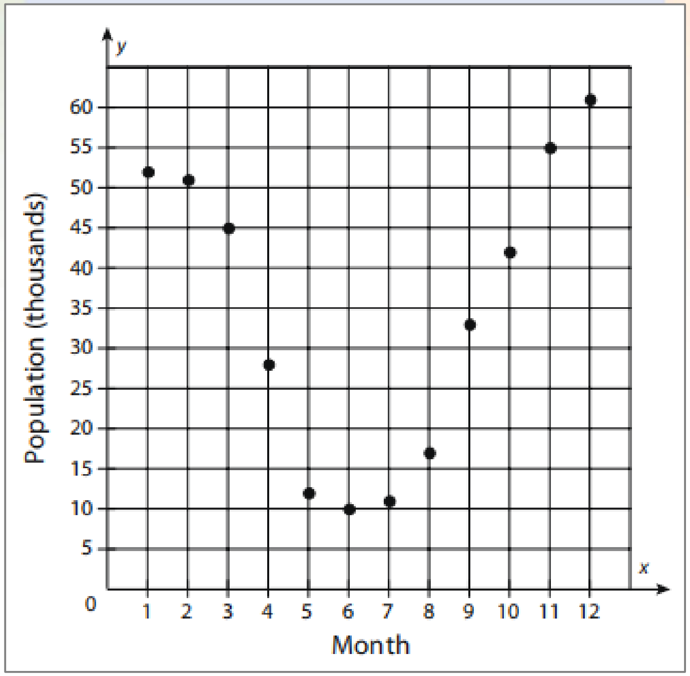
Sinusoidal Regression

Warm-Up

The population of wild rabbits in a national park varies over the year due to natural cycles.



1. Describe the shape of the data.
2. Would you expect similar data for next year? Why or why not?
3. How many rabbits might there be at the end of next year?

1. **Describe the shape of the data.**

The data decreases slowly then quickly in the beginning, then changes direction to increase quickly then slowly.

This forms a U-shape with flared ends, which resembles a curve on a sine function.

2. Would you expect similar data for next year? Why or why not?

Since the pattern is due to natural cycles, it is reasonable to expect it to repeat next year. This expectation is further supported by the way the data seems to form a sine curve.

3. How many rabbits might there be at the end of next year?

If the data pattern repeats next year, the population will be high at the end of it. There could be between 50 and 60 thousand rabbits in the nature preserve at the end of the year.

Instruction

Introduction

- Many things happen in cycles. For example, in temperature zones, the number of hours of daylight will increase steadily through the year until they reach a maximum in the summer, at which point the hours of daylight will begin to decrease.

- This pattern repeats the next year. Periodic phenomena such as this can be modeled using sine functions.

Key Concepts

- Recall that the process of finding a model to fit data is called **regression analysis**. A model fitted to the data is called the **regression model**, or just the **regression**.
- A sine function can be used to model the data if:
  - the data is curved
  - the data shows periods of increase and decrease
  - the data is periodic; that is, if the data pattern repeats, or could be expected to repeat

Key Concepts, *continued*

- A sine curve that has been fitted to a data set is called a **sinusoidal regression**. Sinusoidal is a shorthand description of the shape of the sine curve.
- In contrast with data with a quadratic trend, sinusoidal data tends to have a flared U-shape. While quadratic models are sometimes used to model a single period of data, their fit tends not to be as good on the flared portions of the curve.
- This tendency can be reflected in the residual plots, yielding a U-shaped graph even though visually the graph appear to be a reasonable fit.

Key Concepts, *continued*

- You can use a calculator or other graphing software to find a logarithmic regression model for a data set.

- Note that TI calculators assume the calculator is in radian mode for sinusoidal regression. Be sure to check that your calculator is set to **radian mode** before graphing calculator-generated sinusoidal regressions.

Key Concepts, *continued*

- Note that most graphing software returns sinusoidal regression models in a form equivalent to $y = a \sin(bx + c) + d$. (Some software places the constant d in front of the sine expression.).

On a TI-83/84:

Step 1: Press [STAT] to bring up the statistics menu. The first option, 1: Edit, will already be highlighted. Press [ENTER].

Step 2: Arrow up to L1 and press [CLEAR], then [ENTER], to clear the list. Repeat this process to clear L2 and L3, if needed.

Step 3: Enter the ordered pairs in the L1 and L2 lists. Make sure to enter the *x*-coordinates in L1 and the *y*-coordinates in L2.

Step 4: Press [STAT]. Arrow to the CALC menu. Press C: SinReg.

Step 5: Press [(], then [2ND][1] to type "L1" for Xlist. Press [,], then [2ND][2] to type "L2" for Ylist. Press [)] to close the parentheses.

Step 6: Press [ENTER] to calculate. The parameter values will appear on the screen.

Key Concepts, *continued*

On a TI-Nspire:

Step 1: Press [home]. Arrow over to the spreadsheet icon, the fourth icon from the left, and press [enter].

Step 2: To clear the lists in your calculator, arrow up to the topmost cell of the table to highlight the entire column, then press [menu]. Choose 3: Data, then 4: Clear Data. Repeat for each column as necessary.

Step 3: Arrow up to the topmost cell of the first column, labeled "A." Press [X] [enter] to type *x*. Then arrow over to the second column, labeled "B." Press [Y][enter] to type *y*.

Step 4: Arrow down to cell A1 and enter the first *x*-value from the ordered pairs. Press [enter]. Enter the second *x*-value in cell A2 and so on.

Step 5: Move over to cell B1 and enter the first *y*-value. Press [enter]. Enter the second *y*-value in cell B2 and so on.

Step 6: To fit an equation to the data points, press [menu] and select 4: Statistics, then 1: Stat Calculations, then C for sinusoidal regression. Select "*x*" from the X List pop-up menu. Press [tab] to move to the Y List pop-up menu, then select "*y*" from the options. Tab to "OK" and press [enter]. The parameter values will appear on the screen.

Key Concepts, *continued*

- To compare a data set and a function, plot the function on the same coordinate plane as the scatter plot of the data set.
- Graph a sine function by plotting at least five points and drawing a curve through these points.
- Recall that a function is a good fit for the data if it passes closely to the data points and if some of the data points are above the curve and some are below the curve.
- Recall that a function that is a good fit for the data can be used to make estimates for data not included in the plotted data set.

Key Concepts, *continued*

- Evaluate a function algebraically for a given value of *x* or *y* by substituting the given value for *x* or *y* and solving for the remaining variable.
- Evaluate a function graphically for a given value of *x* or *y* by finding the point on the graph of the function with the known coordinate, then finding the corresponding *x*- or *y*-value of the point.
- You can also analyze the **residuals** to assess how well a logarithmic model fits the data. Recall that a residual is the difference between the function-predicted *y*-coordinate and the *y*-coordinate of the actual data point.

Key Concepts, *continued*

- Recall that a **residual plot** is used to visually assess the fit. To create a residual plot, subtract each predicted *y*-value from the actual *y*-value for each point in the data set. Then plot the differences against their corresponding *x*-values.
- A residual plot with a random shape indicates the function is a good fit for the data.
- A residual plot with a U-shape or other distinct pattern indicates that the function is not a good fit for the data.
- If two separate models both have residual that indicate a good fit, the model whose residuals are closer to 0 is a better fit.

Common Errors/Misconceptions

- using a sine function to model a data set that does not exhibit sinusoidal characteristics
- confusing *x* and *y* when graphing data points or analyzing a graph
- confusing parameters in technology-generated regressions
- using radian inputs for a model generated in degree mode or vice-versa

Guided Practice

Example 1

The average monthly low temperatures for North Carolina are recorded in the table below.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Average low temperature (°F)	29.6	31.9	38.9	46.4	55.3	63.8	68.5	67.2	61	48.2	39.5	32.6

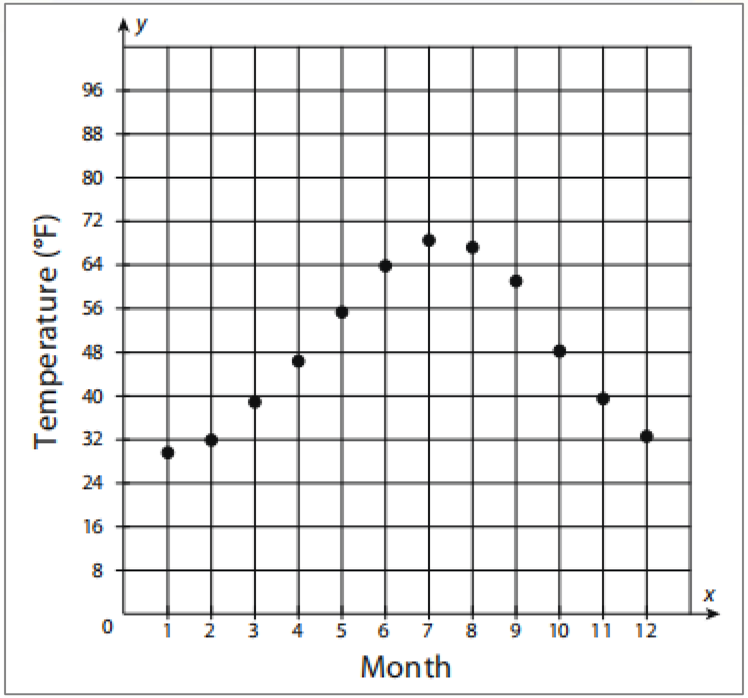
A climatologist determined that the data could be modeled by the sine function $y = 19.11\sin(0.54x - 2.28) + 49.09$.

Determine whether a logarithmic function is a good fit for the data.

Guided Practice: Example 1, *continued*

1. Create a scatter plot of the data set.

Let the *x*-axis represent the month, and the *y*-axis represent the average low temperature for that month.



Guided Practice: Example 1, *continued*

2. Determine if the data can be represented by a sine function.

The data shows an upside-down flared U-shape. Additionally, this shape would likely repeat in any given year due to the seasons.

This indicates a periodic tendency.

A sine function is a good model for this data.

Guided Practice: Example 1, *continued*

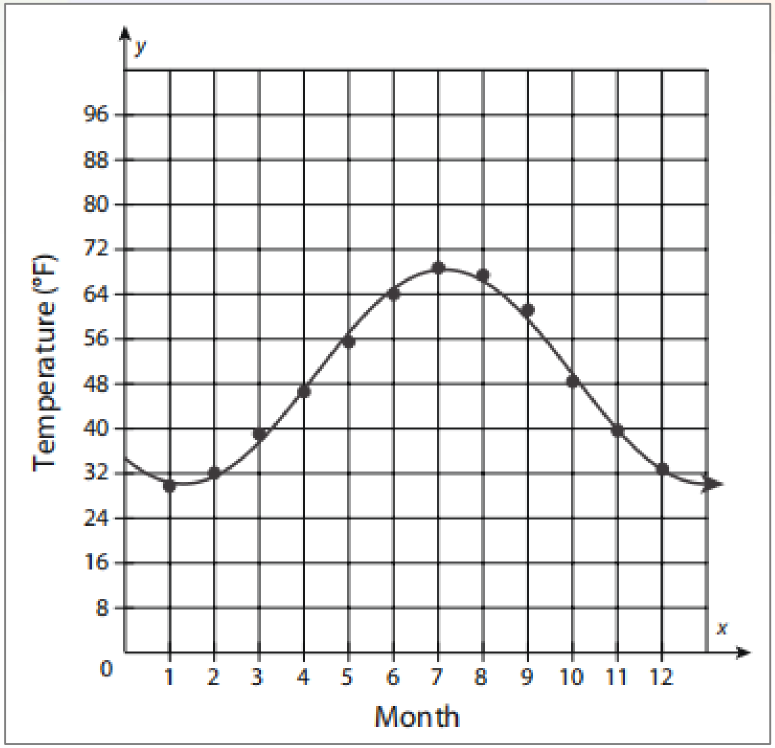
3. Plot the function $y = 19.11 \sin(0.54x - 2.28) + 49.09$ on the scatter plot.

Evaluate the function for at least five values of ***x***. Plot the ***x-y*** coordinates on the scatter plot. Then draw a curve to connect them. Note that x-values are given in radians in the following table.

x	1	2	3	4	5	6	7	8	9	10	11	12
y	30.25	31.28	37.37	46.8	56.88	64.74	68.15	66.13	59.27	49.5	39.62	32.43

Guided Practice: Example 1, *continued*

3. Plot the function $y = 19.11 \sin(0.54x - 2.28) + 49.09$ on the scatter plot.



Guided Practice: Example 1, *continued*

4. **Determine if the function is a good fit for the data.**

The function passes close to the data points, and some points lie above the function, while others lie below the function.

The function appears to be a good fit for the data.

Guided Practice

Example 2

The average closing price per share of a stock over the past 12 months is shown in the following table.

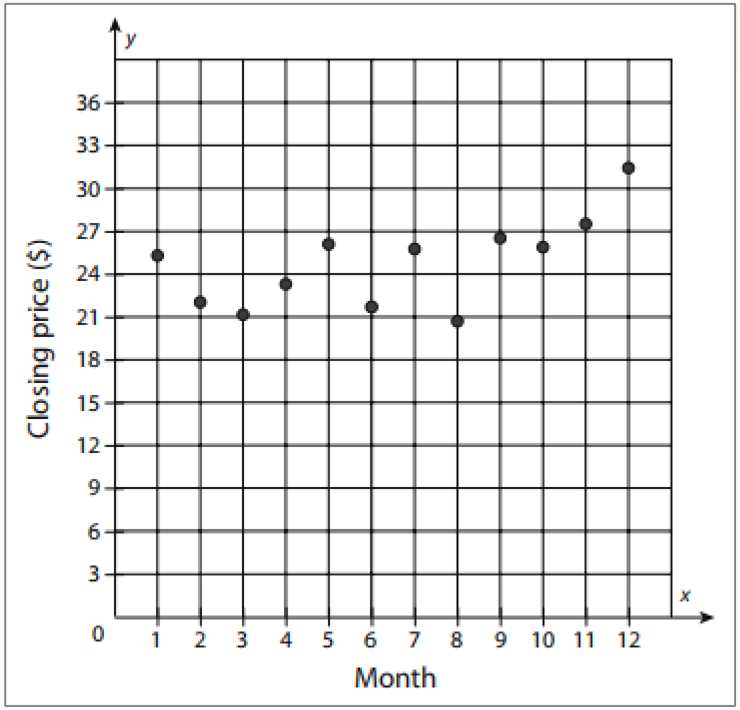
Month	Closing price ($)	Month	Closing price ($)
1	25.30	7	25.76
2	22.03	8	20.69
3	21.16	9	26.53
4	23.30	10	25.89
5	26.10	11	27.52
6	21.70	12	31.41

Determine whether a sine function is a good fit for the data. If so, find a sinusoidal regression for the data and evaluate the fit with a graph.

Guided Practice: Example 2, *continued*

1. Create a scatter plot of the data set.

Let the *x*-axis represent the month, and let the *y*-axis represent the average closing price for that month.



Guided Practice: Example 2, *continued*

2. Determine if the data can be represented by a sine function.

The data seems to have an increasing trend.

The shape is a curve, but it does not appear to have flared ends.

A sine function is not a good model for this data.

Guided Practice

Example 3

For a science project on tides, Marissa selects a large boulder in shallow water off the coast and measures the depth of the water above the boulder every hour. Her results are recorded in the following table.

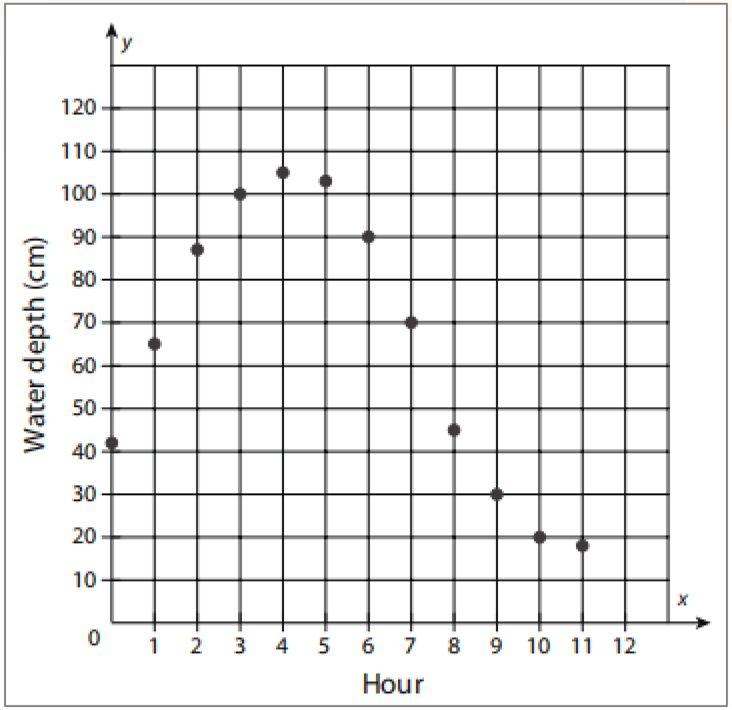
Hour	Water depth (cm)	Hour	Water depth (cm)
0	42	6	90
1	65	7	70
2	87	8	45
3	100	9	30
4	105	10	20
5	103	11	18

Determine whether a sine function is a good fit for the data. If so, find a sinusoidal regression for the data and evaluate the fit with a graph. Then use the model to predict the water depth 15 hours after Marissa began her measurements.

Guided Practice: Example 3, *continued*

1. Create a scatter plot of the data set.

Let the *x*-axis represent the hour and the *y*-axis represent the depth of the water at that time.



Guided Practice: Example 3, *continued*

2. Determine if the data can be represented by a sine function.

The data seems to show an upside-down U-shape with a flared base. Additionally, tides rise and fall every day, which indicates periodic tendencies.

A sine function is a good model for this data.

Guided Practice: Example 3, *continued*

3. Find the equation of the regression model that best fits the data.

Use a calculator or other technology to find the regression equation.

On a TI-83/84:

Step 1: Press [STAT] to bring up the statistics menu. The first option, 1: Edit, will already be highlighted. Press [ENTER].

Step 2: Arrow up to L1 and press [CLEAR], then [ENTER], to clear the list. Repeat this process to clear L2 and L3, if needed.

Step 3: Enter the ordered pairs in the L1 and L2 lists. Make sure to enter the *x*-coordinates in L1 and the *y*-coordinates in L2.

Step 4: Press [STAT]. Arrow to the CALC menu. Press C: SinReg.

Step 5: Press [(], then [2ND][1] to type "L1" for Xlist. Press [,], then [2ND][2] to type "L2" for Ylist. Press [)] to close the parentheses.

Step 6: Press [ENTER] to calculate. The parameter values will appear on the screen.

Guided Practice: Example 3, *continued*

On a TI-Nspire:

Step 1: Press [home]. Arrow over to the spreadsheet icon, the fourth icon from the left, and press [enter].

Step 2: To clear the lists in your calculator, arrow up to the topmost cell of the table to highlight the entire column, then press [menu]. Choose 3: Data, then 4: Clear Data. Repeat for each column as necessary.

Step 3: Arrow up to the topmost cell of the first column, labeled "A." Press [X][enter] to type *x*. Then arrow over to the second column, labeled "B." Press [Y][enter] to type *y*.

Step 4: Arrow down to cell A1 and enter the first *x*-value from the ordered pairs. Press [enter]. Enter the second *x*-value in cell A2 and so on.

Step 5: Move over to cell B1 and enter the first *y*-value. Press [enter]. Enter the second *y*-value in cell B2 and so on.

Step 6: To fit an equation to the data points, press [menu] and select 4: Statistics, then 1: Stat Calculations, then C for sinusoidal regression. Select "*x*" from the X List pop-up menu. Press [tab] to move to the Y List pop-up menu, then select "*y*" from the options. Tab to "OK" and press [enter]. The parameter values will appear on the screen for a logarithmic regression in base *e*.

Guided Practice: Example 3, *continued*

3. Find the equation of the regression model that best fits the data.

The equation of the sine function is:

$$y = 44.76\sin(0.49x - 0.43) + 62.09$$

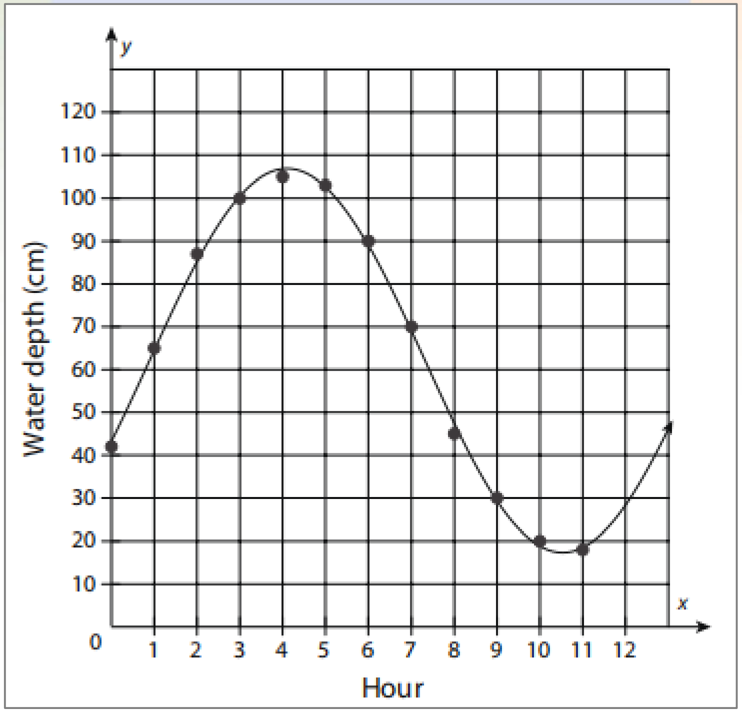
Guided Practice: Example 3, *continued*

4. Plot the equation on the scatter plot and evaluate the fit.

Evaluate the function for at least five values of ***x***. Plot the ***x*-*y*** coordinates on the scatter plot. Then draw a curve to connect them. Note that *x*-coordinates are in radians.

x	0	2	4	6	8	10
y	43.43	85.49	106.81	88.52	46.81	18.64

Guided Practice: Example 3, *continued*



The function appears to be a good fit for the data.

Guided Practice: Example 3, *continued*

5. Estimate the water depth 15 hours after Marissa began her measurements.

Evaluate the equation $y = 44.76 \sin(0.49x - 0.43) + 62.09$ for $x = 15$ to estimate the depth of the water after 15 hours.

$y = 44.76 \sin(0.49(15) - 0.43) + 62.09$	Substitute 15 for x.
$y = 44.76 \sin(6.92) + 62.09$	Simplify the expression inside the parenthesis.
$y \approx 44.76 \bullet 0.59 + 62.09$	Evaluate the sine function in radians.
$y \approx 26.62 + 62.09$	Multiply.
$y \approx 88.71$	Add.

According to the model, after 15 hours, the water depth will be approximately 89 cm.

Guided Practice

Example 4

The following table shows the percentage of the moon that is visible each day in March 2020.

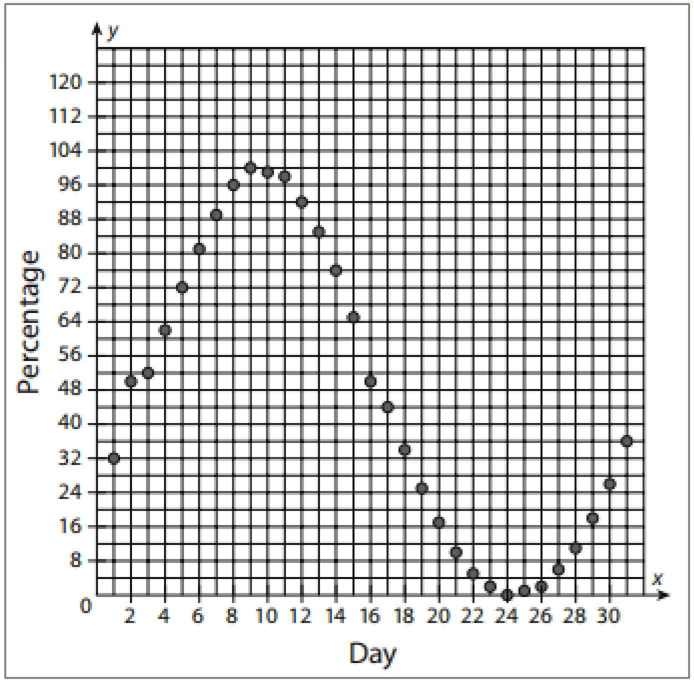
Day	Percentage	Day	Percentage	Day	Percentage
1	32	12	92	22	5
2	50	13	85	23	2
3	52	14	76	24	0
4	62	15	65	25	1
5	72	16	50	26	2
6	81	17	44	27	6
7	89	18	34	28	11
8	96	19	25	29	18
9	100	20	17	30	26
10	99	21	10	31	36
11	98				

Kayla came up with the function $y = 50 + 50\sin(0.21x - 0.47)$. Plot the data and the equation on a scatter plot to evaluate the fit. Then use a residual plot to evaluate the fit. Is the function a good fit for the data?

Guided Practice: Example 4, *continued*

1. Create a scatter plot of the data set.

Let the *x*-axis represent the day and the *y*-axis represent the percentage of the moon that is visible.



Guided Practice: Example 4, *continued*

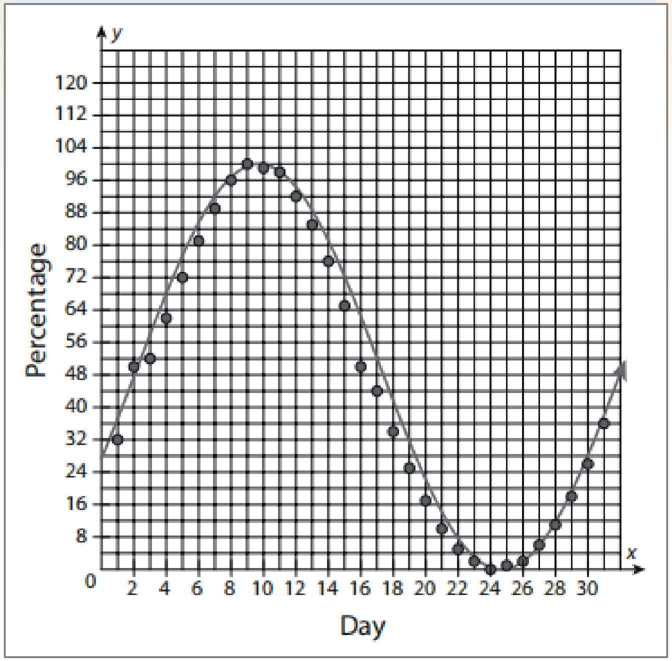
2. Plot the equation on the scatter plot and evaluate the fit.

Evaluate the function $y = 50 + 50 \sin(0.21x - 0.47)$ for each value of x. Plot the x-y coordinates on the scatter plot. Then draw a curve to connect them.

Day	Actual percentage	Predicted percentage	Day	Actual percentage	Predicted percentage
1	32	37.1	17	44	52.1
2	50	47.5	18	34	41.6
3	52	58	19	25	31.5
4	62	68.1	20	17	22.2
5	72	77.4	21	10	14.2
6	81	85.5	22	5	7.7
7	89	92.1	23	2	3.1
8	96	96.8	24	0	0.5
9	100	99.4	25	1	0.1
10	99	99.9	26	2	1.9
11	98	98.2	27	6	5.8
12	92	94.4	28	11	11.7
13	85	88.6	29	18	19.2
14	76	81.1	30	26	28.1
15	65	72.3	31	36	38
16	50	62.4			

Guided Practice: Example 4, *continued*

2. Plot the equation on the scatter plot and evaluate the fit.



The function follows the curve of the data and passes reasonably close to each value. The function appears to be a good fit for the data, although it does appear that a large number of data points lie below the curve.

Guided Practice: Example 4, *continued*

3. Find the residuals.

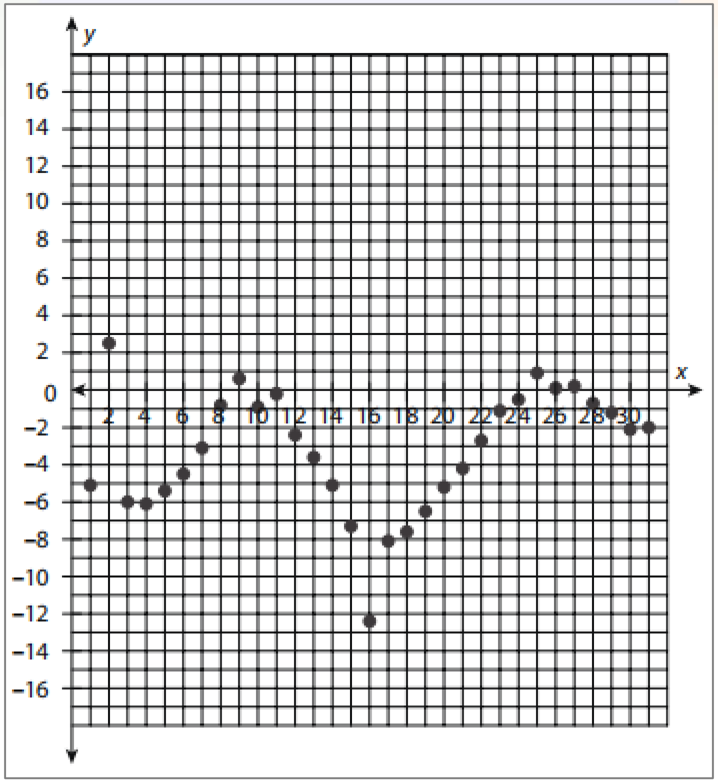
Subtract the estimated *y*-value from the actual *y*-value for each data point.

Day	Actual percentage	Predicted percentage	Residual
1	32	37.1	–5.1
2	50	47.5	2.5
3	52	58	–6
4	62	68.1	–6.1
5	72	77.4	–5.4
6	81	85.5	–4.5
7	89	92.1	–3.1
8	96	96.8	–0.8
9	100	99.4	0.6
10	99	99.9	–0.9
11	98	98.2	–0.2
12	92	94.4	–2.4
13	85	88.6	–3.6
14	76	81.1	–5.1
15	65	72.3	–7.3
16	50	62.4	–12.4
17	44	52.1	–8.1
18	34	41.6	–7.6
19	25	31.5	–6.5
20	17	22.2	–5.2
21	10	14.2	–4.2
22	5	7.7	–2.7
23	2	3.1	–1.1
24	0	0.5	–0.5
25	1	0.1	0.9
26	2	1.9	0.1
27	6	5.8	0.2
28	11	11.7	–0.7
29	18	19.2	–1.2
30	26	28.1	–2.1
31	36	38	–2

Guided Practice: Example 4, *continued*

4. Create a residual plot.

Plot each residual against its corresponding ***x***-value.



Guided Practice: Example 4, *continued*

5. Determine if the function is a good estimate for the data.

Most of the points on the residual plot lie below the *x*-axis. This indicates that the given function underestimates the data in the majority of cases.

Looking at the scatter plot, this seems to correspond to the long, straighter portions of the curve, where the moon is waxing or waning. Additionally, there seems to be a distinct sine-wave pattern on the graph.

This may indicate that a simple sine regression may not be the best fit for the data.

✔