatter plot, students find the point that is closest to a given value and then compare this point to the predicted value based on a linear model.

Student Task Statement

Here is a scatter plot that shows lengths and widths of 20 left feet, together with the graph of a model of the relationship between foot length and width.



1. Draw a box around the point that represents the foot with length closest to 29 cm.
2. What is the approximate width of this foot?

   About 10.4 cm

3. What width does the model predict for a foot with length 29 cm?

   About 11.1 cm

Lesson 4

Practice Problems

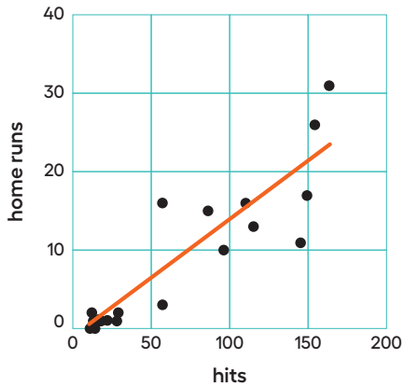
2 Problems

Problem 1

The scatter plot shows the number of hits and home runs in 2015 for 20 baseball players who had at least 10 hits last season. The table shows the values for 15 of those players.

The model, represented by $y = 0.15x - 1.5$, is graphed with a scatter plot.

hits	home runs	predicted home runs
12	2	0.3
22	1	1.8
154	26	21.6
145	11	20.3
110	16	15
57	3	7.1
149	17	20.9
29	2	2.9
13	1	0.5
18	1	1.2
86	15	11.4
163	31	23
115	13	15.8
57	16	7.1
96	10	12.9



Use the graph and the table to answer the questions.

a. Player A had 154 hits in 2015. How many home runs did he have? How many was he predicted to have?

Home runs: 26; predicted home runs: 21.6

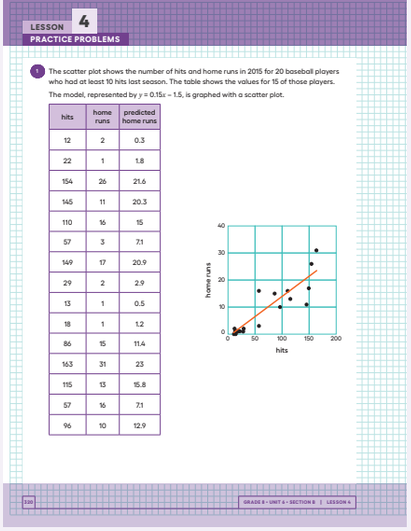
b. Player B was the player who most outperformed the prediction. How many hits did Player B have last season?

57

c. What would you expect to see in the graph for a player who had many fewer home runs than the model predicted?

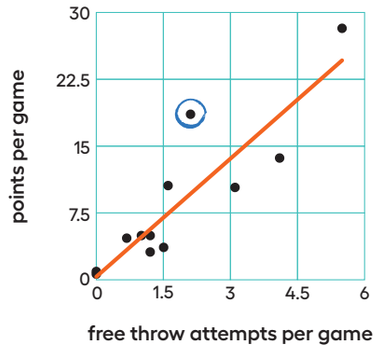
Sample response: The point should be much lower on the scatter plot than the point on the line with the same x-value.

Student Workbook



Problem 2

Here is a scatter plot that compares points per game to free throw attempts per game for basketball players in a tournament. The model, represented by $y = 4.413x + 0.377$, is graphed with the scatter plot. Here, x represents free throw attempts per game, and y represents points per game.



a. Circle any data points that appear to be outliers.

b. What does it mean for a point to be far above the line in this situation?

Sample response: A point above the line represents a player who scores more points per game than predicted by their number of free throw attempts.

c. Based on the model, how many points per game would you expect a player who attempts 4.5 free throws per game to have? Round your answer to the nearest tenth of a point per game.

20.2 points per game, because 4.413(4.5) + 0.377 is roughly equal to 20.2.

d. One of the players scored 13.7 points per game with 4.1 free throw attempts per game. How does this compare to what the model predicts for this player?

Sample response: The player is scoring less than the model predicts they should. The model predicts that with 4.1 free throw attempts per game, the player should score 4.413(4.1) + 0.377, or about 18.5, points per game.