

# MATHEMATICAL MODELS

## MATHEMATICAL MODELS

A **mathematical model** of a physical law is a description of that law in the language of mathematics. Such models make it possible to use mathematical methods to deduce results about the physical world that are not evident or have never been observed. For example, the possibility of placing a satellite in orbit around the Earth was deduced mathematically from Issac Newton's model of mechanics nearly 200 years before the launching of *Sputnik*, and Albert Einstein (1879–1955) gave a relativistic model of mechanics in 1915 that explained a precession (position shift) in the perihelion of the planet Mercury that was not confirmed by physical measurement until 1967.

In a typical modeling situation, a scientist wants to obtain a mathematical relationship between two variables  $x$  and  $y$  using a set of  $n$  ordered pairs of measurements

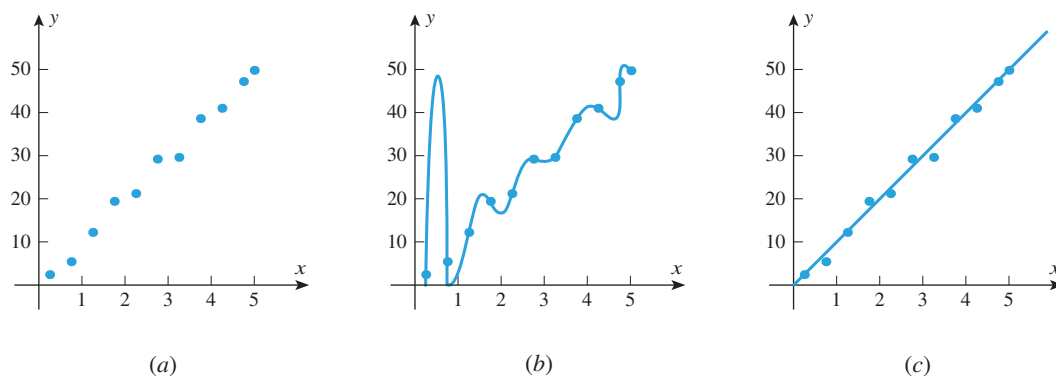
$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n) \quad (1)$$

that relate corresponding values of the variables. We distinguish between two types of physical phenomena—**deterministic phenomena** in which each value of  $x$  determines one value of  $y$ , and **probabilistic phenomena** in which the value of  $y$  associated with a specific  $x$  is not uniquely determined. For example, if  $y$  is the amount that a force  $x$  stretches a certain spring, then each value of  $x$  determines a unique  $y$ , so this is a deterministic phenomenon. In contrast, if  $y$  is the weight of a person whose height is  $x$ , then  $y$  is not uniquely determined by  $x$ , since people with the same height can have different weights. Nevertheless, there is a “correlation” between weight and height that makes it more likely for a taller person to weigh more, so this is a probabilistic phenomenon. We will be concerned with deterministic phenomena only.

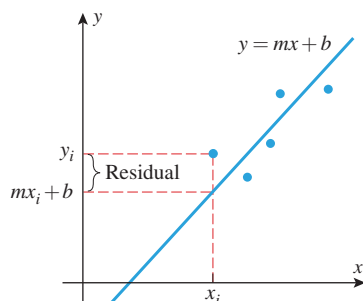
In a deterministic model the variable  $y$  is a function of the variable  $x$ , and the goal is to find a formula  $y = f(x)$  that best describes the data. One way to model a set of deterministic data is to look for a function  $f$  whose graph passes through all of the data points; this is called an **interpolating function**. Although interpolating functions are appropriate in certain situations, they do not adequately account for measurement errors in data. For example, suppose that the relationship between  $x$  and  $y$  is known to be linear but accuracy limitations in the measuring devices and random variations in experimental conditions produce the data plotted in Figure N.1a. One could use a computer program to find a polynomial of degree 10 whose graph passes through all of the data points (Figure N.1b). However, such a polynomial model does not successfully convey the underlying linear relationship. A better approach is to look for a linear equation  $y = mx + b$  whose graph more accurately describes the linear relationship, even if its graph does not pass through all (or any) of the data points (Figure N.1c).

## LINEAR MODELS

The most important methods for finding linear models are based on the following idea: For any proposed linear model  $y = mx + b$ , draw a vertical connector from each data point  $(x_i, y_i)$  to the line, and consider the differences  $y_i - y$  (Figure N.2). These differences,



▲ Figure N.1



▲ Figure N.2

Most graphing calculators, CAS programs, and spreadsheet programs provide methods for finding regression lines. You will need access to one of these technologies for the examples in this section.

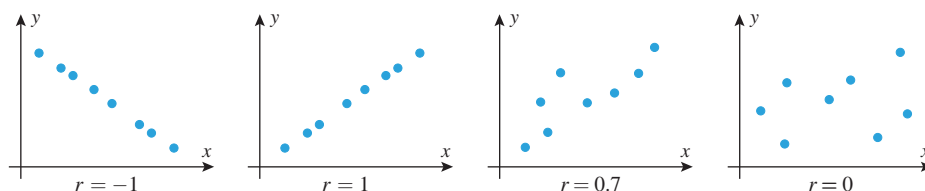
which are called **residuals**, may be viewed as “errors” that result when the line is used to model the data. Points above the line have positive errors, points below the line have negative errors, and points on the line have no error.

One way to choose a linear model is to look for a line  $y = mx + b$  in which the sum of the residuals is zero, the logic being that this makes the positive and negative errors balance out. However, one can find examples where this procedure produces unacceptably poor models, so for reasons that we cannot discuss here the most common method for finding a linear model is to look for a line  $y = mx + b$  in which the *sum of the squares* of the residuals is as small as possible. This is called the **least squares line of best fit** or the **regression line**.

It is possible to compute a regression line, even in cases where the data have no apparent linear pattern. Thus, it is important to have some quantitative method of determining whether a linear model is appropriate for the data. The most common measure of linearity in data is called the **correlation coefficient**. Following convention, we denote the correlation coefficient by the letter  $r$ . Although a detailed discussion of correlation coefficients is beyond the scope of this text, here are some of the basic facts:

- The values of  $r$  are in the interval  $-1 \leq r \leq 1$ , where  $r$  has the same sign as the slope of the regression line.
- If  $r$  is equal to 1 or  $-1$ , then the data points all lie on a line, so a linear model is a perfect fit for the data.
- If  $r = 0$ , then the data points exhibit no linear tendency, so a linear model is inappropriate for the data.

The closer  $r$  is to 1 or  $-1$ , the more tightly the data points hug the regression line and the more appropriate the regression line is as a model; the closer  $r$  is to 0, the more scattered the points and the less appropriate the regression line is as a model (Figure N.3).



▲ Figure N.3

Roughly stated, the value of  $r^2$  is a measure of the percentage of data points that fall in a “tight linear band.” Thus  $r = 0.5$  means that 25% of the points fall in a tight linear band, and  $r = 0.9$  means that 81% of the points fall in a tight linear band. (A precise explanation of what is meant by a “tight linear band” requires ideas from statistics.)

Table N.1

TEMPERATURE $T$ (°C)	PRESSURE $p$ (atm)
0	2.54
50	3.06
100	3.46
150	4.00
200	4.41

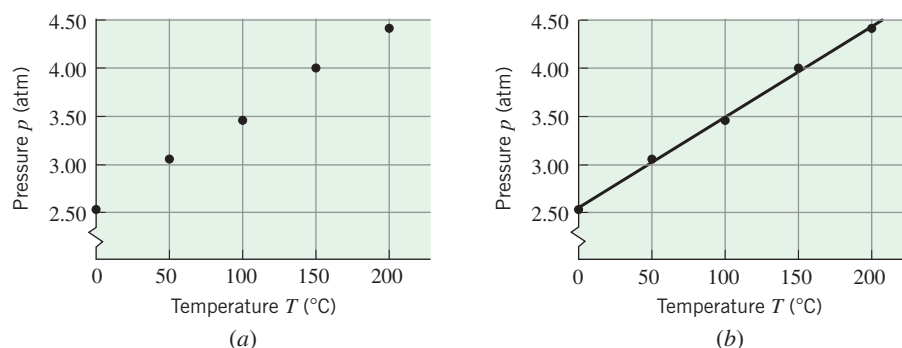
► **Example 1** Table N.1 gives a set of data points relating the pressure  $p$  in atmospheres (atm) and the temperature  $T$  (in °C) of a fixed quantity of carbon dioxide in a closed cylinder. The associated plot in Figure N.4a suggests that there is a linear relationship between the pressure and the temperature.

- Use your calculating utility to find the least squares line for the data. If your utility can produce the correlation coefficient, then find it.
- Use the model obtained in part (a) to predict the pressure when the temperature is 250°C.
- Use the model obtained in part (a) to predict a temperature at which the pressure of the gas will be zero.

**Solution (a).** The least squares line is given by  $p = 0.00936T + 2.558$  (Figure N.4b) with correlation coefficient  $r = 0.998979$ .

**Solution (b).** If  $T = 250$ , then  $p = (0.00936)(250) + 2.558 = 4.898$  (atm).

**Solution (c).** Solving the equation  $0 = p = 0.00936T + 2.558$  yields  $T \approx -273.291$  °C. ◀



▲ Figure N.4

It is not always convenient (or necessary) to obtain the least squares line for a linear phenomenon in order to create a model. In some cases, more elementary methods suffice. Here is an example.

► **Example 2** Figure N.5a shows a graph of temperature versus altitude that was transmitted by the *Magellan* spacecraft when it entered the atmosphere of Venus in October 1991. The graph strongly suggests that there is a linear relationship between temperature and altitude for altitudes between 35 and 60 km.

- Use the graph transmitted by the *Magellan* spacecraft to find a linear model of temperature versus altitude in the Venusian atmosphere that is valid for altitudes between 35 and 60 km.
- Use the model obtained in part (a) to estimate the temperature at the surface of Venus, and discuss the assumptions you are making in obtaining the estimate.

**Solution (a).** Let  $T$  be the temperature in kelvins and  $h$  the altitude in kilometers. We will first estimate the slope  $m$  of the linear portion of the graph, then estimate the coordinates of a data point  $(h_1, T_1)$  on that portion of the graph, and then use the point-slope form of a line

$$T - T_1 = m(h - h_1) \quad (2)$$

The graph nearly passes through the point  $(60, 250)$ , so we will take  $h_1 \approx 60$  and  $T_1 \approx 250$ . In Figure N.5b we have sketched a line that closely approximates the linear portion of the

data. Using the intersections of that line with the edges of the grid box, we estimate the slope to be

$$m \approx \frac{100 - 490}{78 - 30} = -\frac{390}{48} = -8.125 \text{ K/km}$$

Substituting our estimates of  $h_1$ ,  $T_1$ , and  $m$  into (2) yields the equation

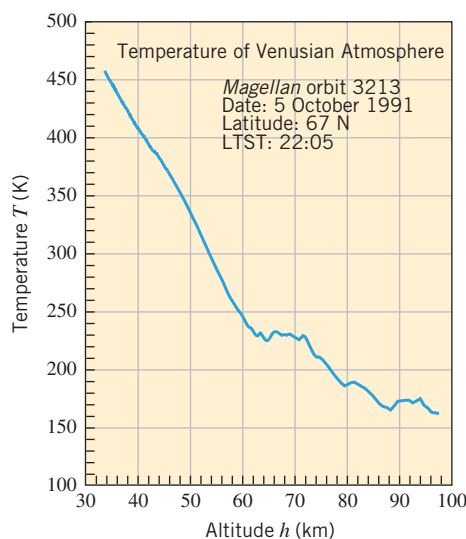
$$T - 250 = -8.125(h - 60)$$

or, equivalently,

$$T = -8.125h + 737.5 \quad (3)$$

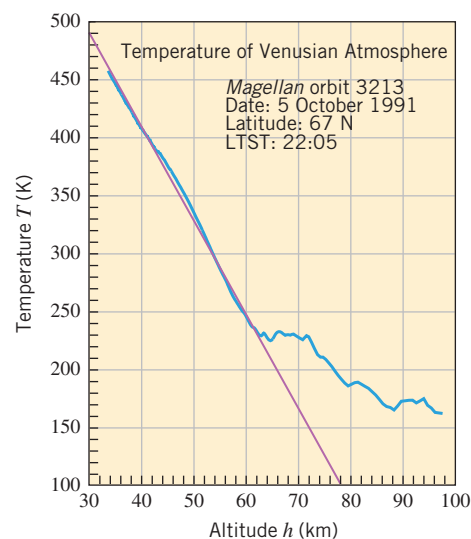
The method of Example 2 is crude, at best, since it relies on extracting rough estimates of numerical data from a graph. Nevertheless, the final result is quite good, since information from NASA places the surface temperature of Venus at about 737 K (hot enough to melt lead).

**Solution (b).** The *Magellan* spacecraft stopped transmitting data at an altitude of approximately 35 km, so we cannot be certain that the linear model applies at lower altitudes. However, if we *assume* that the model is valid at all lower altitudes, then we can approximate the temperature at the surface of Venus by setting  $h = 0$  in (3). We obtain  $T \approx 737.5$  K. ◀



Source: NASA

(a)



Source: NASA

(b)

▲ Figure N.5

### QUADRATIC AND TRIGONOMETRIC FUNCTIONS AS MODELS

Although linear models are simple, they are not always appropriate. It may be, for example, that a quadratic model  $y = ax^2 + bx + c$  is suggested by the graphical form of the data or by some known law. Most calculators, CAS programs, and spreadsheets can compute a **quadratic regression curve** for a set of  $(x, y)$  data pairs that minimizes the sum of the squares of the residuals.

► **Example 3** To study the equations of motion of a falling body, a student in a physics laboratory collects the data in Table N.2 showing the height of a body at various times over a 0.15-s time interval. If air resistance is neglected and if the acceleration due to gravity is assumed constant, then known principles of physics state that the height  $h$  should be a quadratic function of time  $t$ . This is consistent with Figure N.6a, in which the plotted data suggest the shape of an inverted parabola.

- Determine the quadratic regression curve for the data in Table N.2.
- According to the model obtained in part (a), when will the object strike the ground?

Table N.2

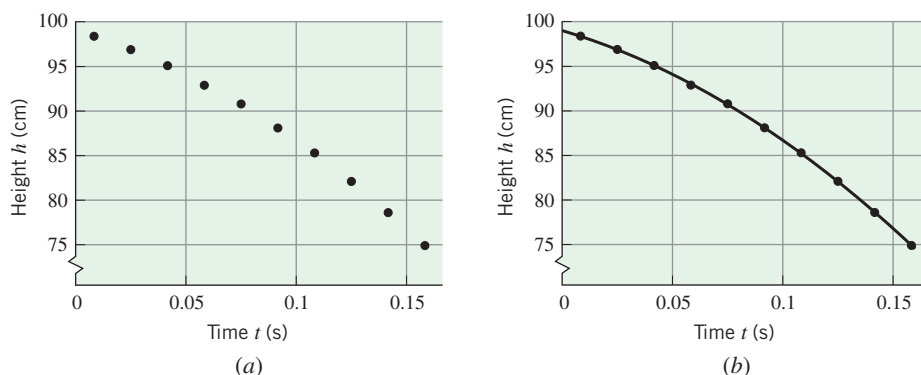
TIME $t$ (s)	HEIGHT $h$ (cm)
0.008333	98.4
0.025	96.9
0.04167	95.1
0.05833	92.9
0.075	90.8
0.09167	88.1
0.10833	85.3
0.125	82.1
0.14167	78.6
0.15833	74.9

**Solution (a).** Using the quadratic regression routine on a calculator, we find that the quadratic curve that best fits the data in Table N.2 has equation

$$h = 99.02 - 73.21t - 499.13t^2$$

Figure N.6b shows the data points and the graph of this quadratic function on the same set of axes. It appears that we have excellent agreement between our curve and the data.

**Solution (b).** Solving the equation  $0 = h = 99.02 - 73.21t - 499.13t^2$ , we find that the object will strike the ground at  $t \approx 0.38$  s. ◀



▲ Figure N.6

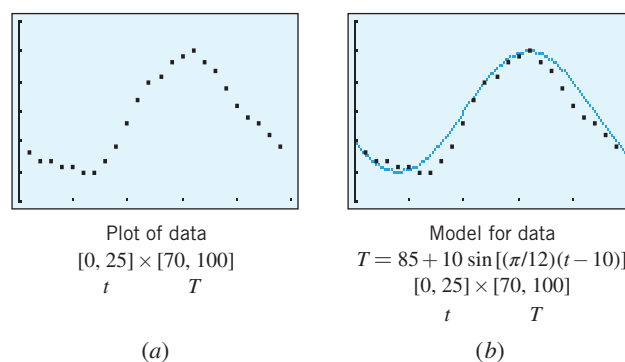
The trigonometric functions  $y = A \sin(Bx - C)$  and  $y = A \cos(Bx - C)$  are particularly useful for modeling periodic phenomena.

► **Example 4** Figure N.7a shows a table and a plot of temperature data recorded over a 24-hour period in the city of Philadelphia.\* Find a function that models the data, and graph your function and data together.

PHILADELPHIA TEMPERATURES FROM 1:00 A.M. TO 12:00 MIDNIGHT ON 27 AUGUST 1993 ( $t$ = HOURS AFTER MIDNIGHT AND $T$ = DEGREES FAHRENHEIT)				
	A.M.		P.M.	
	$t$	$T$	$t$	$T$
1:00	1	78°	13	91°
2:00	2	77°	14	93°
3:00	3	77°	15	94°
4:00	4	76°	16	95°
5:00	5	76°	17	93°
6:00	6	75°	18	92°
7:00	7	75°	19	89°
8:00	8	77°	20	86°
9:00	9	79°	21	84°
10:00	10	83°	22	83°
11:00	11	87°	23	81°
12:00	12	90°	24	79°

**Source:** Philadelphia Inquirer, 28 August 1993.

▲ Figure N.7



\*This example is based on the article “Everybody Talks About It!—Weather Investigations,” by Gloria S. Dion and Iris Brann Fetta, *The Mathematics Teacher*, Vol. 89, No. 2, February 1996, pp. 160–165.

**Solution.** The pattern of the data suggests that the relationship between the temperature  $T$  and the time  $t$  can be modeled by a sinusoidal function that has been translated both horizontally and vertically, so we will look for an equation of the form

$$T = D + A \sin[Bt - C] = D + A \sin\left[B\left(t - \frac{C}{B}\right)\right] \quad (4)$$

Since the highest temperature is  $95^\circ\text{F}$  and the lowest temperature is  $75^\circ\text{F}$ , we take  $2A = 20$  or  $A = 10$ . The midpoint between the high and low is  $85^\circ\text{F}$ , so we have a vertical shift of  $D = 85$ . The period seems to be about 24, so  $2\pi/B = 24$  or  $B = \pi/12$ . The horizontal shift appears to be about 10 (verify), so  $C/B = 10$ . Substituting these values in (4) yields the equation

$$T = 85 + 10 \sin\left[\frac{\pi}{12}(t - 10)\right] \quad (5)$$

whose graph is shown in Figure N.7b. ◀

### TECHNOLOGY MASTERY

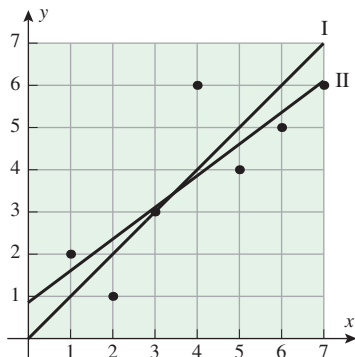
Note that Equation (5) in this example was *not* obtained by minimizing the sum of the squares of the residuals; rather, we used the calculator's graphing capability to see that the proposed model gave a *reasonable* fit to the data. Using a computer program, we can show that the curve of the form  $T = D + A \sin(Bt - C)$  that minimizes the sum of the squares of the residuals for the data in Figure N.7 is

$$y = 84.2037 + 9.5964 \sin(0.2849t - 2.9300)$$

Use your technology utility to compare the graph of this best-fit curve to the graph of (5).

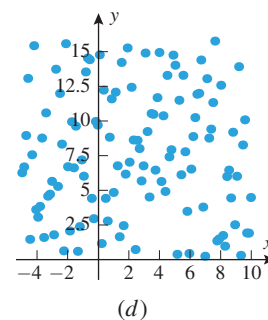
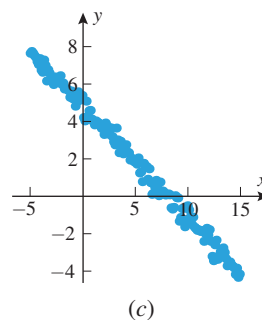
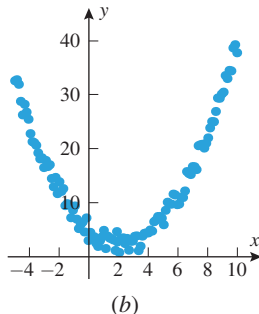
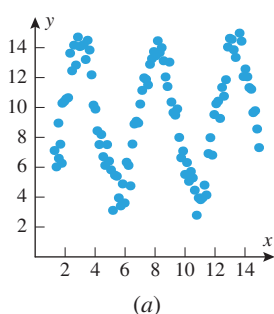
## EXERCISE SET N Graphing Utility

1. One of the lines in the accompanying figure is the regression line. Which one is it?



◀ Figure Ex-1

2. In each part, determine whether a model (linear, quadratic, or trigonometric) reasonably describes the plot of data in the accompanying figure.



▲ Figure Ex-2

3. Table B.1 provides data for the top qualifying speeds at the Indianapolis 500 from 1994 to 2011. Find the least squares line for these data. What is the correlation coefficient? Sketch the least squares line on a plot of the data points.
4. A 25-liter container holds 150 g of  $\text{O}_2$ . The pressure  $p$  of the gas is measured at various temperatures  $T$  (see the accompanying table).
- Determine the least squares line for the data given in the table.
  - Use the model obtained in part (a) to estimate the pressure of the gas at a temperature of  $-50^\circ\text{C}$ .

TEMPERATURE $T$ ( $^{\circ}\text{C}$ )	PRESSURE $p$ (atm)
0	4.18
50	4.96
100	5.74
150	6.49
200	7.26

▲ Table Ex-4

5. A 20-liter container holds 100 g of  $\text{N}_2$ . The pressure  $p$  of this gas is measured at various temperatures  $T$  (see the accompanying table).
- Find the least squares line for this collection of data points. If your calculating utility can produce the correlation coefficient, then find it.
  - Use the model obtained in part (a) to predict the pressure of the gas at a temperature of  $-50^{\circ}\text{C}$ .
  - Use the model obtained in part (a) to predict a temperature at which the pressure of the gas will be zero.

TEMPERATURE $T$ ( $^{\circ}\text{C}$ )	PRESSURE $p$ (atm)
0	3.99
25	4.34
50	4.70
75	5.08
100	5.45

▲ Table Ex-5

6. A 40-liter container holds 20 g of  $\text{H}_2$ . The pressure  $p$  of this gas is measured at various temperatures  $T$  (see the accompanying table).
- Find the least squares line for this collection of data points. If your calculating utility can produce the correlation coefficient, then find it.
  - Use the model obtained in part (a) to predict a temperature at which the pressure of the gas will be zero.
  - At approximately what temperature of the gas will a  $10^{\circ}\text{C}$  increase in temperature result in a 5% increase in pressure?

TEMPERATURE $T$ ( $^{\circ}\text{C}$ )	PRESSURE $p$ (atm)
0	5.55
30	6.13
60	6.75
90	7.35
120	7.98

▲ Table Ex-6

7. The **resistivity** of a metal is a measure of the extent to which a wire made from the metal will resist the flow of electrical current. (The actual *resistance* of the wire will depend on both the resistivity of the metal and the dimensions of the wire.) A common unit for resistivity is the ohm-meter ( $\Omega \cdot \text{m}$ ). Experiments show that lowering the temperature of

a metal also lowers its resistivity. The accompanying table gives the resistivity of copper at various temperatures.

- Find the least squares line for this collection of data points.
- Using the model obtained in part (a), at what temperature will copper have a resistivity of zero?

TEMPERATURE ( $^{\circ}\text{C}$ )	RESISTIVITY ( $10^{-8} \Omega \cdot \text{m}$ )
-100	0.82
-50	1.19
0	1.54
50	1.91
100	2.27
150	2.63

▲ Table Ex-7

8. The accompanying table gives the resistivity of tungsten at various temperatures.
- Find the least squares line for this collection of data points.
  - Using the model obtained in part (a), at what temperature will tungsten have a resistivity of zero?

TEMPERATURE ( $^{\circ}\text{C}$ )	RESISTIVITY ( $10^{-8} \Omega \cdot \text{m}$ )
-100	2.43
-50	3.61
0	4.78
50	5.96
100	7.16
150	8.32

▲ Table Ex-8

9. The accompanying table gives the number of inches that a spring is stretched by various attached weights.
- Use linear regression to express the amount of stretch of the spring as a function of the weight attached.
  - Use the model obtained in part (a) to determine the weight required to stretch the spring 8 in.

WEIGHT (lb)	STRETCH (in)
0	0
2	0.99
4	2.01
6	2.99
8	4.00
10	5.03
12	6.01

◀ Table Ex-9

10. The accompanying table gives the number of inches that a spring is stretched by various attached weights.
- Use linear regression to express the amount of stretch of the spring as a function of the weight attached.
  - Suppose that the spring has been stretched a certain amount by a weight and that adding another 5 lb to the



## N8 Appendix N: Mathematical Models

weight doubles the stretch of the spring. Use the model obtained in part (a) to determine the original amount that the spring was stretched.

11. The accompanying table provides the heights and rebounds per minute for players on the 2002–2003 Davidson College women’s basketball team who played more than 100 minutes during the season.

- Find the least squares line for these data. If your calculating utility can produce the correlation coefficient, then find it.
- Sketch the least squares line on a plot of the data points.
- Is the least squares line a good model for these data? Explain.

WEIGHT (lb)	STRETCH (in)
0	0
1	0.73
2	1.50
3	2.24
4	3.02
5	3.77

▲ Table Ex-10

HEIGHT	REBOUNDS PER MINUTE
6'0"	0.132
6'0"	0.252
5'6"	0.126
5'10"	0.139
6'2"	0.227
6'3"	0.299
6'0"	0.170
5'5"	0.071
6'2"	0.222

▲ Table Ex-11

12. The accompanying table provides the heights and weights for players on the 2002–2003 Davidson College men’s basketball team.

- Find the least squares line for these data. If your calculating utility can produce the correlation coefficient, then find it.
- Sketch the least squares line on a plot of the data points.
- Use this model to predict the weight of the team’s new 7 ft recruit.

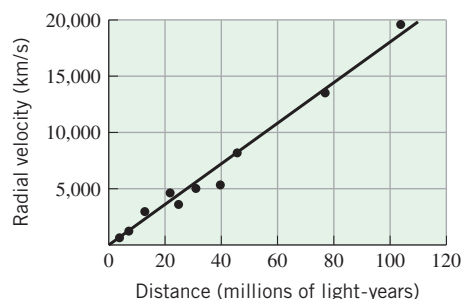
HEIGHT	WEIGHT (lb)	HEIGHT	WEIGHT (lb)
6'2"	185	6'5"	210
6'6"	215	6'4"	185
6'1"	175	6'5"	190
6'1"	180	6'3"	180
6'6"	210	6'8"	235
5'10"	175	6'8"	215
6'9"	210	6'10"	235
6'1"	180		

▲ Table Ex-12

13. (The Age of the Universe) In the early 1900s the astronomer Edwin P. Hubble (1889–1953) noted an unexpected relationship between the radial velocity of a galaxy and its distance  $d$  from any reference point (Earth, for example). That relationship, now known as **Hubble’s law**, states that the galaxies are receding with a velocity  $v$  that is directly proportional to the distance  $d$ . This is usually expressed as  $v = Hd$ , where  $H$  (the constant of proportionality) is called **Hubble’s constant**. When applying this formula it is usual to express  $v$  in kilometers per second

(km/s) and  $d$  in millions of light-years (Mly), in which case  $H$  has units of km/s/Mly. The accompanying figure shows an original plot and trend line of the velocity-distance relationship obtained by Hubble and a collaborator Milton L. Humason (1891–1972).

- Use the trend line in the figure to estimate Hubble’s constant.
- An estimate of the age of the universe can be obtained by assuming that the galaxies move with constant velocity  $v$ , in which case  $v$  and  $d$  are related by  $d = vt$ . Assuming that the Universe began with a “big bang” that initiated its expansion, show that the Universe is roughly  $1.5 \times 10^{10}$  years old. [Use the conversion  $1 \text{ Mly} \approx 9.048 \times 10^{18} \text{ km}$  and take  $H = 20 \text{ km/s/Mly}$ , which is in keeping with current estimates that place  $H$  between 15 and 27 km/s/Mly. (Note that the current estimates are significantly less than that resulting from Hubble’s data.)]
- In a more realistic model of the Universe, the velocity  $v$  would decrease with time. What effect would that have on your estimate in part (b)?



▲ Figure Ex-13

14. The accompanying table gives data for the U.S. population at 10-year intervals from 1790 to 1850. Use quadratic regression to model the U.S. population as a function of time since 1790. What does your model predict as the population of the United States in the year 2000? How accurate is this prediction?

U.S. POPULATION

YEAR $t$	POPULATION $P$ (millions)
1790	3.9
1800	5.3
1810	7.2
1820	9.6
1830	12
1840	17
1850	23

Source: The World Almanac.

◀ Table Ex-14

15. The accompanying table gives the minutes of daylight predicted for Davidson, North Carolina, in 10-day increments during the year 2000. Find a function that models the data in this table, and graph your function on a plot of the data.



DAY	DAYLIGHT (min)	DAY	DAYLIGHT (min)
10	716	190	986
20	727	200	975
30	744	210	961
40	762	220	944
50	783	230	926
60	804	240	905
70	826	250	883
80	848	260	861
90	872	270	839
100	894	280	817
110	915	290	795
120	935	300	774
130	954	310	755
140	969	320	738
150	982	330	723
160	990	340	712
170	993	350	706
180	992	360	706

▲ Table Ex-15

16. The accompanying table gives the fraction of the Moon that is illuminated (as seen from Earth) at midnight (Eastern Standard Time) in 2-day intervals for the first 60 days of 1999. Find a function that models the data in this table, and graph your function on a plot of the data.

DAY	ILLUMINATION	DAY	ILLUMINATION
2	1	32	1
4	0.94	34	0.93
6	0.81	36	0.79
8	0.63	38	0.62
10	0.44	40	0.43
12	0.26	42	0.25
14	0.12	44	0.10
16	0.02	46	0.01
18	0	48	0.01
20	0.07	50	0.11
22	0.22	52	0.29
24	0.43	54	0.51
26	0.66	56	0.73
28	0.85	58	0.90
30	0.97	60	0.99

▲ Table Ex-16

17. The accompanying table provides data about the relationship between distance  $d$  traveled in meters and elapsed time  $t$  in seconds for an object dropped near the Earth's surface. Plot time versus distance and make a guess at a "square-root function" that provides a reasonable model for  $t$  in terms of  $d$ . Use a graphing utility to confirm the reasonableness of your guess.

$d$ (meters)	0	2.5	5	10	15	20	25
$t$ (seconds)	0	0.7	1.0	1.4	1.7	2	2.3

▲ Table Ex-17

18. (a) The accompanying table provides data on five moons of the planet Saturn. In this table  $r$  is the **orbital radius** (the average distance between the moon and Saturn) and  $t$  is the time in days required for the moon to complete one orbit around Saturn. For each data pair calculate  $tr^{-3/2}$ , and use your results to find a reasonable model for  $r$  as a function of  $t$ .
- (b) Use the model obtained in part (a) to estimate the orbital radius of the moon Enceladus, given that its orbit time is  $t \approx 1.370$  days.
- (c) Use the model obtained in part (a) to estimate the orbit time of the moon Tethys, given that its orbital radius is  $r \approx 2.9467 \times 10^5$  km.

MOON	RADIUS (100,000 km)	ORBIT TIME (days)
1980S28	1.3767	0.602
1980S27	1.3935	0.613
1980S26	1.4170	0.629
1980S3	1.5142	0.694
1980S1	1.5147	0.695

▲ Table Ex-18