



Lesson 2: Analyzing a Data Set

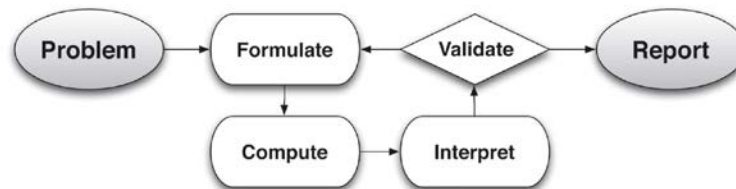
Student Outcomes

- Students recognize linear, quadratic, and exponential functions when presented as a data set or sequence and formulate a model based on the data.

Lesson Notes

This lesson asks students to use the first steps in the modeling cycle when a mathematical relationship is represented as a table of values. They will look for trends and patterns in the data and formulate a model. First, they will look for common differences between the function values of regular intervals, i.e., the first differences for linear and the second differences for quadratic. For exponential functions, they will look for a common ratio and/or a constant percent change. For this lesson, all data will be based on functions' inputs and outputs, and the numbers will be friendly.

Throughout this module, refer to the modeling cycle below (found on page 61 of the CCLS and page 72 of the CCSS):



MP.1
&
MP.4

The Opening Exercise and Discussion for this lesson involve students learning to use the first steps in the modeling cycle. When presented with a problem, they will make sense of the given information, analyze the presentation, define the variables involved, look for entry points to a solution, and create an equation or graph to model the context. Students represent the data with appropriate symbolic representation, interpret their mathematical results in the context, and validate their conclusions.

MP.8

All of the examples, exercises, and discussion in the lesson require that students search for regularity or trends in tables of values and recognize the correspondence between equations, verbal descriptions, and tables.

Classwork

Opening Exercise (8 minutes)

Display the following three sets of data on the board or screen. Ask students to study each data set and make notes about what they see. Then, have a brainstorming session where students share whatever they observe. Write down all suggestions on the board or screen so that you can come back to them later to see how many were helpful. Use the questions that follow to guide the discussion.

Opening Exercise

When tables are used to model functions, we typically have just a few sample values of the function and therefore have to do some detective work to figure out what the function might be. Look at these three tables:

x	$f(x)$
0	6
1	12
2	18
3	24
4	30
5	36

x	$g(x)$
0	0
1	14
2	24
3	30
4	32
5	30

x	$h(x)$
0	1
1	3
2	9
3	27
4	81
5	243

- What do you notice about the three sets of data? Can you identify the type of function they represent?

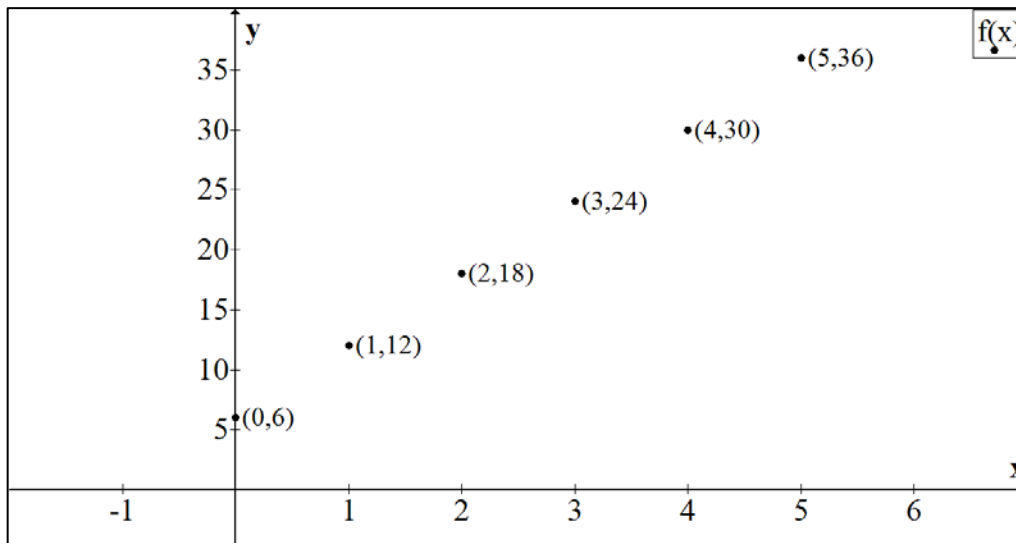
Students may observe some of the examples listed below. It is also possible that they will not notice much of anything, or they may have ideas that are incorrect. Just let them brainstorm, and then come back to their ideas later to see how many were helpful.

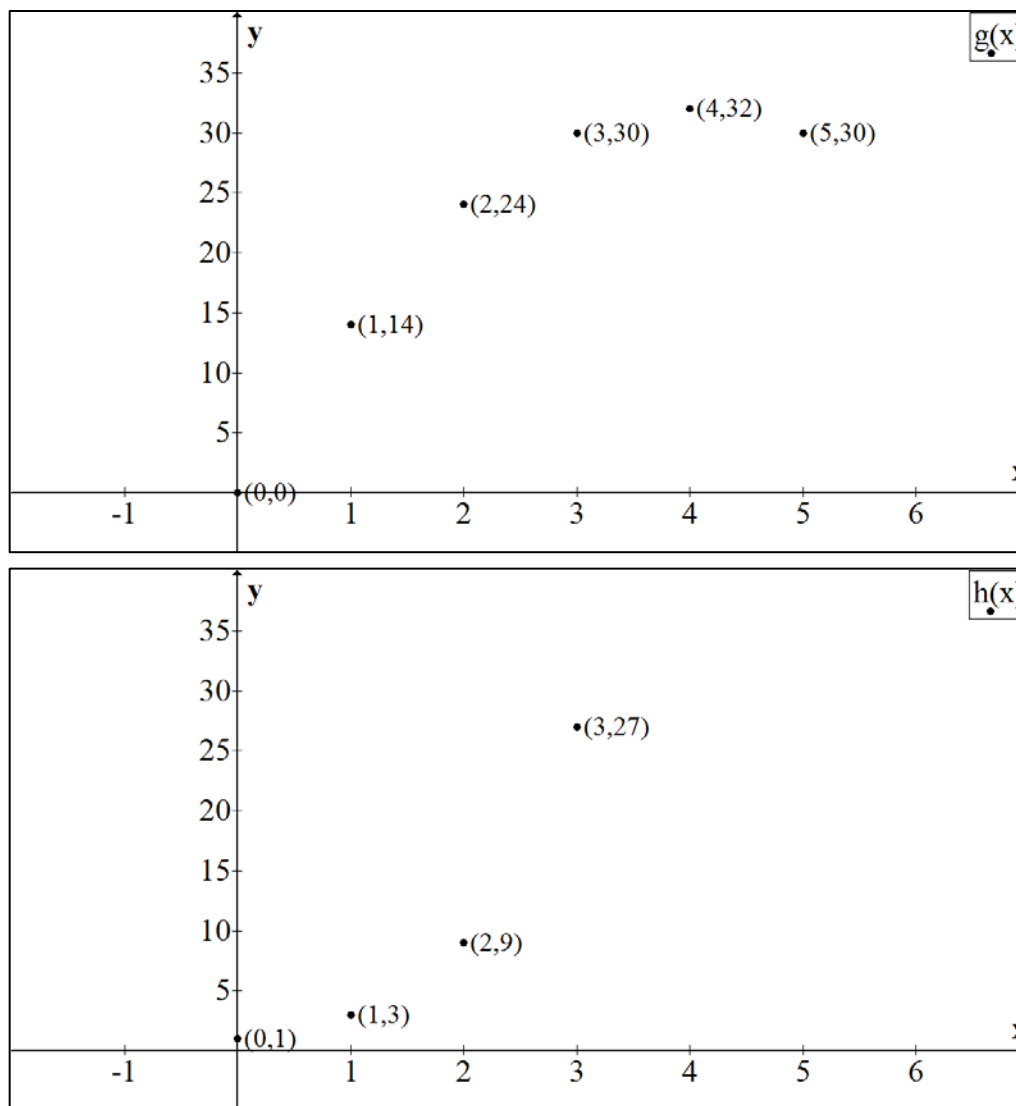
- The orange set is growing much faster than the first two. Many of the function values in the blue and green are the same. It is difficult to know the function by looking at only these numbers for the first two. However, I notice that the function values in the green set were growing steadily and then decreased at the end. It might help if we had more data points.*

MP.3

Have students make conjectures about the type of function they believe each data set represents and support their conjectures with evidence, which might include a graph.

Now show a data plot for each:





- Now can you see the trends more clearly?
 - Yes, it is obvious that the first is linear. It is clearer that the second is quadratic. And the third looks to be exponential.
- We learned to determine the type of function by looking at the shapes of graphs. Can we also determine the type of function by just using its table of values? Analyze each table and find any special patterns that may help determine the type of function it represents.
 - The table of values for the linear function has a constant first difference of 6, indicating the graph will have a slope of 6. The quadratic function does not have a constant difference like a linear function. However, the difference of differences (second difference) is constant. The exponential function does not have constant first or second differences. However, the y -value is multiplied by a constant value of 3. It should be noted that the x -values are all increasing by one. Not all data are presented where the x -values are increasing by a constant value, so it is necessary to interpolate values to produce equal-sized intervals.

Discussion (8 minutes)

Have students take notes in their math journals or notebooks about the discoveries made in the discussion below. With each guiding question, allow enough time for students to make a journal entry. Encourage them to keep their notes organized because they may be of later use as a reference tool.

- How can we identify a linear function from its data?
 - *First, make sure the x -values have a constant difference. Then, check to see if the first differences (changes) between successive y -values are constant (i.e., the successive y -values are found by adding a constant value).*
- How can we identify a quadratic function from its data?
 - *First, make sure the x -values have a constant difference. Then, if the first differences between successive y -values are not constant, look to see if the second differences are (i.e., the changes in the changes are constant or the differences between the first differences are constant and not zero).*
- How can we identify an exponential function from its data?
 - *First, make sure the x -values have a constant difference. Then, look to see if there is a constant ratio between successive y -values (i.e., that the y -values are found by multiplying by a constant).*
- How can we identify an absolute value function from a data set?
 - *We would first notice the same phenomenon that happens with a quadratic—the values go up and then down, or vice-versa if the vertex is included in, or straddled by, the data set. However, when we check for constant first differences, we find them to be constant. That tells us that the pieces of the graph on each side of the minimum and maximum point are parts of lines and that the data set can be modeled by an absolute value.*
- How about square root and cube root? How might we recognize them from a set of data?
 - *For square root, we check to see if squaring the y -value results in the corresponding x -value. For cube root parent functions, we see if cubing the y -value results in the corresponding x -value. If the graphs have been transformed, then squaring or cubing the output will not directly yield the input value. One needs to take care of the transformation to identify whether it is a square root or cube root function.*

Scaffolding:

- Having students keep a math journal or notebook helps to solidify their procedural knowledge base and provides more opportunities to deepen their conceptual understanding. Allowing students to use their notebooks as a reference encourages them to be thorough in their note-taking and improves their attention to and involvement in discussions.
- Occasionally ask a couple of students to share what they have written in their notebooks to give others ideas on how to take good notes.

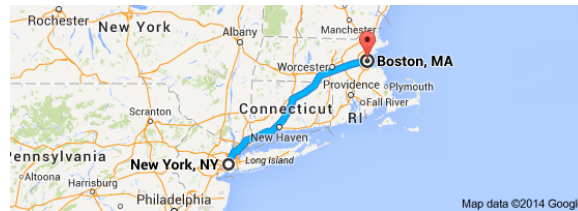
Make sure students have time to make entries in their math journals before continuing with the following example.

Example 1 (8 minutes)

Have the students read the prompt and analyze the data sets for this problem, then have them work with a partner or small group to answer the questions that follow.

Example 1

Noam and Athena had an argument about whether it would take longer to get from NYC to Boston and back by car or by train. To settle their differences, they made separate, non-stop round trips from NYC to Boston. On the trip, at the end of each hour, both recorded the number of miles they had traveled from their starting points in NYC. The tables below show their travel times, in hours, and the distances from their starting points, in miles. The first table shows Noam's travel time and distance from the starting point, and the second represents Athena's. Use both data sets to justify your answers to the questions below.



Time in Hours	Noam's Distance in miles
0	0
1	55
2	110
3	165
4	220
5	165
6	110
7	55
8	0

Time in Hours	Athena's Distance in miles
0	0
1	81
2	144
3	189
4	216
5	225
6	216
7	189
8	144
9	81
10	0

Scaffolding:

You might want to demonstrate the repeated calculations for the first table (MP.8):

$$0(55) = 0$$

$$1(55) = 55$$

$$2(55) = 110$$

$$3(55) = 165, \text{ etc.}$$

- a. Who do you think is driving, and who is riding the train? Explain your answer in the context of the problem.

This is an open-ended question with no right or wrong answer. Check for mathematically sound reasoning based on the data.

Sample Response: It appears that Athena is riding the train since she was able to go 81 miles in the first hour, and Noam was able to go only 55, which is a typical highway speed. Also, Athena made the round trip in 10 hours, while Noam made it in 8, so it looks like the train may have had to stop at stations when it was near Boston, slowing its progress considerably during the 2 hours on either side of Boston.

- b. According to the data, how far apart are Boston and New York City? Explain mathematically.

This is an open-ended question with no right or wrong answer. Check for mathematically sound reasoning based on the data.

Sample Response: Based on the symmetry of the values in the table, Noam's maximum distance was 220 miles (at 4 hours). Using the same symmetry, Athena's maximum value was 225. They may have started or ended at different places, but it is more likely that the route for the train was slightly different from that of the car.

- c. How long did it take each of them to make the round trip?

Let's call Noam's distance $N(t)$ and Athena's distance $A(t)$. Noam was traveling for 8 hours and Athena for 10 hours. The zeros for $N(t)$ are $(0, 0)$ and $(8, 0)$ and for $A(t)$ are $(0, 0)$ and $(10, 0)$. It took Noam 8 hours; it took Athena 10 hours.

- d. According to their collected data, which method of travel was faster?

If we assume that Noam was traveling by car, then the car was faster by 2 hours, overall. However, the speed of the train was faster for the first and the last hour, and then slowed.

- e. What was the average rate of change for Athena for the interval from 3 to 4 hours? How might you explain that in the context of the problem?

She only traveled 27 miles per hour: $\frac{216 - 189}{4 - 3}$. Since she was likely on the train, there may have been stops during that time period.

- f. Noam believes a quadratic function can be used as a model for both data sets. Do you agree? Use and describe the key features of the functions represented by the data sets to support your answer.

The two data sets have several things in common. Both are symmetric, both have a vertex, and both have $(0, 0)$ as one of the x -intercepts and the y -intercept. The other x -intercept is $(8, 0)$ for Noam and $(10, 0)$ for Athena. However, Noam's data set has a positive constant difference of $+55$ on one side of the vertex, which then changes to -55 on the other side. Noam's data set can be best modeled by an absolute value or other piecewise function. Athena's data does not have a constant first difference, so it is not linear. When we check the second differences, we find it to be constant (-18) . Therefore, Athena's trip can be modeled with a quadratic function.

Scaffolding:

As an extension for this task, you might ask students to create the function equation for each of the data sets.

Solutions:

$$N(t) = -55|x - 4| + 220$$

or

$$N(t) = \begin{cases} 55t, & 0 \leq t \leq 4 \\ -55t + 440, & 4 < t \leq 8 \end{cases}$$

$$A(t) = -9(x - 5)^2 + 225$$

Exercises (13 minutes)

Exercises

1. Explain why each function can or cannot be used to model the given data set.

a. $f(x) = 3x + 5$

This function cannot be used to model the data set. The y -intercept is 5, but the first difference is not constant, and the data set is not a linear function.

b. $f(x) = -(x - 2)^2 + 9$

This function can be used to model the data set. The second difference has a constant value of -2 ; therefore, it is a quadratic function. The vertex is $(2, 9)$ and it is a maximum since the leading coefficient is negative.

c. $f(x) = -x^2 + 4x - 5$

This function cannot be used to model the data set. The y -intercept for this equation is -5 instead of 5.

d. $f(x) = 3^x + 4$

This function cannot be used to model the data set. The y -intercept is five, but $f(x)$ values are not being multiplied by a constant value. It is not an exponential function.

x	$f(x)$
0	5
1	8
2	9
3	8
4	5
5	0
6	-7

e. $f(x) = (x - 2)^2 + 9$

This function cannot be used to model the data set. The vertex is $(2, 9)$; however, it is a minimum in this function.

f. $f(x) = -(x + 1)(x - 5)$

This function can be used to model the data set. One of the x -intercepts is $x = 5$, and the second x -intercept is $x = -1$ by following the pattern of the data. The function equation indicates x -intercepts of -1 and 5 , where the vertex is a maximum value of the function.

2. Match each table below to the function and the context, and explain how you made your decision.

A	
x	y
1	9
2	18
3	27
4	18
5	9

Equation h
Context 5

B	
x	y
1	12
2	24
3	36
4	48
5	60

Equation f
Context 4

C	
x	y
0	160
1	174
2	156
3	106
4	24

Equation g
Context 2

D	
x	y
1	2
2	4
3	8
4	16
5	32

Equation p
Context 1

E	
x	y
2	8
3	9
4	8
5	5
6	0

Equation q
Context 3

Equations:

$$f(x) = 12x$$

$$h(x) = -9|x - 3| + 27$$

$$g(x) = -(x)(x - 6)$$

$$p(x) = 2^x$$

$$q(x) = -16x^2 + 30x + 160$$

Contexts:

1. The population of bacteria doubled every month, and the total population vs. time was recorded.
2. A ball was launched upward from the top of a building, and the vertical distance of the ball from the ground vs. time was recorded.
3. The height of a certain animal's vertical leap was recorded at regular time intervals of one second; the animal returned to ground level after six seconds.
4. Melvin saves the same amount of money every month. The total amount saved after each month was recorded.
5. Chris ran at a constant rate on a straight-line path and then returned at the same rate. His distance from his starting point was recorded at regular time intervals.

Closing (3 minutes)

- How do we determine the type of function from a given set of data by analyzing the y -values when the x -values are increasing by a constant increment?
 - *If the first difference is constant, then the data set could be modeled by a linear function.*
 - *If the second difference is constant, then the data set could be modeled by a quadratic function.*
 - *If each y -value is the previous y -value multiplied by a constant, then the data set could be modeled by an exponential function.*

Lesson Summary

The following methods can be used to determine the appropriate model for a given data set as a linear, quadratic, or exponential function:

- If the first difference is constant, then the data set could be modeled by a linear function.
- If the second difference is constant, then the data set could be modeled by a quadratic function.
- If the subsequent y -values are multiplied by a constant, then the data set could be modeled by an exponential function.

Exit Ticket (5 minutes)

Name _____

Date _____

Lesson 2: Analyzing a Data Set

Exit Ticket

Analyze these data sets, recognizing the unique pattern and key feature(s) for each relationship. Then use your findings to fill in the missing data, match to the correct function from the list on the right, and describe the key feature(s) that helped you choose the function.

Table A		Table B		Table C		Table D	
x	y	x	y	x	y	x	y
0	6	0	6	-1	$\frac{1}{6}$	-1	
1	10	1	15	0	1	0	6
2	14	2	18	1		1	8
3		3	15	2	36	2	6
4	22	4		3		3	0
5		5		4	1296	4	
						5	-24

Equations:

$$f(x) = 6^x$$

$$h(x) = -3(x - 2)^2 + 18$$

$$g(x) = -2(x + 1)(x - 3)$$

$$r(x) = 4x + 6$$

Table A: _____ Key Feature(s): _____

Table B: _____ Key Feature(s): _____

Table C: _____ Key Feature(s): _____

Table D: _____ Key Feature(s): _____

Exit Ticket Sample Solutions

Analyze these data sets, recognizing the unique pattern and key feature(s) for each relationship. Then use your findings to fill in the missing data, match to the correct function from the list on the right, and describe the key feature(s) that helped you choose the function.

Table A		Table B		Table C		Table D	
x	y	x	y	x	y	x	y
0	6	0	6	-1	$\frac{1}{6}$	-1	0
1	10	1	15	0	1	0	6
2	14	2	18	1	6	1	8
3	18	3	15	2	36	2	6
4	22	4	6	3	216	3	0
5	26	5	-9	4	1296	4	-10
						5	-24

Equations:

$$f(x) = 6^x$$

$$h(x) = -3(x - 2)^2 + 18$$

$$g(x) = -2(x + 1)(x - 3)$$

$$r(x) = 4x + 6$$

Table A: $r(x)$ Key Feature(s): Constant slope (first difference) is 4

Table B: $h(x)$ Key Feature(s): (2, 18) is the maximum point; Second differences are -6

Table C: $f(x)$ Key Feature(s): Exponential growth; Common ratio is 6

Table D: $g(x)$ Key Feature(s): Includes x-intercepts (-1, 0) and (3, 0); Second differences are -4

Problem Set Sample Solutions

1.

- a. Determine the function type that could be used to model the data set at the right and explain why.

Quadratic—the second difference is 4.

- b. Complete the data set using the special pattern of the function you described above.

See the completed table for answers.

- c. If it exists, find the minimum or maximum value for the function model. If there is no minimum or maximum, explain why.

Minimum value occurs at $(4, -8)$.

x	y
0	24
1	10
2	0
3	-6
4	-8
5	-6
6	0

2.

- a. Determine the function type that could be used to model the data set and explain why.

Exponential— y -value is being multiplied by constant value 4.

- b. Complete the data set using the special pattern of the function you described above.

See responses in the completed table.

- c. If it exists, find the minimum or maximum value for the function model. If there is no minimum or maximum, explain why.

Since this is an exponential function, the y -values will increase as x -values increase, and there will be no maximum value. As x gets smaller and smaller (i.e., moves further to the left on the number line), the y -values approach 0 but will never reach 0 and never become negative. Therefore, there is no minimum value of this function.

x	y
-1	$1/4$
0	1
1	4
2	16
3	64
4	256
5	1024

3.

- a. Determine the function type that could be used to model the data set and explain why.

Linear—the first difference is 6.

- b. Complete the data set using the special pattern of the function you described above.

See responses in the completed table.

- c. If it exists, find the minimum or maximum value for the function model. If there is no minimum or maximum, explain why.

There is no minimum or maximum value for a linear function (except for a horizontal line). For this function, the y -values decrease as the x -values decrease, and the y -values increase as the x -values increase.

x	y
-1	6
0	12
1	18
2	24
3	30
4	36
5	42

4. Circle all the function types that could possibly be used to model a context if the given statement applies.
- a. When x -values are at regular intervals, the first difference of y -values is not constant.
- Linear Function Quadratic Function Exponential Function Absolute Value Function
- b. When x -values are at regular intervals, the second difference of y -values is not constant.
- Linear Function Quadratic Function Exponential Function Absolute Value Function
- c. When x -values are at regular intervals, the quotient of any two consecutive y -values is a constant that is not equal to 0 or 1.
- Linear Function Quadratic Function Exponential Function Absolute Value Function
- d. There may be up to two different x -values for $y = 0$.
- Linear Function Quadratic Function Exponential Function Absolute Value Function