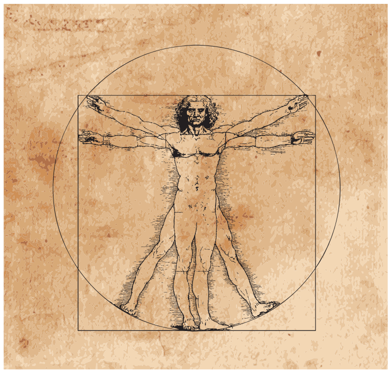
P.4 Fitting Models to Data

- Fit a linear model to a real-life data set.
- Fit a quadratic model to a real-life data set.
- Fit a trigonometric model to a real-life data set.



A computer graphics drawing based on the pen and ink drawing of Leonardo da Vinci's famous study of human proportions, called *Vitruvian Man*

Fitting a Linear Model to Data

A basic premise of science is that much of the physical world can be described mathematically and that many physical phenomena are predictable. This scientific outlook was part of the scientific revolution that took place in Europe during the late 1500s. Two early publications connected with this revolution were *On the Revolutions of the Heavenly Spheres* by the Polish astronomer Nicolaus Copernicus and *On the Fabric of the Human Body* by the Belgian anatomist Andreas Vesalius. Each of these books was published in 1543, and each broke with prior tradition by suggesting the use of a scientific method rather than unquestioned reliance on authority.

One basic technique of modern science is gathering data and then describing the data with a mathematical model. For instance, the data in Example 1 are inspired by Leonardo da Vinci's famous drawing that indicates that a person's height and arm span are equal.

EXAMPLE 1 Fitting a Linear Model to Data

See LarsonCalculus.com for an interactive version of this type of example.

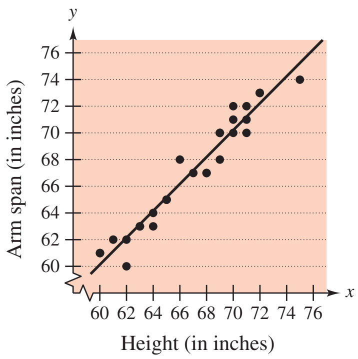
A class of 28 people collected the data shown below, which represent their heights x and arm spans y (rounded to the nearest inch).

$$(60, 61), (65, 65), (68, 67), (72, 73), (61, 62), (63, 63), (70, 71),$$
$$(75, 74), (71, 72), (62, 60), (65, 65), (66, 68), (62, 62), (72, 73),$$
$$(70, 70), (69, 68), (69, 70), (60, 61), (63, 63), (64, 64), (71, 71),$$
$$(68, 67), (69, 70), (70, 72), (65, 65), (64, 63), (71, 70), (67, 67)$$

Find a linear model to represent these data.

Solution There are different ways to model these data with an equation. The simplest would be to observe that x and y are about the same and list the model as simply $y = x$. A more careful analysis would be to use a procedure from statistics called linear regression. (You will study this procedure in Section 13.9.) The least squares regression line for these data is

$$y = 1.006x - 0.23. \qquad \text{Least squares regression line}$$



Linear model and data
Figure P.32

The graph of the model and the data are shown in Figure P.32. From this model, you can see that a person's arm span tends to be about the same as his or her height.

TECHNOLOGY Many graphing utilities have built-in least squares regression programs. Typically, you enter the data into the calculator and then run the linear regression program. The program usually displays the slope and y-intercept of the best-fitting line and the *correlation coefficient* r. The correlation coefficient gives a measure of how well the data can be modeled by a line. The closer $|r|$ is to 1, the better the data can be modeled by a line. For instance, the correlation coefficient for the model in Example 1 is $r \approx 0.97$, which indicates that the linear model is a good fit for the data. If the r-value is positive, then the variables have a positive correlation, as in Example 1. If the r-value is negative, then the variables have a negative correlation.

Hal_P/Shutterstock.com

Fitting a Quadratic Model to Data

A function that gives the height s of a falling object in terms of the time t is called a *position function*. If air resistance is not considered, then the position of a falling object can be modeled by

$$s(t) = \frac{1}{2}gt^2 + v_0t + s_0$$

where g is the acceleration due to gravity, v_0 is the initial velocity, and s_0 is the initial height. The value of g depends on where the object is dropped. On Earth, g is approximately -32 feet per second per second, or -9.8 meters per second per second.

To discover the value of g experimentally, you could record the heights of a falling object at several increments, as shown in Example 2.

EXAMPLE 2 Fitting a Quadratic Model to Data

A basketball is dropped from a height of about $5\frac{1}{4}$ feet. The height of the basketball is recorded 23 times at intervals of about 0.02 second. The results are shown in the table.

Time	0.0	0.02	0.04	0.06	0.08	0.099996
Height	5.23594	5.20353	5.16031	5.0991	5.02707	4.95146

Time	0.119996	0.139992	0.159988	0.179988	0.199984	0.219984
Height	4.85062	4.74979	4.63096	4.50132	4.35728	4.19523

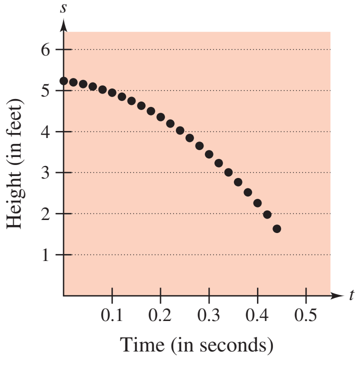
Time	0.23998	0.25993	0.27998	0.299976	0.319972	0.339961
Height	4.02958	3.84593	3.65507	3.44981	3.23375	3.01048

Time	0.359961	0.379951	0.399941	0.419941	0.439941
Height	2.76921	2.52074	2.25786	1.98058	1.63488

Find a model to fit these data. Then use the model to predict the time when the basketball will hit the ground.

Solution Begin by sketching a scatter plot of the data, as shown in Figure P.33. From the scatter plot, you can see that the data do not appear to be linear. It does appear, however, that they might be quadratic. To check this, enter the data into a graphing utility that has a quadratic regression program. You should obtain the model

$$s = -15.45t^2 - 1.302t + 5.2340. \qquad \text{Least squares regression quadratic}$$



Scatter plot of data
Figure P.33

Using this model, you can predict the time when the basketball hits the ground by substituting 0 for s and solving the resulting equation for t.

$$0 = -15.45t^2 - 1.302t + 5.2340 \qquad \text{Let } s = 0.$$

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \text{Quadratic Formula}$$

$$t = \frac{-(-1.302) \pm \sqrt{(-1.302)^2 - 4(-15.45)(5.2340)}}{2(-15.45)} \qquad \text{Substitute } a = -15.45,\ b = -1.302, \text{ and } c = 5.2340.$$

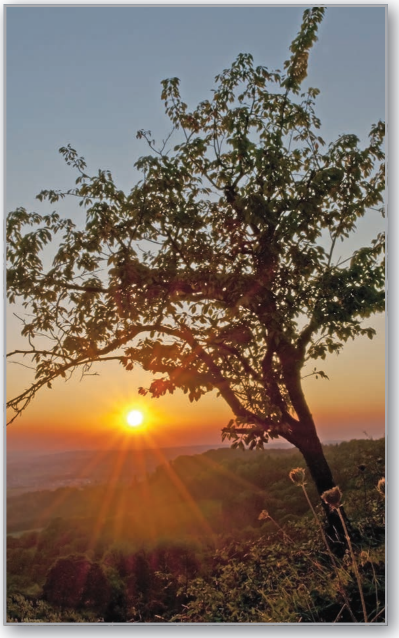
$$t \approx 0.54 \qquad \text{Choose positive solution.}$$

The solution is about 0.54 second. In other words, the basketball will continue to fall for about 0.1 second more before hitting the ground. (Note that the experimental value of g is $\frac{1}{2}g = -15.45$, or $g = -30.90$ feet per second per second.)

Fitting a Trigonometric Model to Data

What is mathematical modeling? This is one of the questions that is asked in the book *Guide to Mathematical Modelling*. Here is part of the answer.*

1. Mathematical modeling consists of applying your mathematical skills to obtain useful answers to real problems.
2. Learning to apply mathematical skills is very different from learning mathematics itself.
3. Models are used in a very wide range of applications, some of which do not appear initially to be mathematical in nature.
4. Models often allow quick and cheap evaluation of alternatives, leading to optimal solutions that are not otherwise obvious.
5. There are no precise rules in mathematical modeling and no "correct" answers.
6. Modeling can be learned only by *doing*.



The amount of daylight received by locations on Earth varies with the time of year.

REMARK For a review of trigonometric functions, see Appendix C.

EXAMPLE 3 **Fitting a Trigonometric Model to Data**

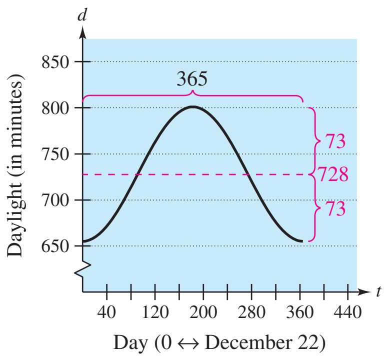
The number of hours of daylight on a given day on Earth depends on the latitude and the time of year. Here are the numbers of minutes of daylight at a location of 20°N latitude on the longest and shortest days of the year: June 21, 801 minutes; December 22, 655 minutes. Use these data to write a model for the amount of daylight d (in minutes) on each day of the year at a location of 20°N latitude. How could you check the accuracy of your model?

Solution Here is one way to create a model. You can hypothesize that the model is a sine function whose period is 365 days. Using the given data, you can conclude that the amplitude of the graph is $(801 - 655)/2$, or 73. So, one possible model is

$$d = 728 - 73\sin\left(\frac{2\pi t}{365} + \frac{\pi}{2}\right).$$

In this model, t represents the number of each day of the year, with December 22 represented by $t = 0$. A graph of this model is shown in Figure P.34. To check the accuracy of this model, a weather almanac was used to find the numbers of minutes of daylight on different days of the year at the location of 20°N latitude.

Date	Value of t	Actual Daylight	Daylight Given by Model
Dec 22	0	655 min	655 min
Jan 1	10	657 min	656 min
Feb 1	41	676 min	672 min
Mar 1	69	705 min	701 min
Apr 1	100	740 min	739 min
May 1	130	772 min	773 min
Jun 1	161	796 min	796 min
Jun 21	181	801 min	801 min
Jul 1	191	799 min	800 min
Aug 1	222	782 min	785 min
Sep 1	253	752 min	754 min
Oct 1	283	718 min	716 min
Nov 1	314	685 min	681 min
Dec 1	344	661 min	660 min



Graph of model
Figure P.34

You can see that the model is fairly accurate.

* Text from Dilwyn Edwards and Mike Hamson, *Guide to Mathematical Modelling* (Boca Raton: CRC Press, 1990), p. 4. Used by permission of the authors.

hjschneider/iStockphoto.com

P.4 Exercises

See CalcChat.com for tutorial help and worked-out solutions to odd-numbered exercises.

1. Wages Each ordered pair gives the average weekly wage x for federal government workers and the average weekly wage y for state government workers for 2001 through 2009. *(Source: U.S. Bureau of Labor Statistics)*

(941, 727), (1001, 754), (1043, 770), (1111, 791), (1151, 812), (1198, 844), (1248, 883), (1275, 923), (1303, 937)

(a) Plot the data. From the graph, do the data appear to be approximately linear?

(b) Visually find a linear model for the data. Graph the model.

(c) Use the model to approximate y when $x = 1075$.

2. Quiz Scores The ordered pairs represent the scores on two consecutive 15-point quizzes for a class of 15 students.

(7, 13), (9, 7), (14, 14), (15, 15), (10, 15), (9, 7), (11, 14), (7, 14), (14, 11), (14, 15), (8, 10), (15, 9), (10, 11), (9, 10), (11, 10)

(a) Plot the data. From the graph, does the relationship between consecutive scores appear to be approximately linear?

(b) If the data appear to be approximately linear, find a linear model for the data. If not, give some possible explanations.

3. Hooke's Law Hooke's Law states that the force F required to compress or stretch a spring (within its elastic limits) is proportional to the distance d that the spring is compressed or stretched from its original length. That is, $F = kd$, where k is a measure of the stiffness of the spring and is called the *spring constant*. The table shows the elongation d in centimeters of a spring when a force of F newtons is applied.

F	20	40	60	80	100
d	1.4	2.5	4.0	5.3	6.6

(a) Use the regression capabilities of a graphing utility to find a linear model for the data.

(b) Use a graphing utility to plot the data and graph the model. How well does the model fit the data? Explain.

(c) Use the model to estimate the elongation of the spring when a force of 55 newtons is applied.

4. Falling Object In an experiment, students measured the speed s (in meters per second) of a falling object t seconds after it was released. The results are shown in the table.

t	0	1	2	3	4
s	0	11.0	19.4	29.2	39.4

(a) Use the regression capabilities of a graphing utility to find a linear model for the data.

(b) Use a graphing utility to plot the data and graph the model. How well does the model fit the data? Explain.

(c) Use the model to estimate the speed of the object after 2.5 seconds.

5. Energy Consumption and Gross National Product The data show the per capita energy consumptions (in millions of Btu) and the per capita gross national incomes (in thousands of U.S. dollars) for several countries in 2008. *(Source: U.S. Energy Information Administration and The World Bank)*

Argentina	(81, 7.19)	India	(17, 1.04)
Australia	(274, 40.24)	Italy	(136, 35.46)
Bangladesh	(6, 0.52)	Japan	(172, 38.13)
Brazil	(54, 7.30)	Mexico	(66, 9.99)
Canada	(422, 43.64)	Poland	(101, 11.73)
Ecuador	(35, 3.69)	Turkey	(57, 9.02)
Hungary	(110, 12.81)	Venezuela	(121, 9.23)

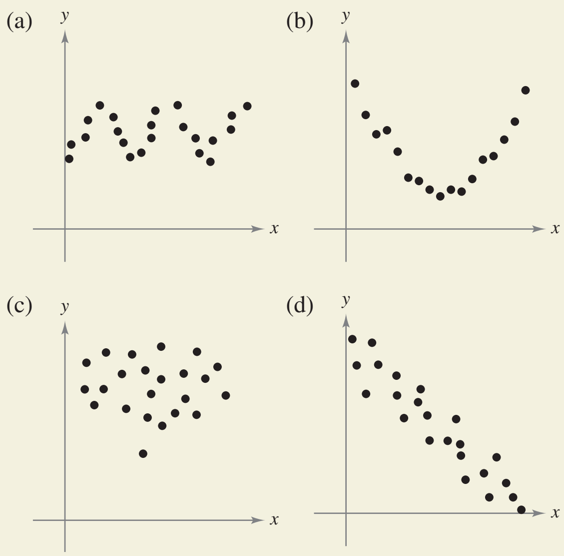
(a) Use the regression capabilities of a graphing utility to find a linear model for the data. What is the correlation coefficient?

(b) Use a graphing utility to plot the data and graph the model.

(c) Interpret the graph in part (b). Use the graph to identify the three countries that differ most from the linear model.

(d) Delete the data for the three countries identified in part (c). Fit a linear model to the remaining data and give the correlation coefficient.

6. HOW DO YOU SEE IT? Determine whether the data can be modeled by a linear function, a quadratic function, or a trigonometric function, or that there appears to be no relationship between x and y.



7. **Beam Strength** Students in a lab measured the breaking strength S (in pounds) of wood 2 inches thick, x inches high, and 12 inches long. The results are shown in the table.

x	4	6	8	10	12
S	2370	5460	10,310	16,250	23,860

(a) Use the regression capabilities of a graphing utility to find a quadratic model for the data.

(b) Use a graphing utility to plot the data and graph the model.

(c) Use the model to approximate the breaking strength when $x = 2$.

(d) How many times greater is the breaking strength for a 4-inch-high board than for a 2-inch-high board?

(e) How many times greater is the breaking strength for a 12-inch-high board than for a 6-inch-high board? When the height of a board increases by a factor, does the breaking strength increase by the same factor? Explain.

8. **Car Performance** The time t (in seconds) required to attain a speed of s miles per hour from a standing start for a Volkswagen Passat is shown in the table. *(Source: Car & Driver)*

s	30	40	50	60	70	80	90
t	2.7	3.8	4.9	6.3	8.0	9.9	12.2

(a) Use the regression capabilities of a graphing utility to find a quadratic model for the data.

(b) Use a graphing utility to plot the data and graph the model.

(c) Use the graph in part (b) to state why the model is not appropriate for determining the times required to attain speeds of less than 20 miles per hour.

(d) Because the test began from a standing start, add the point (0, 0) to the data. Fit a quadratic model to the revised data and graph the new model.

(e) Does the quadratic model in part (d) more accurately model the behavior of the car? Explain.

9. **Engine Performance** A V8 car engine is coupled to a dynamometer, and the horsepower y is measured at different engine speeds x (in thousands of revolutions per minute). The results are shown in the table.

x	1	2	3	4	5	6
y	40	85	140	200	225	245

(a) Use the regression capabilities of a graphing utility to find a cubic model for the data.

(b) Use a graphing utility to plot the data and graph the model.

(c) Use the model to approximate the horsepower when the engine is running at 4500 revolutions per minute.

10. **Boiling Temperature** The table shows the temperatures T (in degrees Fahrenheit) at which water boils at selected pressures p (in pounds per square inch). *(Source: Standard Handbook for Mechanical Engineers)*

p	5	10	14.696 (1 atmosphere)	20
T	162.24°	193.21°	212.00°	227.96°

p	30	40	60	80	100
T	250.33°	267.25°	292.71°	312.03°	327.81°

(a) Use the regression capabilities of a graphing utility to find a cubic model for the data.

(b) Use a graphing utility to plot the data and graph the model.

(c) Use the graph to estimate the pressure required for the boiling point of water to exceed 300°F.

(d) Explain why the model would not be accurate for pressures exceeding 100 pounds per square inch.

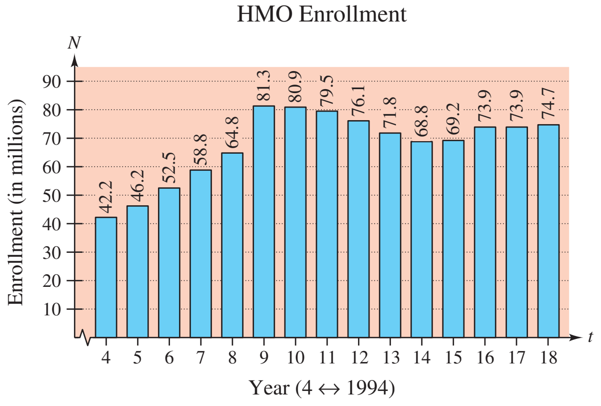
11. **Automobile Costs** The data in the table show the variable costs of operating an automobile in the United States for 2000 through 2010, where t is the year, with $t = 0$ corresponding to 2000. The functions y_1, y_2, and y_3 represent the costs in cents per mile for gas, maintenance, and tires, respectively. *(Source: Bureau of Transportation Statistics)*

t	y_1	y_2	y_3
0	6.9	3.6	1.7
1	7.9	3.9	1.8
2	5.9	4.1	1.8
3	7.2	4.1	1.8
4	6.5	5.4	0.7
5	9.5	4.9	0.7
6	8.9	4.9	0.7
7	11.7	4.6	0.7
8	10.1	4.6	0.8
9	11.4	4.5	0.8
10	12.3	4.4	1.0

(a) Use the regression capabilities of a graphing utility to find cubic models for y_1 and y_3, and a quadratic model for y_2.

(b) Use a graphing utility to graph y_1, y_2, y_3, and $y_1 + y_2 + y_3$ in the same viewing window. Use the model to estimate the total variable cost per mile in 2014.

12. Health Maintenance Organizations The bar graph shows the numbers of people N (in millions) receiving care in HMOs for the years 1994 through 2008. *(Source: HealthLeaders-InterStudy)*



(a) Let t be the time in years, with $t = 4$ corresponding to 1994. Use the regression capabilities of a graphing utility to find linear and cubic models for the data.

(b) Use a graphing utility to plot the data and graph the linear and cubic models.

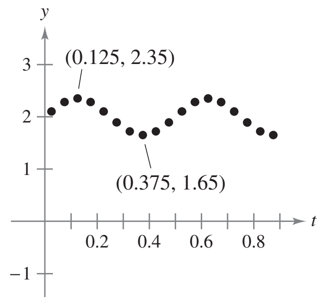
(c) Use the graphs in part (b) to determine which is the better model.

(d) Use a graphing utility to find and graph a quadratic model for the data. How well does the model fit the data? Explain.

(e) Use the linear and cubic models to estimate the number of people receiving care in HMOs in the year 2014. What do you notice?

(f) Use a graphing utility to find other models for the data. Which models do you think best represent the data? Explain.

13. Harmonic Motion The motion of an oscillating weight suspended by a spring was measured by a motion detector. The data collected and the approximate maximum (positive and negative) displacements from equilibrium are shown in the figure. The displacement y is measured in centimeters, and the time t is measured in seconds.



(a) Is y a function of t? Explain.

(b) Approximate the amplitude and period of the oscillations.

(c) Find a model for the data.

(d) Use a graphing utility to graph the model in part (c). Compare the result with the data in the figure.

14. Temperature The table shows the normal daily high temperatures for Miami M and Syracuse S (in degrees Fahrenheit) for month t, with $t = 1$ corresponding to January. *(Source: National Oceanic and Atmospheric Administration)*

t	1	2	3	4	5	6
M	76.5	77.7	80.7	83.8	87.2	89.5
S	31.4	33.5	43.1	55.7	68.5	77.0

t	7	8	9	10	11	12
M	90.9	90.6	89.0	85.4	81.2	77.5
S	81.7	79.6	71.4	59.8	47.4	36.3

(a) A model for Miami is

$$M(t) = 83.70 + 7.46 \sin(0.4912t - 1.95).$$

Find a model for Syracuse.

(b) Use a graphing utility to plot the data and graph the model for Miami. How well does the model fit?

(c) Use a graphing utility to plot the data and graph the model for Syracuse. How well does the model fit?

(d) Use the models to estimate the average annual temperature in each city. Which term of the model did you use? Explain.

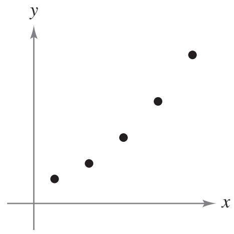
(e) What is the period of each model? Is it what you expected? Explain.

(f) Which city has a greater variability in temperature throughout the year? Which factor of the models determines this variability? Explain.

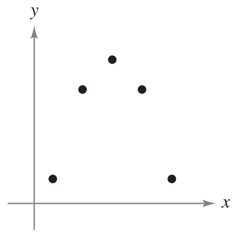
WRITING ABOUT CONCEPTS

Modeling Data **In Exercises 15 and 16, describe a possible real-life situation for each data set. Then describe how a model could be used in the real-life setting.**

15.



16.



PUTNAM EXAM CHALLENGE

17. For $i = 1, 2$, let T_i be a triangle with side lengths a_i, b_i, c_i, and area A_i. Suppose that $a_1 \le a_2$, $b_1 \le b_2$, $c_1 \le c_2$, and that T_2 is an acute triangle. Does it follow that $A_1 \le A_2$?

This problem was composed by the Committee on the Putnam Prize Competition.