Lesson 15: Interpreting Residuals from a Line

Classwork

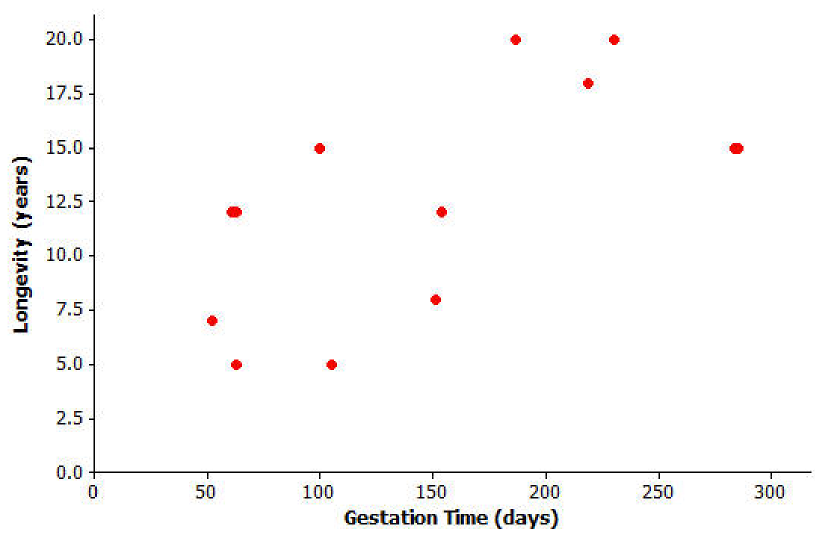
Example 1: Calculating Prediction Errors

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 longevities (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:

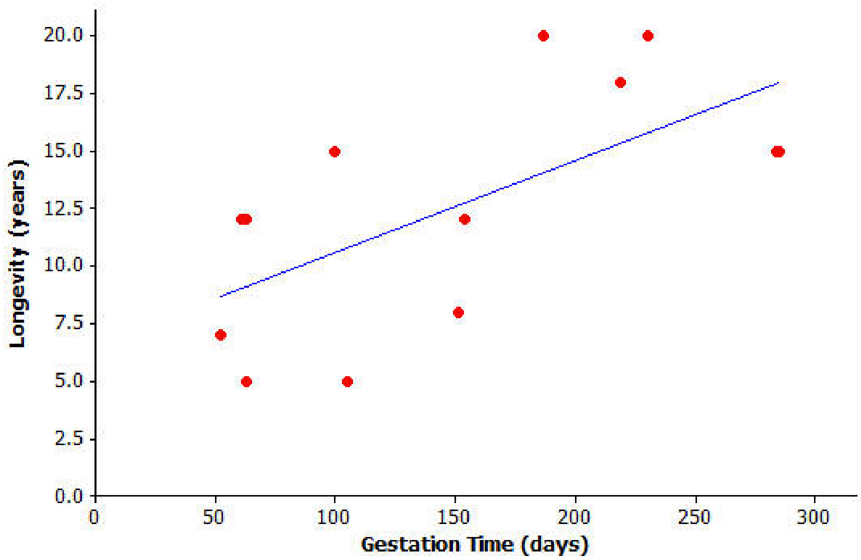


Exercises 1–4

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?

1. 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.



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?

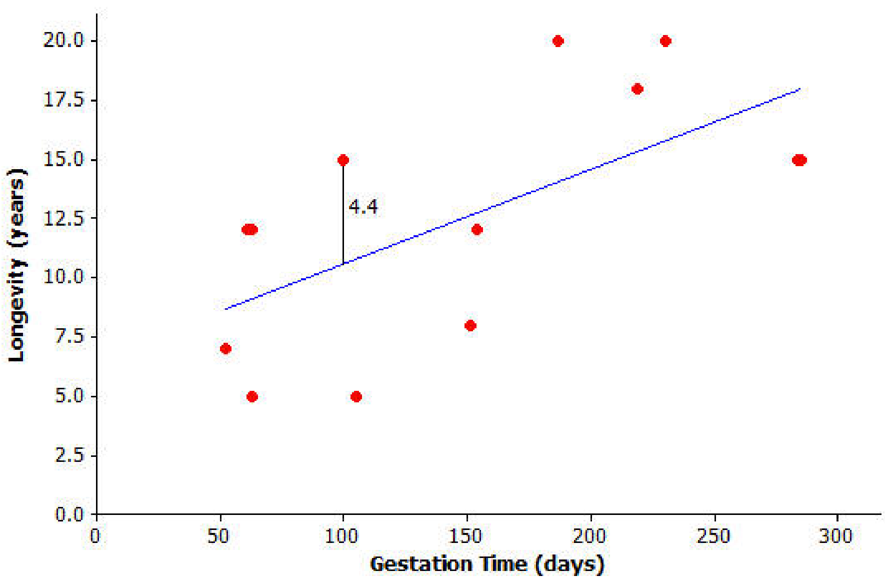
3. Would the value you predicted in Exercise 2 necessarily be the exact value for the longevity of that type of animal? Could the actual longevity of that type of animal be longer than predicted? Could it be shorter?

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?

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



Exercises 5–6

5. Let's continue to think about the gestation times and longevities 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?

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

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.

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

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

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

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.

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.)

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.

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

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

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

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?

Exercises 9–10

Continue to think about the gestation times and longevities 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?

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

Lesson Summary

- When a least squares line is used to calculate a predicted value, the prediction error can be measured by

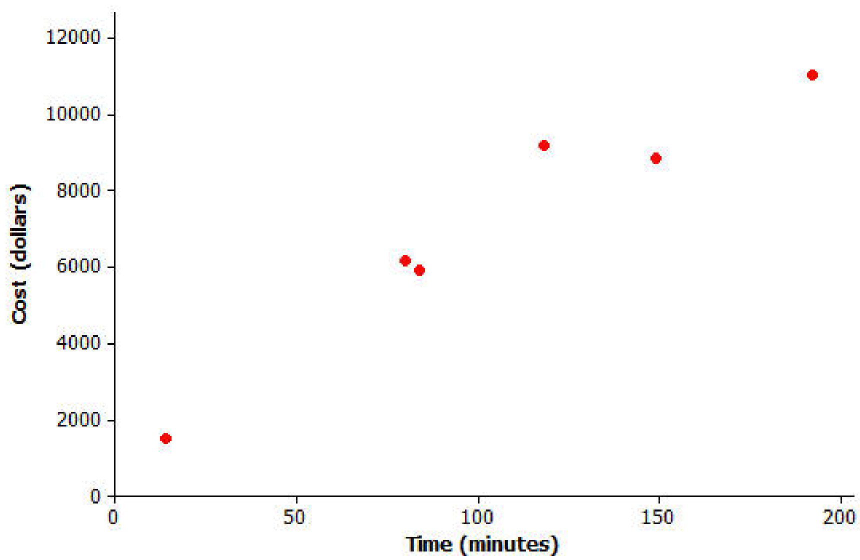
  $$\text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value}.$$

- 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.

Problem Set

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.)

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, your answer) on the graph. Then substitute $x = 180$ into the equation and plot the point. Join the two points with a straightedge.)

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.)

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.)

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.

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.

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

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

8. Suppose that a surgery took 100 minutes.
   a. What does the least squares line predict for the cost of this surgery?
   b. Would you be surprised if the actual cost of this surgery were \$9,000? Why, or why not?
   c. Interpret the slope of the least squares line.