



Lesson 15: Interpreting Residuals from a Line

Student Outcomes

- Students use the least squares line to predict values for a given data set.
- Students use residuals to evaluate the accuracy of predictions based on the least squares line.

Lesson Notes

Students continue their exploration of residuals. Lesson 14 introduced residuals as the rationale for choosing the best-fit line. In this lesson, students calculate prediction errors for a data set and generalize residuals to predict the error of a value not included in the original data set.

Classwork

Example 1 (3 minutes): Calculating Prediction Errors

Introduce the data and plot from the opening paragraph.

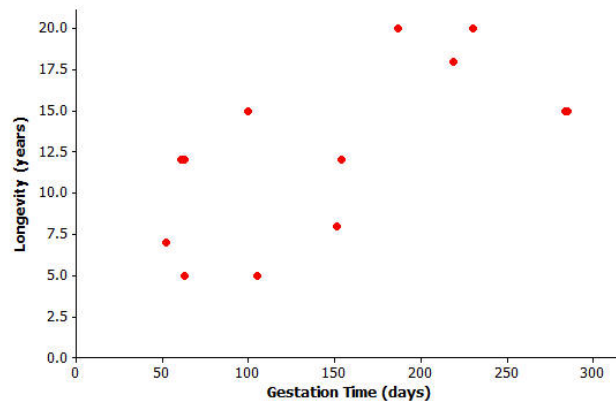
Example 1: Calculating Prediction Errors

The gestation time for an animal is the typical duration between conception and birth. The longevity of an animal is the typical lifespan for that animal. The gestation times (in days) and longevity (in years) for 13 types of animals are shown in the table below.

Animal	Gestation Time (days)	Longevity (years)
Baboon	187	20
Black Bear	219	18
Beaver	105	5
Bison	285	15
Cat	63	12
Chimpanzee	230	20
Cow	284	15
Dog	61	12
Fox (Red)	52	7
Goat	151	8
Lion	100	15
Sheep	154	12
Wolf	63	5

Data Source: *Core Math Tools*, www.nctm.org

Here is the scatter plot for this data set:



Exercises 1–4 (10 minutes)

Let students work independently on finding the regression line in Exercise 1.

Ask students to examine the scatter plot with the regression line. Then discuss:

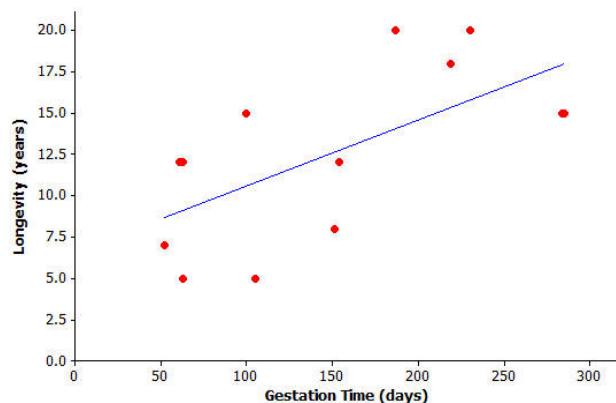
- How accurate will this line be at predicting the longevity of types of animals?
- Encourage students to look at the vertical distances of the points in the plot from the line—smaller distances lead to more accurate predictions.

Exercises 1–4

Finding the equation of the least squares line relating longevity to gestation time for these types of animals provides the equation to predict longevity. How good is the line? In other words, if you were given the gestation time for another type of animal not included in the original list, how accurate would the least squares line be at predicting the longevity of that type of animal?

- Using a graphing calculator, verify that the equation of the least squares line is $y = 6.642 + 0.03974x$, where x represents the gestation time (in days) and y represents longevity (in years).

The least squares line has been added to the scatter plot below.



Let students work independently on Exercises 2–4. Then discuss and confirm as a class.

2. Suppose a particular type of animal has a gestation time of 200 days. Approximately what value does the line predict for the longevity of that type of animal?

$$y = 6.642 + 0.03974(200) = 14.59 \approx 14.6$$

3. Would the value you predicted in Exercise 2 necessarily be the exact value for the longevity of that type of animal? Could the actual longevity of that type of animal be longer than predicted? Could it be shorter?

We do not expect the longevity to be exactly 14.6 years. The actual longevity for this type of animal might be longer or shorter than 14.6 years.

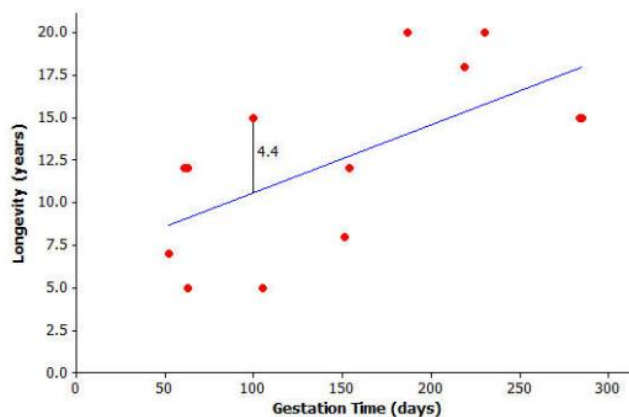
You can investigate further by looking at the types of animals included in the original data set. Take the lion, for example. Its gestation time is 100 days. You also know that its longevity is 15 years, but what does the least squares line *predict* for the lion's longevity?

Substituting $x = 100$ days into the equation, you get $y = 6.642 + 0.03974(100)$ or approximately 10.6. The least squares line predicts the lion's longevity to be approximately 10.6 years.

4. How close is this to being correct? More precisely, how much do you have to add to 10.6 to get the lion's true longevity of 15?

We predicted 10.6 years for the longevity of the lion, and students are asked what needs to be added to this value in order to get the lion's actual longevity of 15 years. Students can do this calculation mentally by adding amounts to 10.6 until they get 15. It is important, then, to see the connection between this adding calculation and the subtraction $15 - 10.6$. The result of this calculation, 4.4, is a residual.

You can show the prediction error of 4.4 years on the graph like this:



MP.4

Exercises 5–6 (10 minutes)

Let students work in pairs on Exercises 5 and 6. Then discuss and confirm as a class.

Exercises 5–6

5. Let's continue to think about the gestation times and longevity of animals. Let's specifically investigate how accurately the least squares line predicted the longevity of the black bear.

- a. What is the gestation time for the black bear?

219 days

- b. Look at the graph. Roughly what does the least squares line predict for the longevity of the black bear?

Around 15 years

- c. Use the gestation time from part (a) and the least squares line $y = 6.642 + 0.03974x$ to predict the black bear's longevity. Round your answer to the nearest tenth.

15.3 years

- d. What is the actual longevity of the black bear?

18 years

- e. How much do you have to add to the predicted value to get the actual longevity of the black bear?

2.7 years

- f. Show your answer to part (e) on the graph as a vertical line segment.

The student should show a vertical line segment of length 2.7 from the least squares line to the point representing the black bear.

6. Repeat this activity for the sheep.

- a. Substitute the sheep's gestation time for x into the equation to find the predicted value for the sheep's longevity. Round your answer to the nearest tenth.

$$y = 6.642 + 0.03974(154) = 12.8 \text{ years}$$

- b. What do you have to add to the predicted value in order to get the actual value of the sheep's longevity? (Hint: Your answer should be negative.)

$$12 - 12.8 = -0.8 \text{ years}$$

- c. Show your answer to part (b) on the graph as a vertical line segment. Write a sentence describing points in the graph for which a negative number would need to be added to the predicted value in order to get the actual value.

The student should show a vertical line segment of length 0.8 from the least squares line to the point representing the sheep. For points that fall below the least squares line you would need to add a negative number to the predicted value in order to get the actual value.

Example 2 (5 minutes): Residuals as Prediction Errors

Explain the concept of residuals and show students how the residuals in the table have been calculated.

To be sure students understand what the residual represents, ask the following:

- What do these residuals tell us?
 - *How much needs to be added to the predicted longevity to get to the actual longevity.*
- Which type of animal has the largest positive residual? What does this tell us?
 - *Baboon, the actual longevity is 5.9 years more than the predicted value.*
- Which type of animal has the “biggest” negative residual? What does this tell us?
 - *Beaver, the actual longevity is 5.8 years less than the predicted value.*

MP.2

Discuss the scenario presented in the text; discuss how to find the predicted longevity of an animal not included in the data.

Example 2: Residuals as Prediction Errors

In each exercise above, you found out how much needs to be added to the predicted value to find the true value of an animal's longevity. In order to find this, you have been calculating

$$\text{actual value} - \text{predicted value.}$$

This quantity is referred to as a residual. It is summarized as

$$\text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value.}$$

You can now work out the residuals for all of the points in our animal longevity example. The values of the residuals are shown in the table below.

Animal	Gestation Time (days)	Longevity (years)	Residual
Baboon	187	20	5.9
Black Bear	219	18	2.7
Beaver	105	5	-5.8
Bison	285	15	-3.0
Cat	63	12	2.9
Chimpanzee	230	20	4.2
Cow	284	15	-2.9
Dog	61	12	2.9
Fox (Red)	52	7	-1.7
Goat	151	8	-4.6
Lion	100	15	4.4
Sheep	154	12	-0.8
Wolf	63	5	-4.1

These residuals show that the actual longevity of an animal should be within six years of the longevity predicted by the least squares line.

Suppose you selected a type of animal that is not included in the original data set, and the gestation time for this type of animal is 270 days. Substituting $x = 270$ into the equation of the least squares line you get

$$\begin{aligned} y &= 6.642 + 0.03974(270) \\ &= 17.4 \text{ years.} \end{aligned}$$

Exercises 7–8 (5 minutes)

Let students work in pairs on Exercises 7–8. Then discuss and confirm as a class. Once students have completed these two questions, tell them that 270 days is the gestation time for humans and ask whether this method produces an accurate predicted longevity for humans.

Exercises 7–8

Think about what the *actual* longevity of this type of animal might be.

7. Could it be 30 years? How about 5 years?

It is unlikely that this type of animal would have a longevity as much as 30 years or as little as 5 years because these values are further from the predicted value (17.4 years) than the errors given by residuals in the table.

8. Judging by the size of the residuals in our table, what kind of values do you think would be reasonable for the longevity of this type of animal?

A sensible range of possible values for the longevity is around 11–23 years.

Exercises 9–10 (5 minutes)

Let students work independently on Exercises 9 and 10. Then discuss and confirm as a class.

Exercises 9–10

Continue to think about the gestation times and longevity of animals. The gestation time for the type of animal called the ocelot is known to be 85 days.

The least squares line predicts the longevity of the ocelot to be

$$y = 6.642 + 0.03974(85) = 10.0 \text{ years.}$$

9. Based on the residuals in Example 3, would you be surprised to find that the longevity of the ocelot was 2 years? Why, or why not? What might be a sensible range of values for the actual longevity of the ocelot?

It is unlikely that the actual longevity of the ocelot is as little as 2 years; this would have a residual of 8 years, which is greater than the ones in the table. A sensible range of possible values might be from 4 years to 16 years.

10. We know that the actual longevity of the ocelot is 9 years. What is the residual for the ocelot?

Residual = 9 – 10 = –1 year

Closing (2 minutes)

Review the term *residual* with students. Have students point out specific residuals found in the exercises and interpret their values. Ask students the following questions:

- If a residual is close to 0, what does that mean about the predicted value?
- If a residual is negative, what does that mean about the predicted value?

Lesson Summary

- When a least squares line is used to calculate a predicted value, the prediction error can be measured by
 $\text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value}.$
- On the graph, the residuals are the vertical distances of the points from the least squares line.
- The residuals give us an idea how close a prediction might be when the least squares line is used to make a prediction for a value that is not included in the data set.

Exit Ticket (5 minutes)

Name _____

Date _____

Lesson 15: Interpreting Residuals from a Line

Exit Ticket

1. Meerkats have a gestation time of 70 days.
 - a. Use the equation of the least squares line from today's class, $y = 6.643 + 0.03974x$, to predict the longevity of the meerkat. Remember x equals the gestation time in days and y equals the longevity in years.
 - b. Approximately how close might your prediction be to the actual longevity of the meerkat? What was it (from class) that told you roughly how close a prediction might be to the true value?
 - c. According to your answers to parts (a) and (b), what is a reasonable range of possible values for the longevity of the meerkat?
 - d. The longevity of the meerkat is actually 10 years. Use this value and the predicted value that you calculated in part (a) to find the residual for the meerkat.

Exit Ticket Sample Solutions

1. Meerkats have a gestation time of 70 days.
 - a. Use the equation of the least squares line from today's class, $y = 6.643 + 0.03974x$, to predict the longevity of the meerkat. Remember x equals the gestation time in days and y equals the longevity in years.

$$y = 6.642 + 0.03974(70) = 9.4 \text{ years}$$
 - b. Approximately how close might your prediction be to the actual longevity of the meerkat? What was it (from class) that told you roughly how close a prediction might be to the true value?

Within about 6 years. We looked at the residuals for the original data set and saw that the largest residual was around 6 years.
 - c. According to your answers to parts (a) and (b), what is a reasonable range of possible values for the longevity of the meerkat?

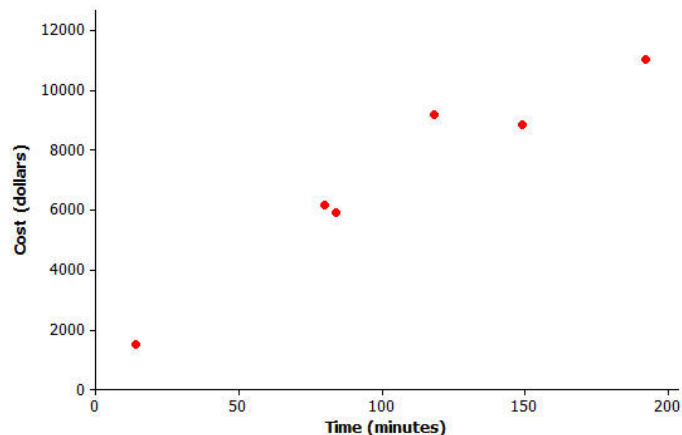
Between 3 and 15 years would be sensible.
 - d. The longevity of the meerkat is actually 10 years. Use this value and the predicted value that you calculated in part (a) to find the residual for the meerkat.

$$10 - 9.4 = 0.6 \text{ years}$$

Problem Set Sample Solutions

The time spent in surgery and the cost of surgery was recorded for six patients. The results and scatter plot are shown below.

Time (minutes)	Cost (\$)
14	1,510
80	6,178
84	5,912
118	9,184
149	8,855
192	11,023

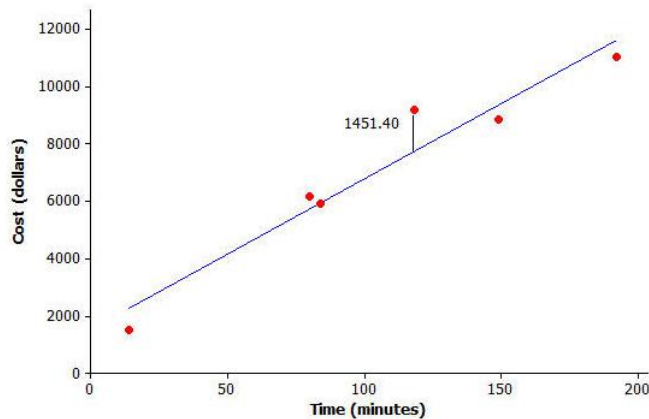


1. Calculate the equation of the least squares line relating cost to time. (Indicate slope to the nearest tenth and y -intercept to the nearest whole number.)

$$y = 1,514 + 52.7x, \text{ where } x = \text{time and } y = \text{cost}$$

2. Draw the least squares line on the graph above. (Hint: Substitute $x = 30$ into your equation to find the predicted y -value. Plot the point $(30, \text{your answer})$ on the graph. Then substitute $x = 180$ into the equation and plot the point. Join the two points with a straightedge.)

Plot the points $(30, 3095)$ and $(180, 11000)$ on the graph.



3. What does the least squares line predict for the cost of a surgery that lasts 118 minutes? (Calculate the cost to the nearest cent.)

$$y = 1,514 + 52.7(118) = \$7,732.60$$

4. How much do you have to add to your answer to question 3 to get the actual cost of surgery for a surgery lasting 118 minutes? (This is the residual.)

$$9,184 - 7,732.60 = \$1,451.40$$

5. Show your answer to question 4 as a vertical line between the point for that person in the scatter plot and the least squares line.

(Shown on the graph above.)

6. Remember that the residual is the actual y -value minus the predicted y -value. Calculate the residual for the surgery that took 149 minutes and cost \$8,855.

$$x = 149; y = 1,514 + 52.7(149) = \$9,366.30$$

$$\text{Residual} = 8,855 - 9,366.3 = -\$511.30$$

7. Calculate the other residuals, and write all the residuals in the table below.

Time (minutes)	Cost (\$)	Predicted Value (\$)	Residual
14	1,510	2,251.8	-741.8
80	6,178	5,730	448.0
84	5,912	5,940.8	-28.8
118	9,184	7,732.6	1,451.4
149	8,855	9,366.3	-511.3
192	11,023	11,632.4	-609.4

8. Suppose that a surgery took 100 minutes.

a. What does the least squares line predict for the cost of this surgery?

$$x = 100; y = 1,514 + 52.7(100) = \$6,784$$

b. Would you be surprised if the actual cost of this surgery were \$9,000? Why, or why not?

$\$9,000 - \$6,784 = \$2,216$. *This is much larger than any of the residuals in the table, so it would be surprising to find that the surgery was this expensive.*

c. Interpret the slope of the least squares line.

The slope is 52.7 dollars per minute. Interpret this slope in the following way: For each additional minute in surgery, the cost of the surgery increases by \$52.70.