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 types of animals are shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Animal** | **Gestation Time (days)** | **Longevity (years)** |
| Baboon |  |  |
| Black Bear |  |  |
| Beaver |  |  |
| Bison |  |  |
| Cat |  |  |
| Chimpanzee |  |  |
| Cow |  |  |
| Dog |  |  |
| Fox (Red) |  |  |
| Goat |  |  |
| Lion |  |  |
| Sheep |  |  |
| Wolf |  |  |

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 , where represents the gestation time (in days) and represents longevity (in years).

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



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

1. 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 days. You also know that its longevity is years, but what does the least squares line *predict* for the lion’s longevity?

Substituting days into the equation, you get or approximately . The least squares line predicts the lion’s longevity to be approximately years.

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

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

Exercises 5–6

1. 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.
	1. What is the gestation time for the black bear?
	2. Look at the graph. Roughly what does the least squares line predict for the longevity of the black bear?
	3. Use the gestation time from part (a) and the least squares line to predict the black bear’s longevity. Round your answer to the nearest tenth.
	4. What is the actual longevity of the black bear?
	5. How much do you have to add to the predicted value to get the actual longevity of the black bear?
	6. Show your answer to part (e) on the graph as a vertical line segment.
2. Repeat this activity for the sheep.
	1. Substitute the sheep’s gestation time for into the equation to find the predicted value for the sheep’s longevity. Round your answer to the nearest tenth.
	2. 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.)
	3. 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

actual value predicted value.

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

residual actual -value predicted -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 |  |  |  |
| Black Bear |  |  |  |
| Beaver |  |  |  |
| Bison |  |  |  |
| Cat |  |  |  |
| Chimpanzee |  |  |  |
| Cow |  |  |  |
| Dog |  |  |  |
| Fox (Red) |  |  |  |
| Goat |  |  |  |
| Lion |  |  |  |
| Sheep |  |  |  |
| Wolf |  |  |  |

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 days. Substituting into the equation of the least squares line you get

**Exercises 7–8**

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

1. Could it be years? How about years?

1. 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 days.

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

 years.

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

residual actual -value predicted -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 ()** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

 |  |

1. Calculate the equation of the least squares line relating cost to time. (Indicate slope to the nearest tenth and -intercept to the nearest whole number.)
2. Draw the least squares line on the graph above. (Hint: Substitute into your equation to find the predicted *-*value. Plot the point (, your answer) on the graph. Then substitute 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 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 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 -value minus the predicted -value. Calculate the residual for the surgery that took minutes and cost .
7. Calculate the other residuals, and write all the residuals in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Time (minutes) | Cost () | Predicted Value () | Residual |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. Suppose that a surgery took minutes.
	1. What does the least squares line predict for the cost of this surgery?
	2. Would you be surprised if the actual cost of this surgery were ? Why, or why not?
	3. Interpret the slope of the least squares line.