Residuals are a measure of how far the actual data points are from the line of best fit. Residuals are measured in the vertical (y) direction from the data point to the line. A single residual is calculated by:

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

A positive residual means that the actual value is greater than the predicted value; a negative residual means that the actual value is less than the predicted value.

A prediction made from a line of best fit gives no indication of the variability in the original data. Upper and lower boundary lines are parallel lines above and below the line of best fit. They give upper and lower limits (a “margin of error”) to predictions made by the line of best fit. For example, predicting a test score of 87 is useful. But a prediction of 87 ± 1 is very different from a prediction of 87 ± 10. The upper and lower boundary lines help us put limits like these on predictions.

The most commonly used techniques for finding upper and lower boundary lines are beyond the scope of this course. However, the bounds can be reasonably approximated by finding the residual with the greatest distance, then adding and subtracting that distance from the prediction.

For additional information, see the Math Notes box in Lesson 4.1.4. For additional examples and more practice on the topics from this chapter, see the Checkpoint 7 materials. Note that the problems below do not use a graphing calculator, while the problems in Checkpoint 7 assume a graphing calculator is available. Graphing calculators are introduced for statistical calculations in Lesson 4.1.4.

**Example**

It seems reasonable that there would be a relationship between the amount of time a student spends studying and their GPA. Suppose you were interested in predicting a student’s GPA based on the hours they study per week. You were able to randomly select 12 students and obtain this information from each student. An estimated line of best fit could be

\[ y = 2.03 + 0.17x. \]

(Note that this is the same data as in the example in Lesson 4.1.1 of this Parent Guide with Extra Practice.)

<table>
<thead>
<tr>
<th>Hours</th>
<th>4</th>
<th>5</th>
<th>11</th>
<th>1</th>
<th>15</th>
<th>2</th>
<th>10</th>
<th>6</th>
<th>7</th>
<th>0</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>2.9</td>
<td>3.3</td>
<td>3.9</td>
<td>2.2</td>
<td>4.1</td>
<td>1.8</td>
<td>4.6</td>
<td>2.9</td>
<td>2.2</td>
<td>3</td>
<td>3.3</td>
<td>4.5</td>
</tr>
</tbody>
</table>

a. Find the residual for the student who studied 10 hours, and interpret it in context. Would a student prefer a positive or negative residual?

*Example continues on next page →*
Example continued from previous page.

From the best-fit line, a student who studies 10 hours is predicted to earn a GPA of 
\[ y = 2.03 + 0.17(10) = 3.73. \]
This student actually received a 4.6 GPA. The residual is:
\[ \text{residual} = \text{actual y-value} - \text{predicted y-value} \]
\[ \text{residual} = 4.6 - 3.73 = 0.87 \]
The student who studied 10 hours earned a GPA that was 0.87 points higher than predicted. A student would prefer a positive residual because it means that he/she earned a higher GPA than was predicted from the number of hours spent studying.

b. Write the equations for the upper and lower boundary lines, and show the boundary lines on a scatterplot.

Looking at the scatterplot, the largest residual (the largest vertical distance, \( \Delta y \)) seems to be at the point \((7, 2.2)\). The predicted GPA for a student who studies 7 hours is 
\[ y = 2.03 + 0.17(7) = 3.22. \]
The residual for this point is:
\[ \text{residual} = \text{actual y-value} - \text{expected y-value} \]
\[ \text{residual} = 2.2 - 3.22 = -1.02 \]
The boundary lines are parallel to the best-fit line, at a distance of 1.02 above and below it. Parallel lines have the same slope as the best-fit line. The y-intercepts of the boundary lines will be 1.02 more or less than the y-intercept of the best-fit line.

Therefore, the upper boundary line is:
\[ y = (2.03 + 1.02) + 0.17x \] or \[ y = 3.05 + 0.17x \]
and the lower boundary line is:
\[ y = (2.03 - 1.02) + 0.17x \] or \[ y = 1.01 + 0.17x. \]

c. Predict the upper and lower bound of the GPA for a student who studies 8 hours per week. Is your prediction useful?

The lower bound is \[ y = 1.01 + 0.17(8) = 2.37, \] and the upper bound is 
\[ y = 3.03 + 0.17(8) = 4.41. \] We predict that a student who studies 8 hours per week has a GPA between 2.37 and 4.41. Due to the large amount of variability in the data collected, this is a large range. The prediction is not particularly useful.
Problems

1. It seems reasonable that the power of a car is related to its gas mileage. Suppose a random sample of 10 car models is selected and the engine horsepower and city gas mileage is recorded for each one. An outlier was removed and a line of best fit was estimated to be $y = 29.5 - 0.05x$. (Note that this is the same data as in Problem 1 in Lesson 4.1.1 of this Parent Guide with Extra Practice.)

| Power (hp) | 197 | 170 | 166 | 230 | 381 | 170 | 326 | 451 | 290 |
| Mileage (mpg) | 16 | 24 | 19 | 15 | 13 | 21 | 11 | 10 | 15 |

a. Find the residual for the vehicle that had 381 horsepower, and interpret it in context. Would a car owner prefer a positive or negative residual?
b. Remove the outlier from the data set. Write the equations for the upper and lower boundary lines, and show the boundary lines on a scatterplot.
c. Predict the upper and lower bounds of the gas mileage for a car with 300 horsepower. Is your prediction useful?

2. Many people believe that students who are strong in music are also strong in mathematics. The principal at University High School wonders if that same connection exists between music students and English students. The principal went through the records for the past year and found 10 students who were enrolled in both Advanced Placement Music and Advanced Placement English. He compared their final exam scores. A line of best fit was estimated to be $y = 132.7 - 0.67x$. (Note that this is the same data as in Problem 2 in Lesson 4.1.1 of this Parent Guide with Extra Practice.)

<table>
<thead>
<tr>
<th>Final Exam Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP Music</td>
</tr>
<tr>
<td>88</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>64</td>
</tr>
<tr>
<td>97</td>
</tr>
<tr>
<td>90</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>78</td>
</tr>
<tr>
<td>62</td>
</tr>
</tbody>
</table>

a. The principal checked the records of a student who just entered the school. She had a perfect score of 100 on the English final and her residual was 30 points. What was her predicted English score? What was her music score?
b. Write the equations for the upper and lower boundary lines, and show the boundary lines on a scatterplot. Do not add the new student from part (a) to your scatterplot.
c. Predict the upper and lower bounds of the AP English score for a student with a perfect score of 100 on the music final. Is your prediction useful?
1. a. From the best-fit line, a vehicle with 381 horsepower is predicted to have a mileage of 
   \[ y = 29.5 - 0.05(381) \approx 10.5 \]. But this vehicle actually got 13 mpg. The residual is:
   \[ \text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value} \]
   \[ \text{residual} = 13 - 10.5 = 2.5 \]
   The vehicle with 381 horsepower got 2.5 mpg more than was predicted. A car owner 
   would prefer a positive residual—a positive residual means the car is getting more 
   miles per gallon than was predicted from its power.

   b. Looking at the scatterplot without the outlier, the 
   largest residual (the largest vertical distance, \( \Delta y \)) 
   seems to be at the point (197, 16). The predicted 
   mileage for a car with 197 horsepower is 
   \[ y = 29.5 - 0.05(197) = 19.7 \] mpg. The residual 
   for this point is:
   \[ \text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value} \]
   \[ \text{residual} = 16 - 19.7 = -3.7 \]
   The boundary lines are parallel to the best-fit line, but a distance of 3.7 above it and 
   below it. Parallel lines have the same slope as the best-fit line. The \( y\)-intercepts of the 
   boundary lines will be 3.7 more or less than the \( y\)-intercept of the best-fit line.

   Therefore, the upper boundary line is: \( y = (29.5 + 3.7) - 0.05x \) or \( y = 33.2 - 0.05x \), 
   and the lower boundary line is: \( y = (29.5 - 3.7) - 0.05x \) or \( y = 25.8 - 0.05x \).

   c. The upper bound is 
   \[ y = 33.2 - 0.05(300) = 18.2 \], and the lower 
   bound is \( y = 25.8 - 0.05(300) = 10.8 \). We 
   predict that a car with 300 horsepower will 
   have gas mileage between 10.8 and 18.2 mpg. 
   The prediction could be useful: it indicates that 
   this is a car with a low miles per gallon ratio.
2. a. residual = actual y-value – predicted y-value
   
   \[ 30 = 100 - \text{predicted y-value} \]
   
   predicted y-value = 100 – 30 = 70
   
   From the best-fit line we can find the music score: \( 70 = 132.7 - 0.67x \) or \( x \approx 94 \).
   
   Her predicted English score was 70 and her music score was 94.

b. Looking at the scatterplot, the largest residual (the largest vertical distance, \( \Delta y \)) seems to be at the point (90, 90). The predicted English score for a student with 90 on the music test is
   
   \( y = 132.7 - 0.67(90) = 72.4 \). The residual for this point is: \( 90 - 72.4 = 17.6 \).
   
   Therefore, the upper boundary line is:
   
   \( y = (132.7 + 17.6) - 0.67x \) or \( y = 150.3 - 0.67x \),
   
   and the lower boundary line is:
   
   \( y = (132.7 - 17.6) - 0.67x \) or \( y = 115.1 - 0.67x \).

c. The upper bound is \( y = 150.3 - 0.67(100) = 83.3 \),
   
   and the lower bound is
   
   \( y = 115.1 - 0.67(100) = 48.1 \). We predict that a student who scores 100 on the AP Music final exam will earn a score between 48 and 83 on the AP English final exam. This range is much too large to be useful. There is a lot of variability in the data the principal collected, making the margin of error on the prediction very large.