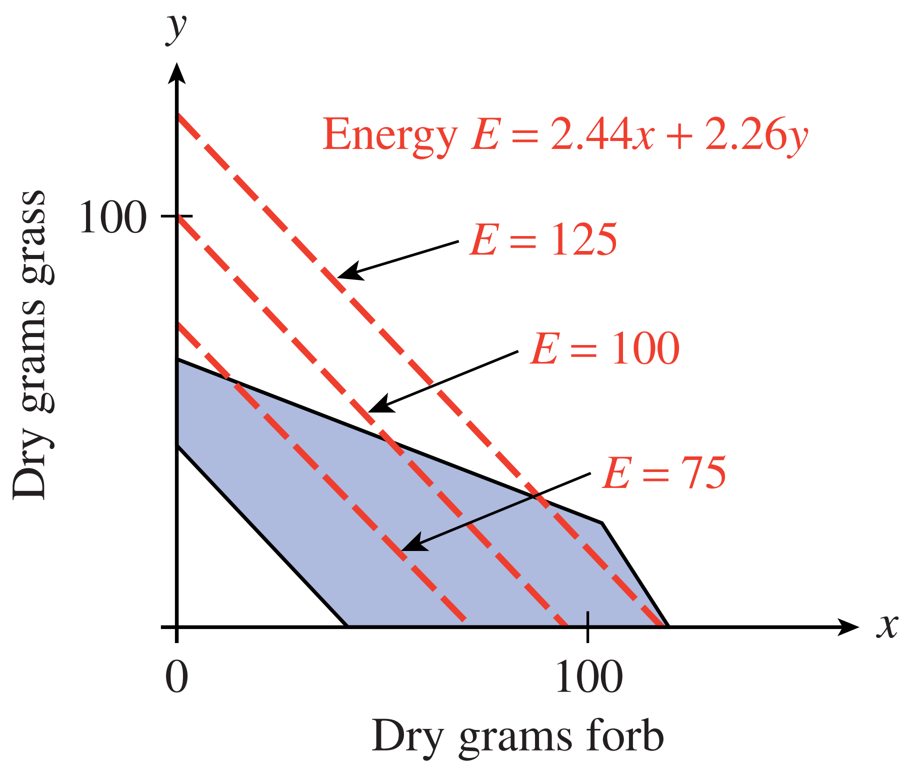
Chapter 8

Models and data

Curve-fitting or regression is an application of linear systems of equations.

In this chapter, we use a graphing technique to solve problems in two unknowns. For example, the diet of the Columbian ground squirrel consists of two foods: grass and forb (a type of flowering weed). Small animals spend most of their time foraging for food, but they must also be alert for predators. Which foraging strategy favors survival: Should the squirrel try to satisfy its dietary requirements in minimum time, thus minimizing its exposure to predators and the elements, or should it try to maximize its intake of nutrients?



Say something

8.1 Linear Regression

We have spent most of this chapter analyzing models described by graphs or equations. To create a model, however, we often start with a quantity of data. Choosing an appropriate function for a model is a complicated process. In this section, we consider only linear models and explore methods for fitting a linear function to a collection of data points. First, we fit a line through two data points.

8.1.1 Fitting a Line through Two Points

If we already know that two variables are related by a linear function, we can find a formula from just two data points. For example, variables that increase or decrease at a constant rate can be described by linear functions.

Example 8.1.1 In 1993, Americans drank 188.6 million cases of wine. Wine consumption increased at a constant rate over the next decade, and we drank 258.3 million cases of wine in 2003. (Source: Los Angeles Times, Adams Beverage Group)

a Find a formula for wine consumption, W, in millions of cases, as a linear function of time, t, in years since 1990.

b State the slope as a rate of change. What does the slope tell us about this problem?

Solution.

a We have two data points of the form (t, W), namely $(3, 188.6)$ and $(13, 258.3)$. We use the point-slope formula to fit a line through these two points. First, we compute the slope.

$$\frac{\Delta W}{\Delta t} = \frac{258.3 - 188.6}{13 - 3} = 6.97$$

Next, we use the slope m = 6.97 and either of the two data points in the point-slope formula.

$$W = W_1 + m(t - t_1)$$
$$W = 188.6 + 6.97(t - 3)$$
$$W = 167.69 + 6.97t$$

Thus, $W = f(t) = 167.69 + 6.97t$.

b The slope gives us the rate of change of the function, and the units of the variables can help us interpret the slope in context.

$$\frac{\Delta W}{\Delta t} = \frac{258.3 - 188.6 \text{ millions of cases}}{13 - 3 \text{ years}} = 6.97 \text{ millions of cases / year}$$

Over the 10 years between 1993 and 2003, wine consumption in the United States increased at a rate of 6.97 million cases per year.

□

To Fit a Line through Two Points:.

1 Compute the slope between the two points.

2 Substitute the slope and either point into the point-slope formula

$$y = y_1 + m(x - x_1)$$

Checkpoint 8.1.2 In 1991, there were 64.6 burglaries per 1000 households in the United States. The number of burglaries reported annually declined at a roughly constant rate over the next decade, and in 2001 there were 28.7 burglaries per 1000 households. (Source: U.S. Department of Justice)

a Find a function for the number of burglaries, B, as a function of time, t, in years, since 1990.

b State the slope as a rate of change. What does the slope tell us about this problem?

Answer.

a $y = 68.19 - 3.59x$

b -3.59 burglaries per 1000 households per year. From 1991 to 2001, the burglary rate declined by 3.59 burglaries per 1000 households every year.

8.1.2 Scatterplots

Empirical data points in a linear relation may not lie exactly on a line. There are many factors that can affect experimental data, including measurement error, the influence of environmental conditions, and the presence of related variable quantities.

Example 8.1.3 A consumer group wants to test the gas mileage of a new model SUV. They test-drive six vehicles under similar conditions and record the distance each drove on various amounts of gasoline.

Gasoline used (gal)	9.6	11.3	8.8	5.2	10.3	6.7
Miles driven	155.8	183.6	139.6	80.4	167.1	99.7

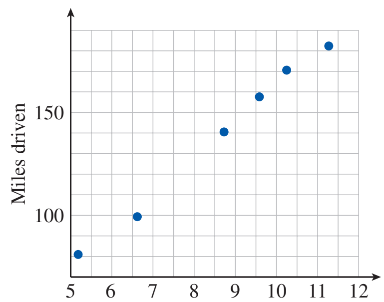
a Are the data linear?

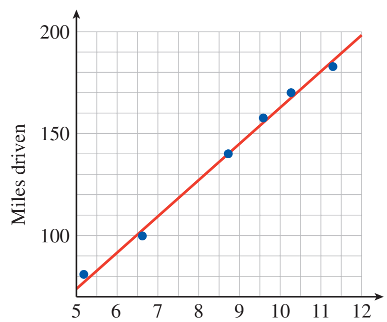
b Draw a line that fits the data.

c What does the slope of the line tell us about the data?

Solution.

a No, the data are not strictly linear. If we compute the slopes between successive data points, the values are not constant. We can see from an accurate plot of the data, shown below, that the points lie close to, but not precisely on, a straight line.





b We would like to draw a line that comes as close as possible to all the data points, even though it may not pass precisely through any of them. In particular, we try to adjust the line so that we have the same number of data points above the line and below the line. One possible solution is shown above.

c To compute the slope of the our estimated line, we first choose two points on the line. Our line appears to pass through one of the data points,$(8.8, 139.6)$. We look for a second point on the line whose coordinates are easy to read, perhaps $(6.5, 100)$. The slope is

$$m = \frac{139.6 - 100}{8.8 - 6.5} = 17.2 \text{ miles per gallon}$$

According to our data, the SUV gets about 17.2 miles to the gallon.

□

Caution 8.1.4 To find the slope of your estimated line, be sure to choose points *on the line*; do not choose any of the data points (unless they happen to lie on your line).

Checkpoint 8.1.5

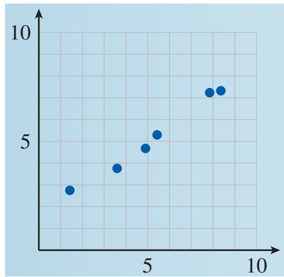
a Plot the data points. Do the points lie on a line?

b Draw a line that fits the data.

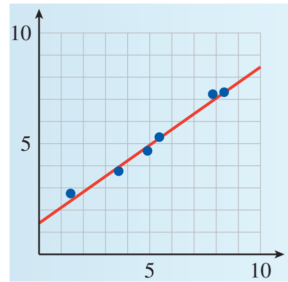
x	1.49	3.68	4.95	5.49	7.88	8.41
y	2.69	3.7	4.6	5.2	7.2	7.3

Answer.

a

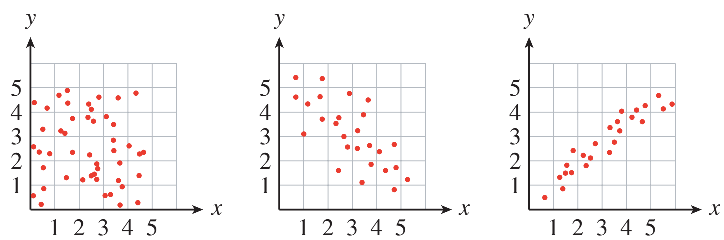


b



The graph in Example 8.1.3, p. 810 is called a **scatterplot**. The points on a scatterplot may or may not show some sort of pattern. Consider the three plots shown below.

- In figure (a), the data points resemble a cloud of gnats; there is no apparent pattern to their locations.
- In figure (b), the data follow a generally decreasing trend, but certainly do not all lie on the same line.
- The points in figure (c) are even more organized; they seem to lie very close to an imaginary line.



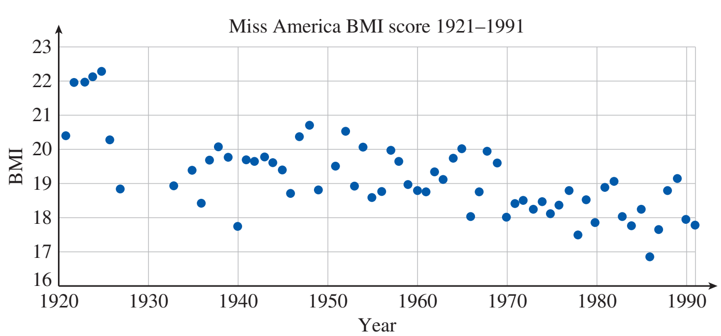
(a) **(b)** **(c)**

If the data in a scatterplot are roughly linear, we can estimate the location of an imaginary **line of best** fit that passes as close as possible to the data points. We can then use this line to make predictions about the data.

8.1.3 Linear Regression

One measure of a person's physical fitness is the *body mass index*, or BMI. Your BMI is the ratio of your weight in kilograms to the square of your height in centimeters. Thus, thinner people have lower BMI scores, and fatter people have higher scores. The Centers for Disease Control considers a BMI between 18.5 and 24.9 to be healthy.

The points on the scatterplot below show the BMI of Miss America from 1921 to 1991. From the data in the scatterplot, can we see a trend in Americans' ideal of female beauty?



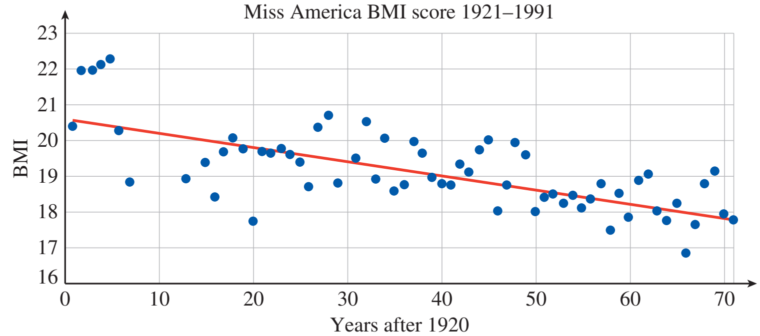
Example 8.1.6

a Estimate a line of best fit for the scatterplot above. (Source: http://www.pbs.org)

b Use your line to estimate the BMI of Miss America 1980.

Solution.

a We draw a line that fits the data points as best we can, as shown below. (Note that we have set $t = 0$ in 1920 on this graph.) We try to end up with roughly equal numbers of data points above and below our line.

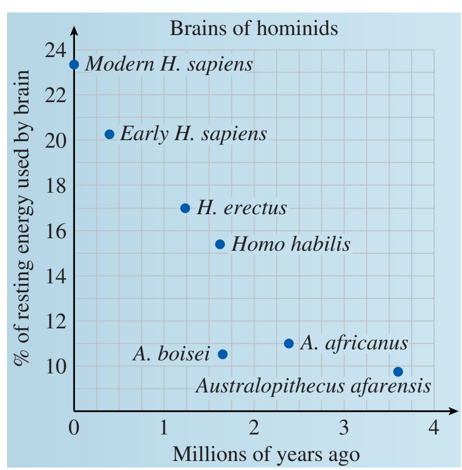


b We see that when $t = 60$ on this line, the y-value is approximately 18.3. We therefore estimate that Miss America 1980 had a BMI of 18.3. (Her actual BMI was 17.85.)

□

Checkpoint 8.1.7 Human brains consume a large amount of energy, about 16 times as much as muscle tissue per unit weight. In fact, brain metabolism accounts for about 25% of an adult human's energy needs, as compared to about 5% for other mammals.

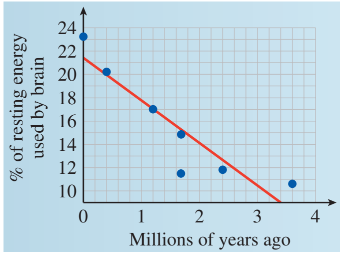
As hominid species evolved, their brains required larger and larger amounts of energy, as shown below. (Source: Scientific American, December 2002)



a Draw a line of best fit through the data points.

b Estimate the amount of energy used by the brain of a hominid species that lived three million years ago.

Answer.

a 

b About 10.5%

The process of predicting an output value based on a straight line that fits the data is called **linear regression**, and the line itself is called the **regression line**. The equation of the regression line is usually used (instead of a graph) to predict values.

Example 8.1.8

a Find the equation of the regression line in Example 8.1.6, p. 812.

b Use the regression equation to predict the BMI of Miss America 1980.

Solution.

a We first calculate the slope by choosing two points on the regression line. The points we choose are not necessarily any of the original data points; instead they should be points on the regression line itself. The line appears to pass through the points $(17, 20)$ and $(67, 18)$. The slope of the line is then

$$m = \frac{18 - 20}{67 - 17} \approx -0.04$$

Now we use the point-slope formula to find the equation of the line. (If you need to review the point-slope formula, see Section 1.5, p. 108.) We substitute $m = -0.04$ and use either of the two points for (x_1, y_1); we will choose $(17, 20)$. The equation of the regression line is

$$\begin{aligned} y &= y_1 + m(x - x_1) \\ y &= 20 - 0.04(x - 17) \quad \text{Simplify.} \\ y &= 20.68 - 0.04t \end{aligned}$$

b We will use the regression equation to make our prediction. For Miss America 1980, $t = 60$ and

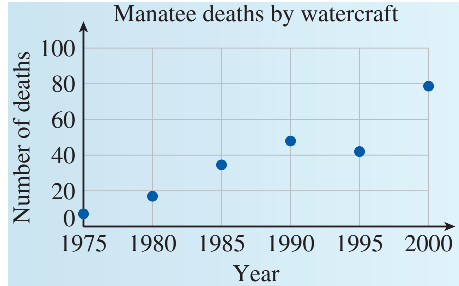
$$y = 20.68 - 0.04(60) = 18.28$$

This value agrees well with the estimate we made in Example 8.1.6, p. 812.

□

Checkpoint 8.1.9 The number of manatees killed by watercraft in Florida waters has been increasing since 1975. Data are given at 5-year intervals in the table. (Source: Florida Fish and Wildlife Conservation Commission)

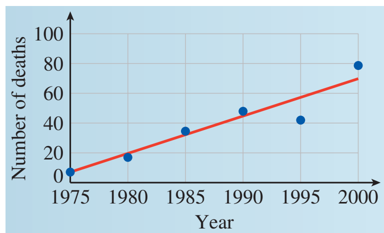
Year	Manatee deaths
1975	6
1980	16
1985	33
1990	47
1995	42
2000	78



a Draw a regression line through the data points shown in the figure.

b Use the regression equation to estimate the number of manatees killed by watercraft in 1998.

Answer.

a 

b $y = 4.7 + 2.6t$

c 65

8.1.4 Linear Interpolation and Extrapolation

Using a regression line to estimate values between known data points is called **interpolation**. Making predictions beyond the range of known data is called **extrapolation**.

Example 8.1.10

a Use linear interpolation to estimate the BMI of Miss America 1960.

b Use linear extrapolation to predict the BMI of Miss America 2001.

Solution.

a For 1960, we substitute $t = 40$ into the regression equation we found in Example 8.1.8, p. 813.

$$y = 20.68 - 0.04(40) = 19.08$$

We estimate that Miss America 1960 had a BMI of 19.08. (Her BMI was actually 18.79.)

b For 2001, we substitute $t = 81$ into the regression equation.

$$y = 20.68 - 0.04(81) = 17.44$$

Our model predicts that Miss America 2001 had a BMI of 17.44. In fact, her BMI was 20.25. By the late 1990s, public concern over the self-image of young women had led to a reversal of the trend toward ever-thinner role models. ☐

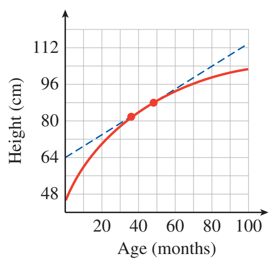
Example 8.1.10, p. 814b illustrates an important fact about extrapolation:

If we try to extrapolate too far, we may get unreasonable results. For example, if we use our model to predict the BMI of Miss America 2520 (when $t = 600$), we get

$$y = 20.68 - 0.04(600) = -3.32$$

Even if the Miss America pageant is still operating in 600 years, the winner cannot have a negative BMI. Our linear model provides a fair approximation for 1920–1990, but if we try to extrapolate too far beyond the known data, the model may no longer apply.

We can also use interpolation and extrapolation to make estimates for nonlinear functions. Sometimes a variable relationship is not linear, but a portion of its graph can be approximated by a line.



The graph at right shows a child's height each month. The graph is not linear because her rate of growth is not constant; her growth slows down as she approaches her adult height. However, over a short time interval the graph is close to a line, and that line can be used to approximate the coordinates of points on the curve.

Checkpoint 8.1.11 Emily was 82 centimeters tall at age 36 months and 88 centimeters tall at age 48 months.

a Find a linear equation that approximates Emily's height in terms of her age over the given time interval.

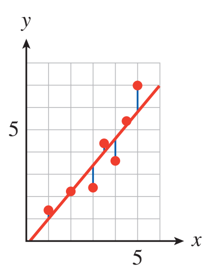
b Use linear interpolation to estimate Emily's height when she was 38 months old, and extrapolate to predict her height at age 50 months.

c Predict Emily's height at age 25 (300 months). Is your answer reasonable?

Answer.

a $y = 64 + 0.5x$ b 83 cm, 89 cm   c 214 cm; No

Estimating a line of best fit is a subjective process. Rather than base their estimates on such a line, statisticians often use the **least squares regression line**.



This regression line minimizes the sum of the squares of all the vertical distances between the data points and the corresponding points on the line, as shown at left. Many calculators are programmed to find the least squares regression line, using an algorithm that depends only on the data, not on the appearance of the graph.

Example 8.1.12

a Find the equation of the least squares regression line for the following data:

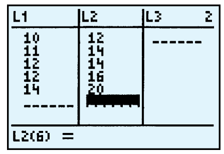
$$(10, 12), (11, 14), (12, 14), (12, 16), (14, 20)$$

b Plot the data points and the least squares regression line on the same axes.

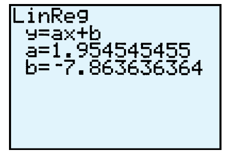
Solution.

a We must first enter the data.

- Press `STATENTER` to select *Edit*.
- If there are data in column L_1 or L_2, clear them out: Use the $\boxed{\uparrow}$ key to select L_1, press `CLEAR`, then do the same for L_2.
- Enter the x-coordinates of the data points in the L_1 column and enter the y-coordinates in the L_2 column, as shown in figure (a) below.



(a)



(b)

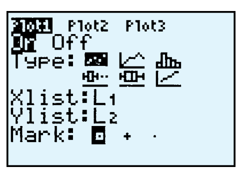
Now we are ready to find the regression equation for our data.

- Press `STAT` $\boxed{\rightarrow}$ 4 to select linear regression, or LinReg (ax + b), then press `ENTER`.
- The calculator will display the equation $y = ax + b$ and the values for a and b, as shown in figure (b).

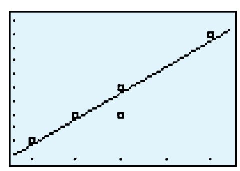
You should find that your regression line is approximately $y = 1.95x - 7.86$.

b First, we first clear out any old definitions in the list.

- Position the cursor after $Y_1 =$ and copy in the regression equation as follows:
- Press `VARS5` $\boxed{\rightarrow}$ $\boxed{\rightarrow}$ `ENTER`.
- To draw a scatterplot, press `2ndY=1` and set the Plot1 menu as shown in figure (a) below.
- Finally, press `ZOOM` 9 to see the scatterplot of the data and the regression line. The graph is shown in figure (b).



(a)


(b)

□

Caution 8.1.13 When you are through with the scatterplot, press `Y=` $\boxed{\uparrow}$ `ENTER` to turn off the *StatPlot*. If you neglect to do this, the calculator will continue to show the scatterplot even after you ask it to plot a new equation.

Checkpoint 8.1.14

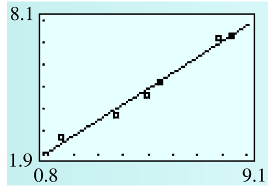
a Use your calculator's statistics features to find the least squares regression equation for the data in Checkpoint 8.1.5, p. 810.

b Plot the data and the graph of the regression equation.

Answer.

a $y = 1.34 + 0.71x$

b



8.1.5 Section Summary

8.1.5.1 Vocabulary

Look up the definitions of new terms in the Glossary.

- Scatterplot
- Extrapolate
- Interpolate
- Least squares regression line
- Regression line
- Linear regression

8.1.5.2 CONCEPTS

1 Data points may not lie exactly on the graph of an equation.

2 Points in a scatterplot may or may not exhibit a pattern.

3 We can approximate a linear pattern by a regression line.

4 We can use interpolation or extrapolation to make estimates and predictions.

5 If we extrapolate too far beyond the known data, we may get unreasonable results.

8.1.5.3 STUDY QUESTIONS

1 What is a regression line?

2 State two formulas you will need to calculate the equation of a line through two points.

3 Explain the difference between interpolation and extrapolation.

4 In general, should you have more confidence in figures obtained by interpolation or by extrapolation? Why?

8.1.5.4 SKILLS

Practice each skill in the Homework 8.1.6, p. 818 problems listed.

1 Find the equation of a line through two points: #1–6, 29–36

2 Draw a line of best fit: #7–18

3 Find the equation of a regression line: #11–28, 37–40

4 Use interpolation and extrapolation to make predictions: #11–40

8.1.6 Linear Regression (Homework 8.1)

In Problems 1–6, we find a linear model from two data points.

a Make a table showing the coordinates of two data points for the model. (Which variable should be plotted on the horizontal axis?)

b Find a linear equation relating the variables.

c State the slope of the line, including units, and explain its meaning in the context of the problem.

1. It cost a bicycle company \$9000 to make 40 touring bikes in its first month of operation and \$15,000 to make 125 bikes during its second month. Express the company's monthly production cost, C, in terms of the number, x, of bikes it makes.

Answer.

a

x	50	125
y	9000	15,000

b $C = 5000 + 80x$

c $m = 80$ dollars/bike, so it costs the company \$80 per bike it manufactures.

2. Flying lessons cost \$645 for an 8-hour course and \$1425 for a 20-hour course. Both prices include a fixed insurance fee. Express the cost, C, of flying lessons in terms of the length, h, of the course in hours.

3. Under ideal conditions, Andrea's Porsche can travel 312 miles on a full tank (12 gallons of gasoline) and 130 miles on 5 gallons. Express the distance, d, Andrea can drive in terms of the amount of gasoline, g, she buys.

Answer.

a

g	12	5
d	312	130

b $d = 26g$

c $m = 26$ miles/gallon, so the Porche's fuel efficiency is 26 miles per gallon.

4. On an international flight, a passenger may check two bags each weighing 70 kilograms, or 154 pounds, and one carry-on bag weighing 50 kilograms, or 110 pounds. Express the weight, p, of a bag in pounds in terms of its weight, k, in kilograms.

5. A radio station in Detroit, Michigan, reports the high and low temperatures in the Detroit/Windsor area as 59°F and 23°F, respectively. A station in Windsor, Ontario, reports the same temperatures as 15°C and −5°C. Express the Fahrenheit temperature, F, in terms of the Celsius temperature, C.

Answer.

a

C	15	-5
F	59	23

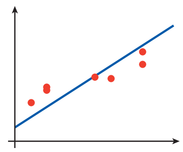
b $F = 32 + \frac{9}{5}C$

c $m = \frac{9}{5}$, so an increase of 1°C is equivalent to an increase of $\frac{9}{5}^\circ$ F.

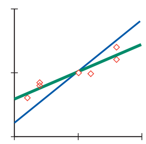
6. Ms. Randolph bought a used car in 2000. In 2002, the car was worth \$9000, and in 2005 it was valued at \$4500. Express the value, V, of Ms. Randolph's car in terms of the number of years, t, she has owned it.

Each regression line can be improved by adjusting either m or b. Draw a line that fits the data points more closely.

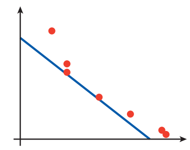
7.



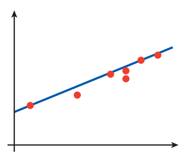
Answer.



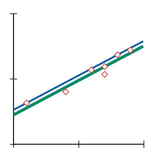
8.



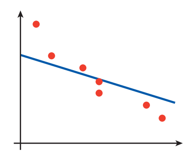
9.



Answer.

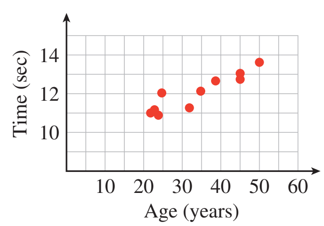


10.



In Problems 11 and 12, use information from the graphs to answer the questions.

11. The scatterplot shows the ages of 10 army drill sergeants and the time it took each to run 100 meters, in seconds.



a What was the hundred-meter time for the 25-year-old drill sergeant?

b How old was the drill sergeant whose hundred-meter time was 12.6 seconds?

c Use a straightedge to draw a line of best fit through the data points.

d Use your line of best fit to predict the hundred-meter time of a 28-year-old drill sergeant.

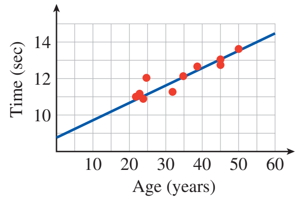
e Choose two points on your regression line and find its equation.

f Use the equation to predict the hundred-meter time of a 40-year-old drill sergeant and a 12 year-old drill sergeant. Are these predictions reasonable?

Answer.

a 12 seconds

b 39

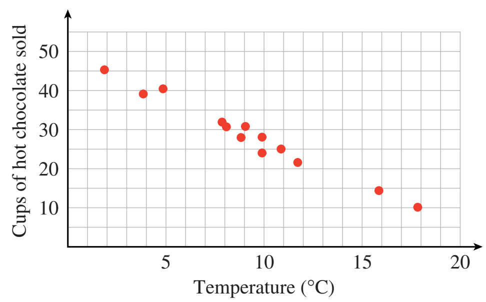
c 

d 11.6 seconds

e $y = 8.5 + 0.1x$

f 12.7 seconds; 10.18 seconds; The prediction for the 40-year-old is reasonable, but not the prediction for the 12-year-old.

12. The scatterplot shows the outside temperature and the number of cups of cocoa sold at an outdoor skating rink snack bar on 13 consecutive nights.



a How many cups of cocoa were sold when the temperature was 2°C?

b What was the temperature on the night when 25 cups of cocoa were sold?

c Use a straightedge to draw a line of best fit through the data points

d Use your line of best fit to predict the number of cups of cocoa that will be sold at the snack bar if the temperature is 7°C.

e Choose two points on your regression line and find its equation.

f Use the equation to predict the number of cups of cocoa that will be sold when the temperature is 10°C and when the temperature is 24°C. Are these predictions reasonable?

13. With Americans' increased use of faxes, pagers, and cell phones, new area codes are being created at a steady rate. The table shows the number of area codes in the United States each year. (Source: USA Today, NeuStar, Inc.)

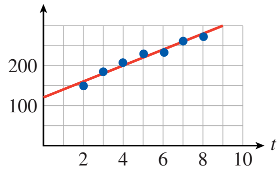
Year	1997	1998	1999	2000	2001	2002	2003
Number of area codes	151	186	204	226	239	262	274

a Let t represent the number of years after 1995 and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c How many area codes do you predict for 2010?

Answer.

a 

b $y = 121 + 19.86t$

c 419

14. The number of mobile homes in the United States has been increasing since 1960. The data in the table are given in millions of mobile homes. (Source: USA Today, U.S. Census Bureau)

Year	1960	1970	1980	1990	2000
Number of mobile homes	0.8	2.1	4.7	7.4	8.8

a Let t represent the number of years after 1960 and plot the data. Draw a line of best fit for the data points

b Find an equation for your regression line.

c How many mobile homes do you predict for 2010?

15. Teenage birth rates in the United States declined from 1991 to 2000. The table shows the number of births per 1000 women in selected years. (Source: U.S. National Health Statistics)

Year	1991	1993	1995	1996	1997	1998
Births	62.1	59.6	56.8	54.4	52.3	51.1

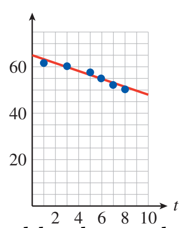
a Let t represent the number of years after 1990 and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c Estimate the teen birth rate in 1994.

d Predict the teen birth rate in 2010.

Answer.

a 

b $y = 64.2 - 1.63t$

c 58 births per 1000 women

d 32 births per 1000 women

16. The table shows the minimum wage in the United States at five-year intervals. (Source: Economic Policy Institute)

Year	1960	1965	1970	1975	1980	1985	1990	1995	2000
Minimum wage	1.00	1.25	1.60	2.10	3.10	3.35	3.80	4.25	5.15

a Let t represent the number of years after 1960 and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c Estimate the minimum wage in 1972.

d Predict the minimum wage in 2010.

17. Life expectancy in the United States has been rising since the nineteenth century. The table shows the U.S. life expectancy in selected years. (Source: http://www.infoplease.com)

Year	1950	1960	1970	1980	1990	2000
Life expectancy at birth	68.2	69.7	70.8	73.7	75.4	77

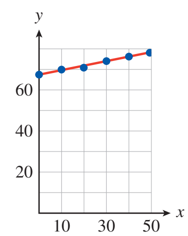
a Let t represent the number of years after 1950, and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c Estimate the life expectancy of someone born in 1987.

d Predict the life expectancy of someone born in 2010.

Answer.

a 

b $y = 0.18t + 67.9$

c 74.9 years

d 79 years

18. The table shows the per capita cigarette consumption in the United States at five-year intervals. (Source: http://www.infoplease.com)

Year	1980	1985	1990	1995	2000
Per capita cigarette consumption	3851	3461	2827	2515	2092

a Let t represent the number of years after 1980, and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c Estimate the per capita cigarette consumption in 1998.

d Predict the per capita cigarette consumption in 2010.

19. "The earnings gap between high-school and college graduates continues to widen, the Census Bureau says. On average, college graduates now earn just over $\$51,000$ a year, almost twice as much as high-school graduates. And those with no high-school diploma have actually seen their earnings drop in recent years." The table shows the unemployment rate and the median weekly earnings for employees with different levels of education. (Source: Morning Edition, National Public Radio, March 28, 2005)

	Years of education	Unemployment rate	Weekly earnings ($)
Some high school no diploma	10	8.8	396
High-school graduate	12	5.5	554
Some college no degree	13	5.2	622
Associate's degree	14	4.0	672
Bachelor's degree	16	3.3	900
Master's degree	18	2.9	1064
Professional degree	20	1.7	1307

a Plot years of education on the horizontal axis and weekly earnings on the vertical axis.

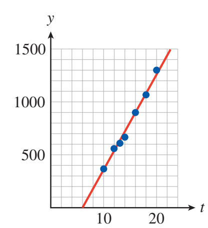
b Find an equation for the regression line.

c State the slope of the regression line, including units, and explain what it means in the context of the data.

d Do you think this model is useful for extrapolation or interpolation? For example, what weekly earnings does the model predict for someone with 15 years of education? For 25 years? Do you think these predictions are valid? Why or why not?

Answer.

a



b $y = 90.49t - 543.7$

c 90.49 dollars/year: Each additional year of education corresponds to an additional $90.49 in weekly earnings.

d No: The degree or diploma attained is more significant than the number of years. So, for example, interpolation for the years of education between a bachelor's and master's degree may be inaccurate because earnings with just the bachelor's degree will not change until the master's degree is attained. And the years after the professional degree will not add significantly to earnings, so extrapolation is inappropriate.

20. The table shows the birth rate (in births per woman) and the female literacy rate (as a percent of the adult female population) in a number of nations. (Source: UNESCO, The World Fact Book, EarthTrends)

Country	Literacy rate	Birth rate
Brazil	88.6	1.93
Egypt	43.6	2.88
Germany	99	1.39
Iraq	53	4.28
Japan	99	1.39
Niger	9.4	6.75
Pakistan	35.2	4.14
Peru	82.1	2.56
Philippines	92.7	3.16
Portugal	91	1.47
Russian Federation	99.2	1.27
Saudi Arabia	69.3	4.05
United States	97	2.08

a Plot the data with literacy rate on the horizontal axis. Draw a line of best fit for the data points.

b Find an equation for the regression line.

c What values for the input variable make sense for the model? What are the largest and smallest values predicted by the model for the output variable?

d State the slope of the regression line, including units, and explain what it means in the context of the data.

21. The table shows the amount of carbon released into the atmosphere annually from burning fossil fuels, in billions of tons, at 5-year intervals from 1950 to 1995. (Source: www.worldwatch.org)

Year	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Carbon emissions	1.6	2.0	2.5	3.1	4.0	4.5	5.2	5.3	5.9	6.2

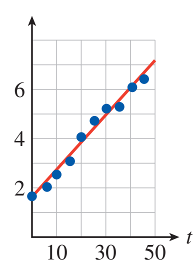
a Let t represent the number of years after 1950 and plot the data. Draw a line of best fit for the data points.

b Find an equation for your regression line.

c Estimate the amount of carbon released in 1992.

Answer.

a



b $y = 1.6 + 0.11t$

c 6.2 billion tons

22. High-frequency radiation is harmful to living things because it can cause changes in their genetic material. The data below, collected by C. P. Oliver in 1930, show the frequency of genetic transmutations induced in fruit flies by doses of X-rays, measured in roentgens. (Source: C. P. Oliver, 1930)

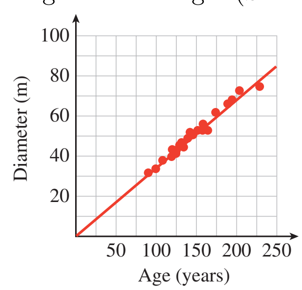
Dosage (roentgens)	285	570	1640	3280	6560
Percentage of mutated genes	1.18	2.99	4.56	9.63	15.85

a Plot the data and draw a line of best fit through the data points.

b Find an equation for your regression line.

c Use the regression equation to predict the percent of mutations that might result from exposure to 5000 roentgens of radiation.

23. Bracken, a type of fern, is one of the most successful plants in the world, growing on every continent except Antarctica. New plants, which are genetic clones of the original, spring from a network of underground stems, or rhizomes, to form a large circular colony. The graph shows the diameters of various colonies plotted against their age. (Source: Chapman et al.,1992)



a Calculate the rate of growth of the diameter of a bracken colony, in meters per year.

b Find an equation for the line of best fit. (What should the vertical intercept of the line be?)

c In Finland, bracken colonies over 450 meters in diameter have been found. How old are these colonies?

Answer.

a 0.34 meters per year

b $y = 0.34x$ ($b = 0$ because the plant has zero size until it begins.)

c Over 1300 years

24. The European sedge warbler can sing several different songs consisting of trills, whistles, and buzzes. Male warblers who sing the largest number of songs are the first to acquire mates in the spring. The data below show the number of different songs sung by several male warblers and the day on which they acquired mates, where day 1 is April 20. (Source: Krebs and Davies, 1993)

Number of songs	41	38	34	32	30	25	24	24	23	14
Pairing day	20	24	25	21	24	27	31	35	40	42

a Plot the data points, with number of songs on the horizontal axis.

A regression line for the data is $y = -0.85x + 53$. Graph this line on the same axes with the data.

b What does the slope of the regression line represent?

c When can a sedge warbler that knows 10 songs expect to find a mate?

d What do the intercepts of the regression line represent? Do these values make sense in context?

25. One of the factors that determines the strength of a muscle is its cross-sectional area. The data below show the cross-sectional area of the arm flexor muscle for several men and women, and their strength, measured by the maximum force they exerted against a resistance. (Source: Davis, Kimmet, Autry, 1986)

Women	Area (sq cm)	11.5	10.8	11.7	12.0	12.5	12.7	14.4	14.4	15.7
	Strength (kg)	11.3	13.2	13.2	14.5	15.6	14.8	15.6	16.1	18.4
Men	Area (sq cm)	13.5	13.8	15.4	15.4	17.7	18.6	20.8	–	–
	Strength (kg)	15.0	17.3	19.0	19.8	20.6	20.8	26.3	–	–

a Plot the data for both men and women on the same graph using different symbols for the data points for men and the data points for women.

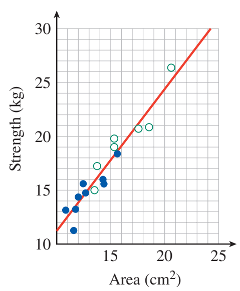
b Are the data for both men and women described reasonably well by the same regression line? Draw a line of best fit through the data.

c Find the equation of your line of best fit, or use a calculator to find the regression line for the data.

d What does the slope mean in this context?

Answer.

a



b Yes

c $y = 1.29x - 1.62$

d The slope, 1.29 kg/sq cm, tells us that strength increases by 1.29 kg when the muscle cross-sectional area increases by 1 sq cm.

26. Astronomers use a numerical scale called **magnitude** to measure the brightness of a star, with brighter stars assigned smaller magnitudes. When we view a star from Earth, dust in the air absorbs some of the light, making the star appear fainter than it really is. Thus, the observed magnitude of a star, m, depends on the distance its light rays must travel through the Earth's atmosphere. The observed magnitude is given by

$$m = m_0 + kx$$

where m_0 is the actual magnitude of the star outside the atmosphere, x is the air mass (a measure of the distance through the atmosphere), and k is a constant called the **extinction coefficient**. To calculate m_0, astronomers observe the same object several times during the night at different positions in the sky, and hence for different values of x. Here are data from such observations. (Source: Karttunen et al., 1987)

Altitude	Air mass, x	Magnitude, m
50°	1.31	0.90
35°	1.74	0.98
25°	2.37	1.07
20°	2.92	1.17

a Plot observed magnitude against air mass, and draw a line of best fit through the data.

b Find the equation of your line of best fit, or use a calculator to find the regression line for the data.

c Find the equation of your line of best fit, or use a calculator to find the regression line for the data.

d What is the value of the extinction coefficient? What is the apparent magnitude of the star outside Earth's atmosphere?

27. Six students are trying to identify an unknown chemical compound by heating the substance and measuring the density of the gas that evaporates. (Density = mass/volume.) The students record the mass lost by the solid substance and the volume of the gas that evaporated from it. They know that the mass lost by the solid must be the same as the mass of the gas that evaporated. (Source: Hunt and Sykes, 1984)

Student	A	B	C	D	E	F
Volume of gas (cm^3)	48	60	24	81	76	54
Loss in mass (mg)	64	81	32	107	88	72

a Plot the data with volume on the horizontal axis. Which student made an error in the experiment?

b Ignoring the incorrect data point, draw a line of best fit through the other points.

c Find an equation of the form $y = kx$ for the data. Why should you expect the regression line to pass through the origin?

d Use your equation to calculate the mass of 1000 cm^3 (one liter) of the gas.

e Here are the densities of some gases at room temperature:

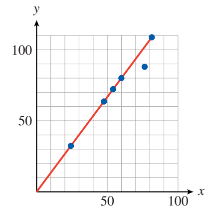
Hydrogen	8	mg/liter
Nitrogen	1160	mg/liter
Oxygen	1330	mg/liter
Carbon dioxide	1830	mg/liter

Which of these might have been the gas that evaporated from the unknown substance?

Hint. Use your answer to part (d) to calculate the density of the gas. 1 cm^3 = 1 milliliter.

Answer.

a E

b 

c $y = 1.33x$; There should be no loss in mass when no gas evaporates.

d 1333 mg

e Oxygen

28. The formulas for many chemical compounds involve ratios of small integers. For example, the formula for water, H_20, means that two atoms of hydrogen combine with one atom of oxygen to make one water molecule. Similarly, magnesium and oxygen combine to produce magnesium oxide. In this problem, we will discover the chemical formula for magnesium oxide. (Source: Hunt and Sykes, 1984)

a Twenty-four grams of magnesium contain the same number of atoms as sixteen grams of oxygen. Complete the table showing the amount of oxygen needed if the formula for magnesium oxide is MgO, Mg_2O, or MgO_2.

Grams of Mg	Grams of O (if MgO)	Grams of O (if Mg_2O)	Grams of O (if MgO_2)
24	16		
48			
12			
6			

b Graph three lines on the same axes to represent the three possibilities, with grams of magnesium on the horizontal axis and grams of oxygen on the vertical axis.

c Here are the results of some experiments synthesizing magnesium oxide.

Experiment	Grams of Magnesium	Grams of oxygen
1	15	10
2	22	14
3	30	20
4	28	18
5	10	6

Plot the data on your graph from part (b). Which is the correct formula for magnesium oxide?

For Problems 29–32,

a Use linear interpolation to give approximate answers.

b What is the meaning of the slope in the context of the problem?

29. The temperature in Encino dropped from 81°F at 1 a.m. to 73°F at 5 a.m. Estimate the temperature at 4 a.m.

Answer.

a 75°F

b The slope of −2 degrees/hour says that temperatures are dropping at a rate of 2° per hour.

30. Newborn blue whales are about 24 feet long and weigh 3 tons. The young whale nurses for 7 months, at which time it is 53 feet long. Estimate the length of a 1-year-old blue whale.

31. A car starts from a standstill and accelerates to a speed of 60 miles per hour in 6 seconds. Estimate the car's speed 2 seconds after it began to accelerate.

Answer.

a 20 mph

b The slope of 10 mph/second says the car accelerates at a rate of 10 mph per second.

32. A truck on a slippery road is moving at 24 feet per second when the driver steps on the brakes. The truck needs 3 seconds to come to a stop. Estimate the truck's speed 2 seconds after the brakes were applied.

In Problems 33–36, use linear interpolation or extrapolation to answer the questions.

33. The temperature of an automobile engine is 9° Celsius when the engine is started and 51°C seven minutes later. Use a linear model to predict the engine temperature for both 2 minutes and 2 hours after it started. Are your predictions reasonable?

Answer. 2 min: 21°C; 2 hr: 729°C; The estimate at 2 minutes is reasonable; the estimate at 2 hours is not reasonable.

34. The temperature in Death Valley is 95° Fahrenheit at 5 a.m. and rises to 110° Fahrenheit by noon. Use a linear model to predict the temperature at 2 p.m. and at midnight. Are your predictions reasonable?

35. Ben weighed 8 pounds at birth and 20 pounds at age 1 year. How much will he weigh at age 10 if his weight increases at a constant rate?

Answer. 128 lb.

36. The elephant at the City Zoo becomes ill and loses weight. She weighed $10,012$ pounds when healthy and only 9641 pounds a week later. Predict her weight after 10 days of illness.

37. Birds' nests are always in danger from predators. If there are other nests close by, the chances of predators finding the nest increase. The table shows the probability of a nest being found by predators and the distance to the nearest neighboring nest. (Source: Perrins, 1979)

Distance to nearest neighbor (meters)	20	40	60	80	100
Probability of predators (%)	47	34	32	17	1.5

a Plot the data and the least squares regression line.

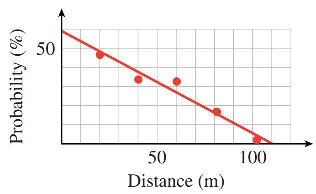
b Use the regression line to estimate the probability of predators finding a nest if its nearest neighbor is 50 meters away.

c If the probability of predators finding a nest is 10%, how far away is its nearest neighbor?

d What is the probability of predators finding a nest if its nearest neighbor is 120 meters away? Is your answer reasonable?

Answer.

a $y \approx -0.54x + 58.7$



b 31.7%

c 90 meters

d The regression line gives a negative probability, which is not reasonable.

38. A trained cyclist pedals faster as he increases his cycling speed, even with a multiple-gear bicycle. The table shows the pedal frequency, p (in revolutions per minute), and the cycling speed, c (in kilometers per hour), of one cyclist. (Source: Pugh, 1974)

Speed (km/hr)	8.8	12.5	16.2	24.4	31.9	35.0
Pedal frequency (rpm)	44.5	50.7	60.6	77.9	81.9	95.3

a Plot the data and the least squares regression line.

b Estimate the cyclist's pedal frequency at a speed of 20 kilometers per hour.

c Estimate the cyclist's speed when he is pedaling at 70 revolutions per minute.

d Does your regression line give a reasonable prediction for the pedaling frequency when the cyclist is not moving? Explain.

39. In this problem we will calculate the efficiency of swimming as a means of locomotion. A swimmer generates power to maintain a constant speed in the water. If she must swim against an opposing force, the power increases. The following table shows the power expended by a swimmer while working against different amounts of force. (A positive force opposes the swimmer, and a negative force helps her.) (Source: diPrampero et al., 1974, and Alexander, 1992)

Force (newtons)	-3.5	0	0	6	8	10	17	17
Metabolic power (watts)	100	190	230	320	380	450	560	600

a Plot the data on the grid, or use the **StatPlot** feature on your calculator. Use your calculator to find the least squares regression line. Graph the regression line on top of the data.

b Use your regression line to estimate the power needed for the swimmer to overcome an opposing force of 15 newtons.

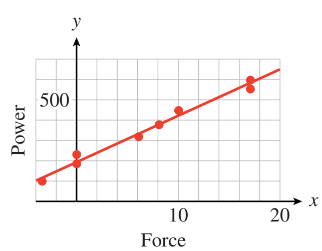
c Use your regression line to estimate the power generated by the swimmer when there is no force either hindering or helping her.

d Estimate the force needed to tow the swimmer at 0.4 meters per second while she rests. (If she is resting, she is not generating any power).

e The swimmer's **mechanical** power (or rate of work) is computed by multiplying her speed times the force needed to tow her at rest. Use your answer to part (d) to calculate the mechanical power she generates by swimming at 0.4 meters per second.

f The ratio of mechanical power to metabolic power is a measure of the swimmer's efficiency. Compute the efficiency of the swimmer when there is no external force opposing or helping her.

Answer.

a 

$y \approx 22.8x + 198.5$

b ≈ 540 watts

c 198.5 watts

d ≈ -8.7 newtons

e 3.5 watts

f about 0.018 or 1.8%

40. In this problem, we calculate the amount of energy generated by a cyclist. An athlete uses oxygen slowly when resting but more quickly during physical exertion. In an experiment, several trained cyclists took turns pedaling on a bicycle ergometer, which measures their work rate. The table shows the work rate of the cyclists, in watts, measured against their oxygen intake, in liters per minute. (Source: Pugh, 1974)

Oxygen consumption (liters/min)	1	1.7	2	3.3	3.9	3.6	4.3	5
Work rate (watts)	40	100	180	220	280	300	320	410

a Plot the data on the grid, or use the **StatPlot** feature on your calculator. Use your calculator to find the least squares regression line. Graph the regression line on top of the data.

b Find the horizontal intercept of the regression line. What does the horizontal intercept tell you about this situation?

c Estimate the power produced by a cyclist consuming oxygen at 5.9 liters per minute.

d What is the slope of the regression line? The slope represents the amount of power, in watts, generated by a cyclist for each liter of oxygen consumed per minute. How many watts of power does a cyclist generate from each liter of oxygen?

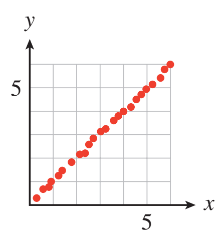
e One watt of power represents an energy output of one joule per second. How many joules of energy does the cyclist generate in one minute?

f How many joules of energy can be extracted from each cubic centimeter of oxygen used? (One liter is equal to 1000 cubic centimeters.)

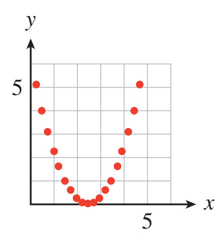
8.2 Curve Fitting

8.2.1 Introduction

In Section 8.1, p. 808, we used linear regression to fit a line to a collection of data points. If the data points do not cluster around a line, it does not make sense to describe them by a linear function. Compare the scatterplots shown below.



(a)



(b)

The points in figure (a) are roughly linear in appearance, but the points in figure (b) are not. However, we can visualize a parabola that would approximate the data we will. In this section, we will see how to fit a quadratic function to a collection of data points.

We will need to solve a special type of 3×3 linear system, that is, a linear system of three equations in three variables. We can solve these systems using the **elimination** method. See Appendix A.5, p. 887 to review the elimination method.

Example 8.2.1 Use elimination to solve the system of equations.

$$\begin{aligned} 3a + 2b + c &= -1 && (1)\\ a - 2b + c &= -3 && (2)\\ 2a + 3b + c &= 4 && (3) \end{aligned}$$

Solution. We first eliminate c from the system by combining the equations in pairs. We can add -1 times Equation (2) to Equation (1) to get a new equation in two variables:

$$\begin{array}{rcrcrcr} 3a & + & 2b & + & c & = & -1 \\ -a & + & 2b & - & c & = & 3 \\ \hline 2a & + & 4b & & & = & 2 \end{array} \qquad \begin{array}{r} (1) \\ -1 \text{ times Equation } (2) \\ (4) \end{array}$$

Next, we add -1 times Equation (2) to Equation (3) to get a second equation in two variables:

$$\begin{array}{rcrcrcr} 2a & + & 3b & + & c & = & -4 \\ -a & + & 2b & - & c & = & 3 \\ \hline a & + & 5b & & & = & 7 \end{array} \qquad \begin{array}{r} (3) \\ -1 \times (2) \\ (5) \end{array}$$

By combining Equations (4) and (5), we have a 2×2 linear system, which we can solve as usual.

$$\begin{aligned} 2a + 4b &= 2 && (4)\\ a + 5b &= 7 && (5) \end{aligned}$$

To eliminate a, we add -2 times Equation (5) to Equation (4):

$$\begin{array}{rcrcr} 2a & + & 4b & = & 2 \\ -2a & - & 10b & = & -14 \\ \hline & & -6b & = & -12 \end{array} \qquad \begin{array}{r} (4) \\ -2 \times (5) \\ \end{array}$$

Solving this last equation gives us $b = 2$. Then we substitute $b = 2$ into either of Equations (4) or (5) to find $a = -3$. Finally, we substitute both values into one of the three original equations to find $c = 4$. The solution of the system is $a = -3$, $b = 2$, $c = 4$. □

Checkpoint 8.2.2 Follow the steps to solve the system

$$\begin{aligned} a + b + c &= 3 && (1)\\ 4a - b + c &= -4 && (2)\\ -3a + 2b + c &= 4 && (3) \end{aligned}$$

1 Eliminate c from Equations (1) and (2) to obtain a new Equation (4).

2 Eliminate c from Equations (2) and (3) to obtain a new Equation (5).

3 Solve the system of Equations (4) and (5).

4 Substitute the values of a and b into one of the original equations to find c.

Answer. $a = 1, b = 5, c = -3$

8.2.2 Finding a Quadratic Function through Three Points

Every linear function can be written in the form

$$y = mx + b$$

To find a specific line, we must find values for the two parameters (constants) m and b. We need two data points in order to find those two parameters. A quadratic function, however, has three parameters, a, b, and c:

$$y = ax^2 + bx + c$$

To find these parameters, we need three data points. We then use the method of elimination to solve a system of three linear equations.

Example 8.2.3 Find values for a, b, and c so that the points $(1, 3)$, $(3, 5)$, and $(4, 9)$ lie on the graph of $y = ax^2 + bx + c$.

Solution. We substitute the coordinates of each of the three points into the equation of the parabola to obtain three equations:

$$\begin{aligned} \mathbf{3} &= a(\mathbf{1})^2 + b(\mathbf{1}) + c \\ \mathbf{5} &= a(\mathbf{3})^2 + b(\mathbf{3}) + c \\ \mathbf{9} &= a(\mathbf{4})^2 + b(\mathbf{4}) + c \end{aligned}$$

or, equivalently,

$$\begin{aligned} a + b + c &= 3 && (1) \\ 9a + 3b + c &= 5 && (2) \\ 16a + 4b + c &= 9 && (3) \end{aligned}$$

This is a system of three equations in the three unknowns a, b, and c. To solve the system, we first eliminate c. Add -1 times Equation (1) to Equation (2) to obtain

$$8a + 2b = 2 \qquad (4)$$

and add -1 times Equation (1) to Equation (3) to get

$$15a + 3b = 6 \qquad (5)$$

We now have a system of two linear equations in two variables:

$$\begin{aligned} 8a + 2b &= 2 && (4) \\ 15a + 3b &= 6 && (5) \end{aligned}$$

We eliminate b from Equations (4) and (5): Add -3 times Equation (4) to 2 times Equation (5) to get

$$\begin{array}{rcrcr} -24a & - & 6b & = & -6 \\ 30a & + & 6b & = & 12 \\ \hline 6a & & & = & 6 \end{array} \qquad \begin{array}{r} -3 \times (4) \\ 2 \times (5) \\ \\ \end{array}$$

or $a = 1$. We substitute 1 for a in Equation (4) to find

$$\begin{aligned} 8(\mathbf{1}) + 2b &= 2 \qquad \text{Solve for } \boldsymbol{b}. \\ b &= -3 \end{aligned}$$

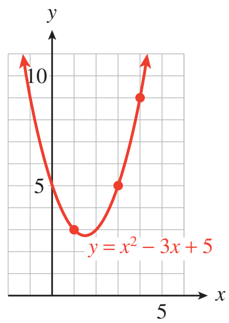
Finally, we substitute -3 for b and 1 for a in Equation (1) to find

$$\begin{aligned} \mathbf{1} + (\mathbf{-3}) + c &= 3 \qquad \text{Solve for } \boldsymbol{c}. \\ c &= 5 \end{aligned}$$

Thus, the equation of the parabola is

$$y = x^2 - 3x + 5$$

The parabola and the three points are shown below.



□

Checkpoint 8.2.4

a Find the equation of a parabola

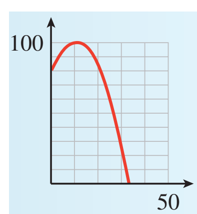
$$y = ax^2 + bx + c$$

that passes through the points $(0, 80)$, $(15, 95)$, and $(25, 55)$.

b Plot the data points and sketch the parabola.

Answer.

a $y = \frac{-1}{5}x^2 + 4x + 80$

b 

The simplest way to fit a parabola to a set of data points is to pick three of the points and find the equation of the parabola that passes through those three points.

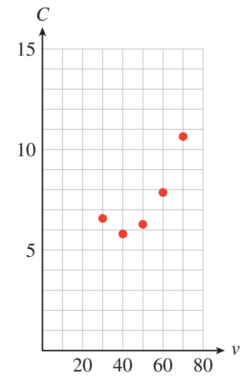
Example 8.2.5 Major Motors Corporation is testing a new car designed for in-town driving. The data below show the cost of driving the car at different speeds. The speeds, v, are given in miles per hour, and the cost, C, includes fuel and maintenance for driving the car 100 miles at that speed.

v	30	40	50	60	70
C	6.50	6.00	6.20	7.80	10.60

Find a possible quadratic model for C as a function of v, $C = av^2 + bv + c$.

Solution.

When we plot the data, it is clear that the relationship between v and C is not linear, but it may be quadratic, as shown at right.
We will use the last three data points, $(50, 6.20)$, $(60, 7.80)$, and $(70, 10.60)$, to fit a parabola to the data. We would like to find the coefficients a, b, and c of a parabola $C = av^2 + bv + c$ that includes the three data points. This gives us a system of equations:



$$\begin{aligned} 2500a + 50b + c &= 6.20 && (1)\\ 3600a + 60b + c &= \ \ 7.8 && (2)\\ 4900a + 70b + c &= 10.6 && (3) \end{aligned}$$

Eliminating c from Equations (1) and (2) yields Equation (4), and eliminating c from Equations (2) and (3) yields Equation (5).

$$\begin{aligned} 1100a + 10b &= 1.60 && (4)\\ 1300a + 10b &= \ \ 2.8 && (5) \end{aligned}$$

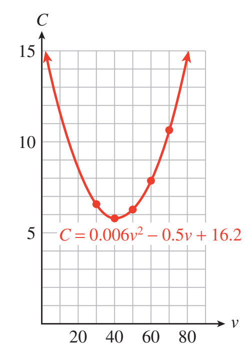
Eliminating b from Equations (4) and (5) gives us

$$\begin{aligned} 200a &= 1.20\\ a &= 0.006 \end{aligned}$$

We substitute this value into Equation (4) to find $b = -0.5$, then substitute both values into Equation (1) to find $c = 16.2$. Thus, our quadratic model is

$$C = 0.006v^2 - 0.5v + 16.2$$

The graph of this function, along with the data points, is shown at right.

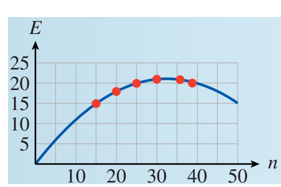


□

Checkpoint 8.2.6 Sara plans to start a side business selling eggs. She finds that the total number of eggs produced each day depends on the number of hens confined in the henhouse, as shown in the table. Use the first three data points to find a quadratic model $E = an^2 + bn + c$. Plot the data and sketch the curve on the same axes.

Number of hens, n	15	20	25	30	36	39
Number of eggs, E	15	18	20	21	21	20

Answer. $E = -0.02n^2 + 1.3n$

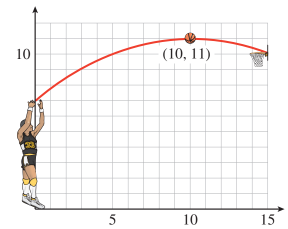


8.2.3 Finding an Equation in Vertex Form

It is easier to find a quadratic model if one of the points we know happens to be the vertex of the parabola. In that case, we need only one other point, and we can use the vertex form to find its equation.

Example 8.2.7

When Andre practices free-throws at the park, the ball leaves his hands at a height of 7 feet and reaches the vertex of its trajectory 10 feet away at a height of 11 feet, as shown at right.



a Find a quadratic function for the ball's trajectory.

b Do you think Andre's free-throw will score on a basketball court where the hoop is 15 feet from the shooter and 10 feet high?

Solution.

a If Andre's feet are at the origin, then the vertex of the ball's trajectory is the point $(10, 11)$, and its y-intercept is $(0, 7)$. Start with the vertex form for a parabola:

$$y = a(x - x_v)^2 + y_v$$
$$y = a(x - 10)^2 + 11$$

We still need to know the value of a. We can substitute the coordinates of any point on the parabola for x and y and solve for a. We will use the point $(0, 7)$:

$$7 = a(0 - 10)^2 + 11$$
$$7 = 100a + 11$$
$$a = -0.04$$

The equation of the trajectory is $y = -0.04(x - 10)^2 + 11$.

b We would like to know if the point $(15, 10)$ is on the trajectory of Andre's free-throw. Substitute $x = 15$ into the equation:

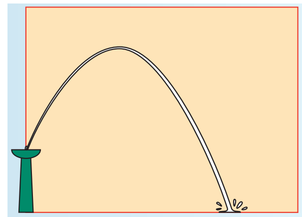
$$y = -0.04(15 - 10)^2 + 11$$
$$= -0.04(25) + 11 = 10$$

Andre's shot will score.

□

Checkpoint 8.2.8 Francine is designing a synchronized fountain display for a hotel in Las Vegas. For each fountain, water emerges in a parabolic arc from a nozzle 3 feet above the ground. Francine would like the vertex of the arc to be 8 feet high and 2 feet horizontally from the nozzle.

a Choose a coordinate system for the diagram below and write a function for the path of the water.



b How far from the base of the nozzle will the stream of water hit the ground?

Answer.

a With the origin on the ground directly below the nozzle, $y = \frac{-5}{4}x^2 + 5x + 3$.

b Approximately 4.53 feet

Example 8.2.9

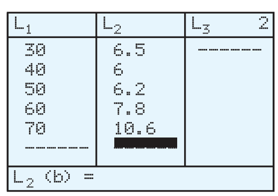
a Use your calculator to find a quadratic fit for the data in Example 8.2.5, p. 835.

b How many of the given data points actually lie on the graph of the quadratic approximation?

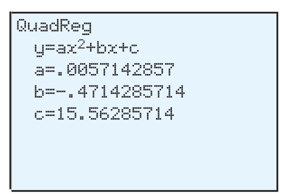
Solution.

a We press `STAT` `ENTER` and enter the data under columns L_1 and L_2, as shown below. Next, we calculate the quadratic regression equation and store it in Y_1 by pressing `STAT` [→] 5 `VARS` [→] 1 1 `ENTER`.

The regression equation has the form $y = ax^2 + bx + c$, where $a = 0.0057$, $b = -0.47$, and $c = 15.56$. Notice that a, b, and c are all close to the values we computed in Example 8.2.5, p. 835.



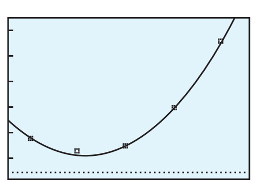
(a)



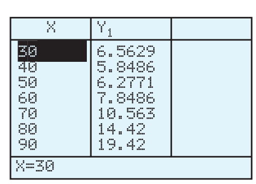
(b)

b Next, we will graph the data and the regression equation. We press `Y=` and select *Plot1*, then press `ZOOM` 9 to see the graph shown below. The parabola seems to pass close to all the data points.

However, try using either the *value* feature or a table to find the y-coordinates of points on the regression curve. By comparing these y-coordinates with our original data points, we find that none of the given data points lies precisely on the parabola.



(a)



(b)

□

Checkpoint 8.2.10 To test the effects of radiation, a researcher irradiated male mice with various dosages and bred them with unexposed female mice. The table below shows the fraction of fertilized eggs that survived, as a function of the radiation dosage. (Source: Strickberger, Monroe W., 1976)

Radiation (rems)	100	300	500	700	900	1100	1500
Relative survival of eggs	0.94	0.700	0.544	0.424	0.366	0.277	0.195

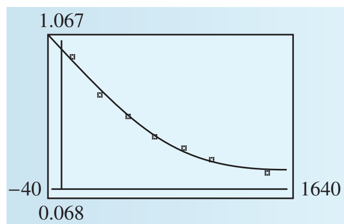
a Enter the data into your calculator and create a scatterplot. Does the graph appear to be linear? Does it appear to be quadratic?

b Fit a quadratic regression equation to the data and graph the equation on the scatterplot.

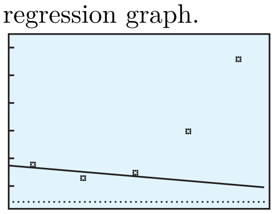
Answer.

a The graph appears to be quadratic.

b $y = 3.65x^2 - 0.001x + 1.02$



Caution 8.2.11 We must be careful that our data set gives a complete picture of the situation we want to model. A regression equation may fit a particular collection of data and still be a poor model if the rest of the data diverge from the regression graph.



In Example 8.2.5, p. 835, suppose Major Motors had collected only the first three data points and fit a line through them, as shown at left. This regression line gives poor predictions for the cost of driving at 60 or 70 miles per hour.

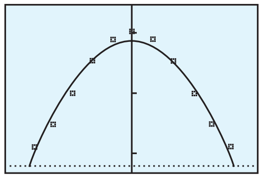
Example 8.2.12 Francine records the height of the tip of the minute hand on the classroom's clock at different times. The data are shown in the table, where time is measured in minutes since noon. (A negative time indicates a number of minutes before noon.) Find a quadratic regression equation for the data and use it to predict the height of the minute hand's tip at 40 minutes past noon. Do you believe this prediction is valid?

Time (minutes)	−25	−20	−15	−10	−5	0	5	10	15	20	25
Height (feet)	7.13	7.50	8.00	8.50	8.87	9.00	8.87	8.50	8.80	7.50	7.13

Solution. We enter the time data under L_1 and the height data under L_2. Then we calculate and store the quadratic regression equation in Y_1, as we did in Example 8.2.9, p. 838. The regression equation is

$$y = -0.00297x^2 + 0x + 8.834$$

From either the graph of the regression equation or from the table (see figure below), we can see that the fit is not perfect, although the curve certainly fits the data better than any straight line could.



(a)

X	Y_1
-25	6.9762
-20	7.645
-15	8.1652
-10	8.5368
-5	8.7597
0	8.834
5	8.7597

X=-25

(b)

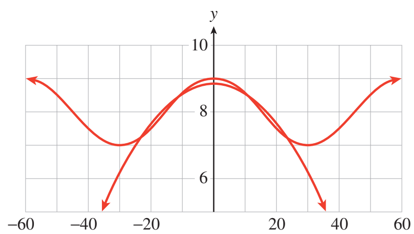
X	Y_1
10	8.5368
15	8.1652
20	7.645
25	6.9762
30	6.1588
35	5.1927
40	4.078

X=40

(c)

If we scroll down the table, we find that this equation predicts a height of approximately 4.08 feet at time 40 minutes. (See figure (c).) This is a preposterous estimate! The position of the minute hand at 40 minutes after noon should be the same as it was exactly one hour earlier (at 20 minutes before noon), when it was 7.50 feet. □

Using the wrong type of function to fit the data is a common error in making predictions. We know that the minute hand of a clock repeats its position every 60 minutes. The graph of the height of its tip oscillates up and down, repeating the same pattern over and over. We cannot describe such a graph using either a linear or a quadratic function.



The graph of the height is shown at left, along with the graph of our quadratic regression equation. You can see that the regression equation fits the actual curve only on a small interval.

Your calculator can always compute a regression equation, but that equation is not necessarily appropriate for your data. Choosing a reasonable type of regression equation for a particular data set requires knowledge of different kinds of models and the physical or natural laws that govern the situation at hand.

Checkpoint 8.2.13 A speeding motorist slams on the brakes when she sees an accident directly ahead of her. The distance she has traveled t seconds after braking is shown in the table.

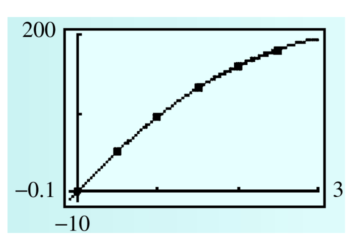
Time (seconds)	0	0.5	1.0	1.5	2.0	2.5
Distance (feet)	0	51	95	131	160	181

a Enter the data into your calculator and create a scatterplot. Fit a quadratic regression equation to the data and graph the equation on the scatterplot.

b Use your regression equation to find the vertex of the parabola. What do the coordinates represent in terms of the problem?

Answer.

a $y = -15x^2 + 110x - 0.07$



b $(3.67, 201)$: The car came to a stop in 3.67 seconds, after sliding 201 feet.

8.2.4 Section Summary

8.2.4.1 Vocabulary

Look up the definitions of new terms in the Glossary.

- Elimination method
- Quadratic regression

8.2.4.2 CONCEPTS

1 We need three points to determine a parabola.

2 We can use the method of elimination to find the equation of a parabola through three points.

3 If we know the vertex of a parabola, we need only one other point to find its equation.

4 We can use quadratic regression to fit a parabola to a collection of data points.

8.2.4.3 STUDY QUESTIONS

1 How many points are necessary to determine a parabola?

2 Why do we need a second point to find the equation of a parabola if we know its vertex?

3 How can you decide whether linear regression, quadratic regression, or neither one is appropriate for a collection of data?

8.2.4.4 SKILLS

Practice each skill in the Homework 8.2.5, p. 841 problems listed.

1 Fit a quadratic equation through three points: #5–12

2 Find a quadratic model in vertex form: #13–30

3 Use quadratic regression to fit a parabola to data: #31–34

8.2.5 Curve-fitting (Homework 8.2)

For Problems 1–4, solve the system by elimination. Begin by eliminating c.

1.
$$\begin{aligned} a + b + c &= -3 \\ a - b + c &= -9 \\ 4a + 2b + c &= -6 \end{aligned}$$

Answer.
$a = -2, b = 3, c = -4$

2.
$$\begin{aligned} a + b + c &= 10 \\ 4a + 2b + c &= 19 \\ 9a + 3b + c &= 38 \end{aligned}$$

3.
$$\begin{aligned} a - b + c &= 12 \\ 4a - 2b + c &= 19 \\ 9a + 3b + c &= 4 \end{aligned}$$

Answer.
$a = 1, b = -4, c = 7$

4.
$$\begin{aligned} 4a + 2b + c &= 14 \\ 9a - 3b + c &= -41 \\ 16a - 4b + c &= -70 \end{aligned}$$

For Problems 5–12, find a quadratic equation that fits the data points.

5. Find values for a, b, and c so that the graph of the parabola $y = ax^2 + bx + c$ includes the points $(-1, 0)$, $(2, 12)$, and $(-2, 8)$.

Answer. $a = 3, b = 1, c = -2$. The equation for the parabola is $y = 3x^2 + x - 2$

6. Find values for a, b, and c so that the graph of the parabola $y = ax^2 + bx + c$ includes the points $(-1, 2)$, $(1, 6)$, and $(2, 11)$.

7. A survey to determine what percent of different age groups regularly use marijuana collected the following data.

Age	15	20	25	30
Percent	4	13	11	7

(a) Use the percentages for ages 15, 20, and 30 to fit a quadratic function to the data, $P = ax^2 + bx + c$, where x represents age.

(b) What does your function predict for the percentage of 25-year-olds who use marijuana?

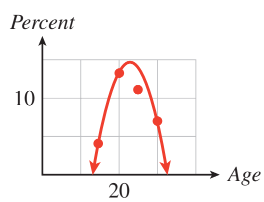
(c) Sketch the graph of your quadratic function and the given data on the same axes.

Answer.

(a) $P = -0.16x^2 + 7.4x - 71$

(b) 14%. It predicts that 14% of the 25-year old population use marijuana on a regular basis.

(c)



8. The following data show the number of people of certain ages who were the victims of homicide in a large city last year.

Age	10	20	30	40
Number of victims	12	62	72	40

(a) Use the first three data points to fit a quadratic function to the data, $N = ax^2 + bx + c$, where x represents age.

(b) What does your function predict for the number of 40-year-olds who were the victims of homicide?

(c) Sketch the graph of your quadratic function and the given data on the same axes.

9. The data below show Americans' annual per capita consumption of chicken for several years since 1985.

Year	1986	1987	1988	1989	1990
Pounds of chicken	51.3	55.5	57.4	60.8	63.6

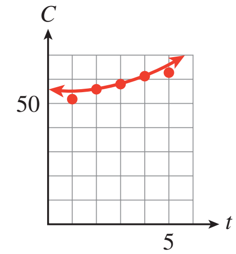
(a) Use the values for 1987 through 1989 to fit a quadratic function to the data, $C = at^2 + bt + c$, where t is measured in years since 1985.

(b) What does your function predict for per capita chicken consumption in 1990?

(c) Sketch the graph of your function and the given data.

Answer.

(a) $C = 0.75t^2 - 1.85t + 56.2$

(c) 

(b) 65.7 lb

10. The data show sales of in-line skates at a sporting goods store at the beach.

Year	1990	1991	1992	1993	1994
Skate sold	54	82	194	446	726

(a) Use the values for 1991 through 1993 to fit a quadratic function to the data, $S = at^2 + bt + c$, where t is measured in years since 1990.

(b) What does your function predict for the number of pairs of skates sold in 1994?

(c) Sketch the graph of your function and the given data.

11. Find a quadratic function for the number of diagonals that can be drawn in a polygon of n sides. Some data are provided.

Sides	4	5	6	7
Diagonals	2	5	9	14

Answer. $D = \frac{1}{2}n^2 - \frac{3}{2}n$

12. You are driving at 60 miles per hour when you step on the brakes. Find a quadratic function for the distance in feet that your car travels in t seconds after braking. Some data are provided.

Seconds	1	2	3	4
Feet	81	148	210	267

13.

(a) Write an equation for a parabola whose vertex is the point $(-2, 6)$. (Many answers are possible.)

(b) Find the value of a if the y-intercept of the parabola in part (a) is 18.

Answer.

(a) $y = a(x + 2)^2 + 6$

(b) 3

14.

(a) Write an equation for a parabola whose vertex is the point $(5, -10)$. (Many answers are possible.)

(b) Find the value of a if the y-intercept of the parabola in part (a) is -5.

15.

(a) Write an equation for a parabola with vertex at $(0, -3)$ and one of its x-intercepts at $(2, 0)$.

(b) Write an equation for a parabola with vertex at $(0, -3)$ and no x-intercepts.

Answer.

(a) $y = \frac{3}{4}x^2 - 3$ (b) $y = ax^2 - 3$ for any $a < 0$

16. Write an equation for a parabola with vertex at $(4, 0)$ and y-intercept at $(0, 4)$. How many x-intercepts does the parabola have?

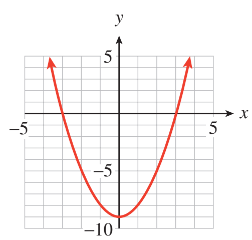
17. Find the equation for a parabola that has a vertex of $(30, 280)$ and passes through the point $(20, 80)$.

Answer. $y = -2(x - 30)^2 + 280$

18. Find the equation for a parabola that has a vertex of $(-12, -40)$ and passes through the point $(6, 68)$.

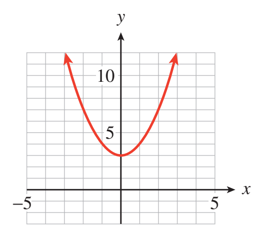
For Problems 19–26, find an equation for each parabola. Use the vertex form or the factored form of the equation, whichever is more appropriate.

19.

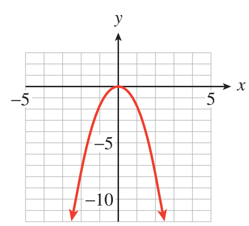


Answer. $y = x^2 - 9$

20.

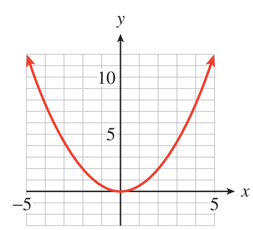


21.

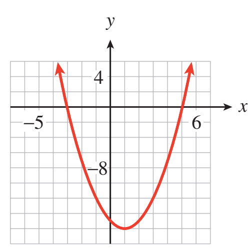


Answer. $y = -2x^2$

22.

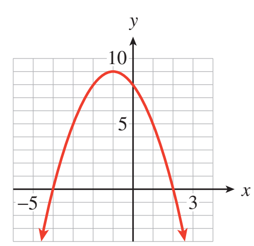


23.

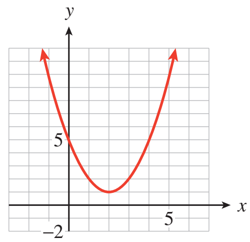


Answer. $y = x^2 - 2x - 15$

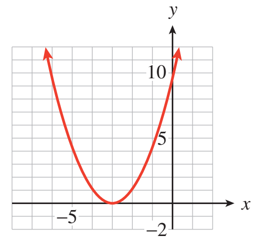
24.



25.



26.



Answer. $y = x^2 - 4x + 5$

27. In skeet shooting, the clay pigeon is launched from a height of 4 feet and reaches a maximum height of 164 feet at a distance of 80 feet from the launch site.

(a) Write a function for the height of the clay pigeon in terms of the horizontal distance it has traveled.

(b) If the shooter misses the clay pigeon, how far from the launch site will it hit the ground?

Answer.

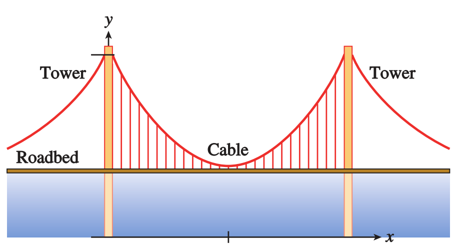
(a) $y = \dfrac{-1}{40}(x - 80)^2 + 164$ (b) 160.99 ft

28. The batter in a softball game hits the ball when it is 4 feet above the ground. The ball reaches the greatest height on its trajectory, 35 feet, directly above the head of the left-fielder, who is 200 feet from home plate.

(a) Write a function for the height of the softball in terms of its horizontal distance from home plate.

(b) Will the ball clear the left field wall, which is 10 feet tall and 375 feet from home plate?

The cables on a suspension bridge hang in the shape of parabolas. For Problems 29–30, imagine a coordinate system superimposed on a diagram of the bridge, as shown in the figure.



29. The Akashi Kaikyo bridge in Japan is the longest suspension bridge in the world, with a main span of 1991 meters. Its main towers are 297 meters tall. The roadbed of the bridge is 14 meters thick and clears the water below by 65 meters.

(a) Find the coordinates of the vertex and one other point on the cable.

(b) Use the points from part (a) to find an equation for the shape of the cable in vertex form.

Answer.

(a) Vertex: $\left(\frac{1991}{2}, 79\right)$; y-intercept: $(0, 297)$

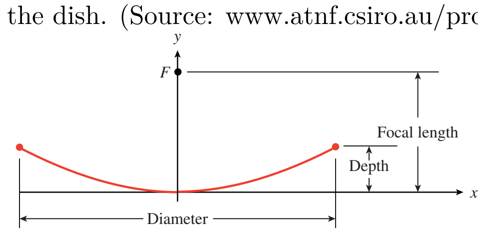
(b) $y = 0.00022(x - 995.5)^2 + 79$

30. A suspension bridge joining Sicily to the tip of Italy over the Straits of Messina has been planned planned and canceled multiple times. The main span of the bridge should be 3300 meters, and its main towers 375 meters tall. The roadbed should be 3 meters thick, clearing the water below by 65 meters.

(a) Find the coordinates of the vertex and one other point on the cable.

(b) Use the points from part (a) to find an equation for the shape of the cable in vertex form.

31. The Square Kilometre Array (SKA) is an international radio telescope project. Project members plan to build a telescope 30 times larger than the largest one currently available. The Australia Telescope National Facility held a workshop in 2005 to design an appropriate antenna. The antenna should be a parabolic dish with diameter from 12 to 20 meters, and the ratio of the focal length to the diameter should be 0.4. The figure shows a cross section of the dish. (Source: www.atnf.csiro.au/projects/ska/)



(a) You want to design a 20-meter-diameter parabolic antenna for the project. What will the focal length of your antenna be?

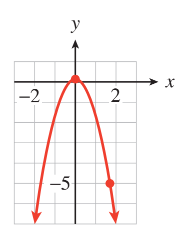
(b) The equation of the dish has the form $y = \dfrac{x^2}{4F}$, where F is the focal length. What is the equation of the parabola for your antenna?

(c) What is the depth of your parabolic antenna?

Answer.

(a) 8 m (b) $y = \dfrac{x^2}{32}$ (c) 3.125 m

32. Some comets move about the sun in parabolic orbits. In 1973, the comet Kohoutek passed within 0.14 AU (astronomical units), or 21 million kilometers, of the Sun. Imagine a coordinate system superimposed on a diagram of the comet's orbit, with the Sun at the origin, as shown in the figure. The units on each axis are measured in AU.



(a) The comet's closest approach to the Sun (called **perihelion**) occurred at the vertex of the parabola. What were the comet's coordinates at perihelion?

(b) When the comet was first discovered, its coordinates were $(1.68, -4.9)$. Find an equation for comet Kohoutek's orbit in vertex form.

Use your calculator's statistics features for Problems 33–38.

33. The table shows the height of a projectile at different times after it was fired.

Time (seconds)	2	4	6	8	10	12	14
Height (meters)	39.2	71.8	98.0	117.8	131.0	137.8	138.0

(a) Find the equation of the least-squares regression line for height in terms of time.

(b) Use the linear regression equation to predict the height of the projectile 15 seconds after it was fired.

(c) Make a scatterplot of the data and draw the regression line on the same axes.

(d) Find the quadratic regression equation for height in terms of time.

(e) Use the quadratic regression equation to predict the height of the projectile 15 seconds after it was fired.

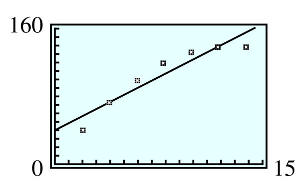
(f) Draw the quadratic regression curve on the graph from part (c).

(g) Which model is more appropriate for the height of the projectile, linear or quadratic? Why?

Answer.

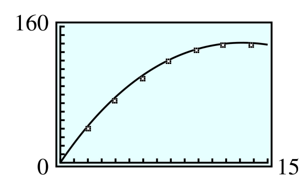
(a) $h = 8.24t + 38.89$

(b) 162.5 m

(c)


(d) $h = -0.81t^2 + 21.2t$

(e) 135.7 m

(f)


(g) Quadratic: Gravity will slow the projectile, giving the graph a concave down shape.

34. The table shows the height of a star-flare at different times after it exploded from the surface of a star.

Time (seconds)	0.2	0.4	0.6	0.8	1.0	1.2
Height (kilometers)	6.8	12.5	17.1	20.5	22.8	23.9

(a) Find the equation of the least-squares regression line for height of the flare in terms of time.

(b) Use the linear regression equation to predict the height of the flare 1.4 seconds after it exploded.

(c) Make a scatterplot of the data and draw the regression line on the same axes.

(d) Find the quadratic regression equation for height in terms of time.

(e) Use the quadratic regression equation to predict the height of the flare 1.4 seconds after it exploded.

(f) Draw the quadratic regression curve on the graph from part (c).

(g) Which model is more appropriate for the height of the star-flare, linear or quadratic? Why?

35. In the 1990s, an outbreak of mad cow disease (Creutzfeldt-Jakob disease) alarmed health officials in England. The table shows the number of deaths each year from the disease.

Year	’94	’95	’96	’97	’98	’99	2000	’01	’02	’03	’04
Deaths	0	3	10	10	18	15	28	20	17	19	9

(Source: www.cjd.ed.ac.uk/vcjdqsep05)

(a) The Health Protection Agency determined that a quadratic model was the best-fitting model for the data. Find a quadratic regression equation for the data.

(b) Use your model to estimate when the peak of the epidemic occurred and how many deaths from mad cow disease were expected in 2005.

Answer.

(a) $y = -0.587t^2 + 7.329t - 2.538$

(b) The predicted peak was in 2000, near the end of March. The model predicts 7 deaths for 2005.

36. The table shows the amount of nitrogen fertilizer applied to a crop of soybeans per hectare of land in a trial in Thailand and the resulting yield.

Nitrogen (kg)	0	15	30	60	120
Yield (tons)	2.12	2.46	2.65	2.80	2.60

(Source: www.arc-avrdc.org)

(a) Fit a quadratic regression equation to the data.

(b) Use your model to predict the maximum yield and the amount of nitrogen needed.

37. The number of daylight hours increases each day from the beginning of winter until the beginning of summer, and then begins to decrease. The table below gives the number of daylight hours in Delbert’s hometown last year in terms of the number of days since January 1.

Days since January 1	0	50	100	150	200	250	300
Hours of daylight	9.8	10.9	12.7	14.1	13.9	12.5	10.7

(a) Find the equation of the least-squares regression line for the number of daylight hours in terms of the number of days since January 1.

(b) Use the linear regression equation to predict the number of daylight hours 365 days after January 1.

(c) Make a scatterplot of the data and draw the regression line on the same axes.

(d) Find the quadratic regression equation for the number of daylight hours in terms of the number of days since January 1.

(e) Use the quadratic regression equation to predict the number of daylight hours 365 days after January 1.

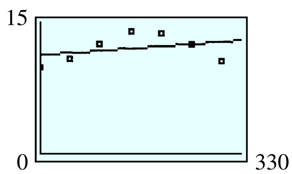
(f) Draw the quadratic regression curve on the graph from part (c).

(g) Predict the number of daylight hours 365 days since January 1 without using any regression equation. What does this tell you about the linear and quadratic models you found?

Answer.

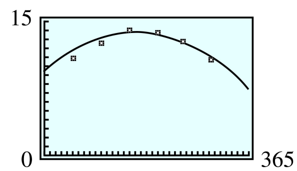
(a) $y = 0.0051t + 11.325$

(b) 13.2 hr

(c) 

(d) $y = -0.00016t^2 + 0.053t + 9.319$

(e) 7.4 hr

(f) 

(g) 9.8 hr (the same as the previous year); Neither model is appropriate.

38. To observers on Earth, the Moon looks like a disk that is completely illuminated at full moon and completely dark at new moon. The table below shows what fraction of the Moon is illuminated at 5-day interval after the last full moon.

Days since full moon	0	5	10	15	20	25
Fraction illuminated	1.000	0.734	0.236	0.001	0.279	0.785

(Source: www.arc-avrdc.org)

(a) Find the equation of the least-squares regression line for the fraction illuminated in terms of days.

(b) Use the linear regression equation to predict the fraction illuminated 30 days after the full moon.

(c) Make a scatterplot of the data and draw the regression line on the same axes.

(d) Find the quadratic regression equation for the fraction illuminated in terms of days.

(e) Use the quadratic regression equation to predict the fraction illuminated 30 days after the full moon.

(f) Draw the quadratic regression curve on the graph from part (c).

(g) Predict the fraction of the disk that is illuminated 30 days after the full moon without using any regression equation. What does this tell you about the linear and quadratic models you found?