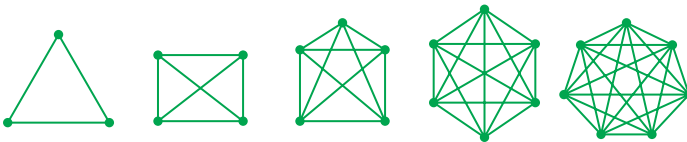## Lesson

11-8

Modeling Data with Polynomials

▶ **BIG IDEA** Given n points in the plane, no two of which have the same first coordinate, it is generally possible to find an equation for a polynomial function that contains all the points.

In Lesson 11-7, you saw how to determine whether a polynomial formula exists for a given function. Now, if the formula exists, you will see how to find the coefficients of the polynomial.

Example 1

In the land of Connectica, it was deemed that all cities built must be joined to every other city directly by a straight road. When the land was first developed, there were three cities and only three roads were needed. After a few years there were more cities and the government found itself having to build more and more roads.

Number of Cities	3	4	5	6	7
Number of Roads	3	6	10	15	21

The city planners wondered if there is a polynomial formula relating r, the number of roads needed, with x, the number of cities built. Does such a formula exist? If so, find the formula.

Solution Use the method of finite differences to determine the degree of a polynomial that will fit the data, if it exists.

Because the 2nd differences are equal, there is a quadratic polynomial that is an exact model for these five data points. That is,

$$r = ax^2 + bx + c.$$

x	3	4	5	6	7
r	3	6	10	15	21

1st differences   3   4   5   6

2nd differences   1   1   1

Now you need to find values of the coefficients a, b, and c. As in Lesson 6-6, we find a, b, and c by solving a system of equations.

First, substitute three known ordered pairs (x, r) into the above equation. We choose $(3, 3)$, $(4, 6)$, and $(5, 10)$.
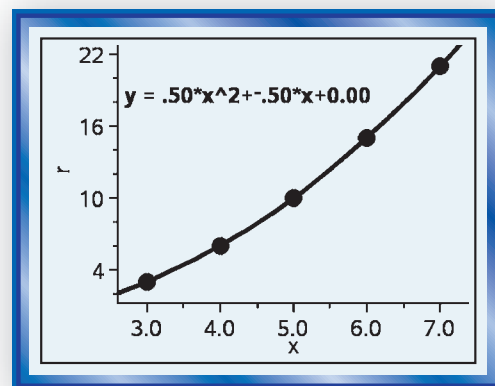
Substitute the ordered pairs into the equation to get the following system.

$$\begin{cases} 3 = a \cdot 3^2 + b \cdot 3 + c \\ 6 = a \cdot 4^2 + b \cdot 4 + c \\ 10 = a \cdot 5^2 + b \cdot 5 + c \end{cases}$$

You may solve this system by using linear combinations or by using matrices. Our matrix solution is shown below. You see that $a = \frac{1}{2}$, $b = -\frac{1}{2}$, and $c = 0$. So, $r = \frac{1}{2}x^2 - \frac{1}{2}x$ models the data from the city planners.

$$\begin{bmatrix} 9 & 3 & 1 \\ 16 & 4 & 1 \\ 25 & 5 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{2} \\ \frac{-1}{2} \\ 0 \end{bmatrix}$$

Check Enter the data into a CAS and run a quadratic regression. We obtain the results at the right. It checks.

y = .50*x^2+-.50*x+0.00

🛑 QY1

Finding Higher-Degree Polynomials

The idea of Example 1 can be used to find a polynomial function of degree greater than 2 that exactly fits some data. Consider the sequence of squares shown on the next page.

Let S be the total number of squares of any size that can be found on an n-by-n checkerboard.

For example, on the 3-by-3 checkerboard there are:

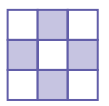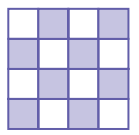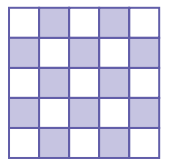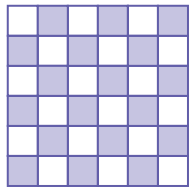9 squares this size: ☐, 4 squares this size: ⊞, and 1 square

this size: ▦ .

So, when $n = 3$, $S = 9 + 4 + 1 = 14$. There are 14 squares on a 3×3 board. The numbers of squares for $n = 1$ through 6 are given on the next page.

n	1	2	3	4	5	6
S	1	5	14	30	55	91

Example 2

How many squares are on an $n \times n$ checkerboard?

Solution First, use the method of finite differences to determine whether a polynomial model fits the data.

n	1	2	3	4	5	6
S	1	5	14	30	55	91
1st differences		4	?	?	?	?
2nd differences			?	?	?	?
3rd differences				?	?	?

The 3rd differences are constant. Therefore, **the data can be modeled by a polynomial function of degree __?__.**

Now use a system of equations to find a polynomial model. You know that the polynomial is of the form $S = an^3 + bn^2 + cn + d$. Substitute $n = 4$, 3, 2, then 1 and the corresponding values of S into the equation and solve the resulting system.

$$\begin{cases} 30 = a(4)^3 + b(4)^2 + c(4) + d \\ 14 = a(3)^3 + b(3)^2 + c(3) + d \\ 5 = a(2)^3 + b(2)^2 + c(2) + d \\ 1 = a(1)^3 + b(1)^2 + c(1) + d \end{cases}$$

$$\begin{bmatrix} 64 & 16 & 4 & 1 \\ 27 & 9 & 3 & 1 \\ 8 & 4 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 30 \\ 14 \\ 5 \\ 1 \end{bmatrix}$$

Thus, $a = $ __?__, $b = $ __?__, $c = $ __?__, and $d = $ __?__.

So, a formula for S in terms of n is $S = $ ___?___.

Check Use *cubic regression* on a CAS to find the model.

Modeling a Finite Set of Points

In Examples 1 and 2, you were asked to find a polynomial function model that related two variables, and in each case the independent variable could take on an infinite number of values. Consequently, you need a way to prove that the polynomial function works for all values. Sometimes you can appeal to the definition of the function (see Question 10). At other times you may need to use methods beyond the scope of this course.

However, when you need a model only to fit a finite number of values in a function, you can always find a polynomial model. Such a model exists even when the data do not follow a simple pattern.

Year	Population
1900	1,850,000
1920	2,284,000
1940	1,890,000
1960	1,698,000
1980	1,428,000
2000	1,537,000

The population of Manhattan Island (part of New York City) has gone up and down over the past 100 years. At the right are a table and a graph showing the population every 20 years from 1900 to 2000.

Let x = the number of 20-year periods since 1900. Let $P(x)$ be the population (in millions) of Manhattan at these times. Here is how to find a polynomial expression for $P(x)$ that contains the six points in the table.

Use the methods of Examples 1 and 2 and find 1st, 2nd, 3rd, 4th, and 5th differences. When you reach the 5th differences, there is only one value. So all the 5th differences are equal! Consequently, by the Polynomial-Difference Theorem, *for these six points* there is a polynomial of degree 5 that fits:
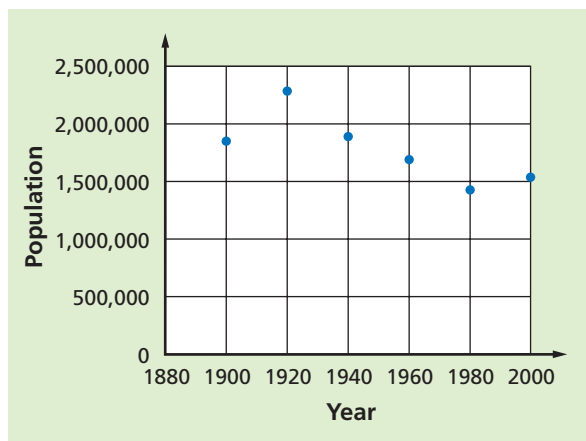
$$P(x) = ax^5 + bx^4 + cx^3 + dx^2 + ex + f.$$

To find a, b, c, d, e, and f, you need to solve a system of six linear equations in six variables. These equations arise from the six given data points. For instance, for the year 1940, $x = 2$, the number of 20-year periods since 1900, and $P(x) = 1.890$, the population of Manhattan in millions. Substituting 2 for x in the general formula above,

$$P(2) = 1.89 = a \cdot 2^5 + b \cdot 2^4 + c \cdot 2^3 + d \cdot 2^2 + e \cdot 2 + f$$
$$= 32a + 16b + 8c + 4d + 2e + f.$$

Similarly, you can find the other five equations. When the system is solved, it turns out that, to four decimal places, $a = 0.0171$, $b = -0.2252$, $c = 1.0962$, $d = -2.3823$, $e = 1.9282$, and $f = 1.85$. So,

$$P(x) = 0.0171x^5 - 0.2252x^4 + 1.0962x^3 - 2.3823x^2 + 1.9282x + 1.85.$$

Unfortunately, polynomial regression on most calculators stops at a 4th-degree polynomial. But some computer spreadsheets and data-analysis applications allow you to find regression equations to 5th and higher degrees. We found the equation for $P(x)$ using a spreadsheet. Although this polynomial model fits the data, because the decimal coefficients are rounded, evaluating the polynomial $P(x)$ will not produce exact populations.

STOP **QY2**

In general, if you have n points of a function, there exists a polynomial formula of some degree less than n that models those n points exactly.

> **▶ QY2**
>
> Substitute 3 for x in the formula for the population of Manhattan. What year's population of Manhattan should you obtain? How close is your result to that year's population?

Limitations of Polynomial Modeling

The model for Manhattan's population is a 5th-degree polynomial because you could take finite differences only five times. If there were more data, then the 5th differences might not be equal and you would need a larger degree polynomial to fit all the points.

In general, if you have n data points through which you want to fit a polynomial function exactly, a polynomial function of any degree $n - 1$ *or more* will work. For instance, suppose you are given only the data at the right. The first differences are 1 and 2, and there is only one 2nd difference. So, there is a quadratic function that fits the data. That function has the formula $y = \frac{x^2 - x + 2}{2}$. If (4, 7) and (5, 11) are two more given data points, then the second differences are still equal and the same formula models all the data.

x	1		2		3
y	1		2		4
1st differences		1		2	
2nd difference			1		

x	1		2		3		4		5
y	1		2		4		7		11
1st differences		1		2		3		4	
2nd differences			1		1		1		

However, if (4, 8) and (5, 15) are the next data points, then the second differences are no longer equal. The 3rd differences are equal, so there is a 3rd-degree polynomial equation modeling the data: $y = \frac{x^3 - 3x^2 + 8x}{6}$.

x	1		2		3		4		5
y	1		2		4		8		15
1st differences		1		2		4		7	
2nd differences			1		2		3		
3rd differences				1		1			

In this way, you can see that $y = \frac{x^2 - x + 2}{2}$ and $y = \frac{x^3 - 3x^2 + 8x}{6}$ are only two of many polynomial formulas fitting the original three data points (1, 1), (2, 2), and (3, 4).

Since many different polynomial functions can fit a set of n data points, polynomial models based on only a few points are usually not good models for making predictions. However, polynomial models do provide an efficient way to store data because the model fits all the given points exactly.

Questions

COVERING THE IDEAS

In 1–3, refer to Example 1.

1. How did the planners know what degree polynomial model would fit the data?

2. Show that the model is correct for $x = 6$.

3. Use the model to calculate how many roads would be needed for 20 cities.

In 4 and 5, refer to Example 2.

4. How many squares of any size are on a standard 8-by-8 checkerboard?

5. How many more squares of any size are created when you turn an 8-by-8 checkerboard into a 9-by-9 checkerboard?

6. Consider the data at the right.
   a. Determine the lowest possible degree of a polynomial function that fits these data.
   b. Find a formula for the function.

x	0	1	2	3	4	5
y	1	2	7	46	173	466

7. Suppose the data in the table at the right are modeled by a formula of the form $y = ax^2 + bx + c$. What three equations are satisfied by a, b, and c?

x	2	5	8
y	12	60	162

8. Refer to the formula modeling Manhattan's population. Verify that the model accurately stores the population for the given year.
   a. 1920
   b. 2000

APPLYING THE MATHEMATICS

9. Ronin did not use matrices to solve the system of Example 1. Instead he reordered the equations so that the largest coefficients are on the top line as shown below, then repeatedly subtracted each equation from the one above it. Use this method to solve the system and check the solution to Example 1.

$$\begin{cases} 25a + 5b + c = 10 \\ 16a + 4b + c = 6 \\ 9a + 3b + c = 3 \end{cases}$$
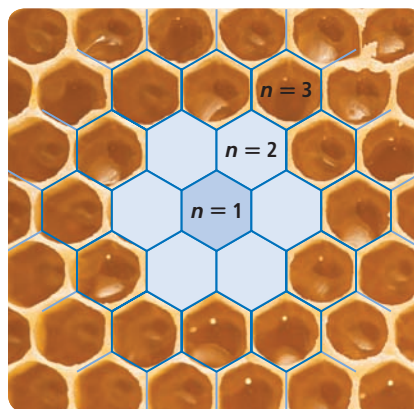
10. Example 1 shows that $r = \frac{1}{2}x^2 - \frac{1}{2}x$ fits the values of the function for integers x with $1 \leq x \leq 6$. Check that this formula works for any integer x by going through the following steps.

   a. Let $r = f(x)$ and calculate $f(n)$ and $f(n-1)$.

   b. The expression $f(n) - f(n-1)$ is the difference between connecting n cities and connecting $n-1$ cities. Find an expression for $f(n) - f(n-1)$ in terms of n.

   c. Interpret your answer to Part b by referring back to the meaning of $f(n)$.

11. The employees at Primo's Pizzeria like to cut pizza into oddly-shaped pieces. In so doing, they noticed that there is a maximum number of pieces that can be formed from a given number of cuts. Write a polynomial model for finding the maximum number of pieces from the number of cuts.

Number of Cuts	0	1	2	3	4
Maximum Number of Pieces	1	2	4	7	11

12. Recall that a *tessellation* is a pattern of shapes that covers a surface completely without overlaps or gaps. One cross section of a honeycomb is a tessellation of regular hexagons, with three hexagons meeting at each vertex. One way to construct the tessellation is to start with 1 hexagon then surround it with 6 more hexagons and then surround these with another "circle" of 12 hexagons, and so on. If this pattern were to continue, find

   a. the number of hexagons in the fourth circle.

   b. the total number of hexagons in the first four circles.

   c. a polynomial equation which expresses the *total* number of hexagons h as a function of the number of circles n.

   d. the total number of hexagons in a honeycomb with 10 circles.

13. Refer to the roller coaster on the first page of the chapter. The table below gives the roller coaster's height $H(x)$ in the picture for different values of horizontal distance x along the picture.

x	0	4	8	12	16
$H(x)$	0	–0.7	2	7.8	11.4

   a. Use the method of Examples 1 and 2 to fit a polynomial function to these data points.

   b. Graph the polynomial you find to see how close its graph is to the picture.

14. The drawings given in Example 1 look similar to the drawings you saw in geometry when you learned that the number d of diagonals in a polygon with n sides is given by $d = \dfrac{n(n-3)}{2}$.

    a. Show that the formula for the number of diagonals and the formula for the number of roads are not equivalent.

    b. Subtract the polynomial for the number of diagonals from the polynomial for the number of connections. What is the difference? What does the difference mean?

REVIEW

15. Consider the data at the right. **(Lesson 11-7)**

    a. Can the data be modeled by a polynomial function?

    b. If so, what degree is the polynomial?

r	1	2	3	4	5	6	7	8	9
t	3	11	31	69	131	223	351	521	739

16. How much larger is the volume of a cube with side $x + 1$ than the volume of a cube with side $x - 1$? **(Lesson 11-2)**

17. May Vary invested different amounts each year for the past 6 years as summarized in the table at the right. Each amount is invested at 4.6% APY. What is the total current value of her investments? **(Lesson 11-1)**

18. Describe three properties of the graph of $f(x) = \sin x$. **(Lesson 10-6)**

19. Express as a logarithm of a single number. **(Lesson 9-9)**

    a. $\log_7 36 - \log_7 6$ b. $4 \log 3$ c. $\dfrac{1}{6} \log 64$

20. Ivan Speeding was driving 25% faster than the speed limit. By what percent must he reduce his speed to be driving at the speed limit? **(Previous Course)**

Years Ago	Amount ($)
5	7100
4	6600
3	5800
2	5100
1	6000
0	5500

EXPLORATION

21. In the last part of this lesson, the functions with equations $y = \dfrac{x^2 - x + 2}{2}$ and $y = \dfrac{x^3 - 3x^2 + 8x}{6}$ are shown to fit the points $(1, 1)$, $(2, 2)$, and $(3, 4)$.

    a. Explain why $y = \dfrac{x^3 - 3x^2 + 8x}{6} + (x - 1)(x - 2)(x - 3)$ describes another 3rd degree polynomial function that contains these points.

    b. Find an equation for a polynomial function of 4th degree that contains these points.

    c. Explain how you could find as many different polynomials as desired that contain these three points.

QY ANSWERS

1. 45 roads

2. 1960; $P(3) = 1.7054$, or 1,705,400 people. This is 7400 more than the actual population.