CHAPTER SEVEN

Beyond Arithmetic

Arithmetical operations are immensely useful but can be taken only so far. We have reached a point where geometry and algebra need to be applied to the concepts of the previous chapters. Hopefully, the reader will see that without these concepts, exploring more involved real-world problems is not possible.

Explanations and definitions are given as the need arises. Though the reader is assumed to know little of the subject, he or she is required to remember well what has been stated. It may be best to have a pencil handy when reading the next two chapters and to proceed slowly. Much can still be gained by reading Chapters Seven and Eight even if all the mathematics is not initially understood.

* * *

Since we relate better to pictures than words or numbers, mathematical equations are usually represented graphically. When it comes to science, the expression, “One picture is worth a thousand words” is invaluable, especially when that “picture” is in the form of a graph. Interestingly enough, the idea of using mathematical relationships to describe reality goes back to antiquity, but the notion of associating a formula with a graph to express real-world phenomena is relatively recent. We begin an exploration of these ideas and their uses by treading back a few hundred years.

* * *

It wasn't until the sixteenth century that the idea of a stationary Earth and geocentric model (Earth centered) of the universe began to crumble. Copernicus's book *De Revolutions*, which advocated a heliocentric (sun centered) solar system, was published at the time of his death. Many believe this was intended to protect Copernicus from difficulties with the Church, which dogmatically held to a geocentric view. For Copernicus, the attraction of a heliocentric model of the solar system was its geometric and mathematical simplicity. This attitude, reducing complex ideas to the simplest possible mathematical and geometrical representations, is at the heart of modern scientific thought.

The structure of the geocentric universe in the sixteenth century was based largely on the writings of Ptolemy, and before him, Aristotle. It was a complicated mixture of circular motions like those inside a spring-wound watch. The need for such complexity arose by insisting that the Earth occupy the center of creation and that all heavenly motion be circular. The reasoning behind this unyielding paradigm (the geocentric concept was accepted for 1800 years) was based on erroneous perception and inflated ego. It **appeared** that all creation circled about the Earth and that God's grandest creation, man, should occupy center stage. It was also terribly difficult to argue for a moving Earth, since "common sense" told people they were not in motion. But Copernicus favored mathematical simplicity over appearance and dogma. Since his heliocentric model gave similar results to the geocentric model, why bother, he reasoned, to keep the more complicated structure. Copernicus's belief in the economy of creation is still echoed today.

Though Copernicus's model was closer to reality than Ptolemy's, it was still seriously flawed. Copernicus had correctly reasoned that the sun was the central body of the solar system but he held to the classical Greek view that the circle, "being the most perfect of forms," must necessarily represent the orbits of all heavenly bodies.

By the year 1600, the Danish astronomer Tycho Brahe (1546-1601) had compiled the most accurate planetary data then known.

Tycho, though an extraordinary astronomer, lacked the mathematical savvy necessary to refine Copernicus's planetary model. But Tycho's assistant, Johannes Kepler, who inherited the data after Tycho's death, did possess the mathematical sophistication and perseverance to properly analyze it. After two decades of work, Kepler had finished the last of what has become known as Kepler's Three Laws of Planetary Motion.

Kepler's first law transformed Copernicus's circular orbits into ellipses which produced a more accurate model for planetary motion. His other two laws involved the changing velocities planets undergo in their motion about the sun and the relationship between a planet's periodicity and its distance from the sun.

The significance of both Copernicus's and Kepler's work was in the preeminent role mathematics played in describing the physical world. They had transcended classical Greek thought by allowing the data to freely determine the geometry, rather than a supposed geometry forcing itself upon the "data." To the Greeks, "data" was meaningless, since they did not usually conduct experiments. Their "data" involved observing nature and then forming conclusions without testing. Curiously, modern physicists are more apt to place geometry (symmetries in nature) above all else. Perhaps we have come full circle.

Galileo Galilei was a contemporary of Kepler. He, more than anyone, is often credited with being the first true experimental scientist. Galileo conducted well designed experiments, wherein data was collected and analyzed mathematically. He sought relationships between variables, such as time and distance. One of Galilieo's many achievements was discovering the relationship between time and distance for falling objects near the Earth's surface. Much of his data analysis and thought experiments disproved a good deal of classical Greek thinking.

You may remember from your school days that it was Galileo who presumably dropped two different weights from the Tower of Pisa and showed that both reached the ground in essentially the same time—thereby proving all objects fall at the same rate. What you may not remember, is that his discovery showed that the mathematical relationship between time and distance for a falling object is nonlinear. That is to say, a falling object (strictly speaking in a vacuum) does not fall equal distances in equal intervals of time. Just as Copernicus and Kepler had done, Galileo let the data interpret the physical world. In the second half of the seventeenth century, the famous Sir Isaac Newton unified the work of Kepler and Galileo into three basic laws of nature which accurately described terrestrial and heavenly motion.

* * *

The key idea in using mathematical data is in defining the correct relationship between the **independent** variable and the **dependent** variable. For instance, if a person is paid five dollars an hour for performing a task, then his earnings are **dependent** on the time he works. **Time is therefore the independent variable, and earnings, the dependent variable**. A simple formula can be written for this relationship. If "t" represents the time worked, then $5.00 × t equals the money earned. The only algebra involved is in replacing the variable "t" with the appropriate number of hours worked. For example, if 4 hours and 45 minutes (4.75 hours) are worked, the pay is $5.00/hour × 4.75 hours = $23.75.

Often, when dealing with algebraic formulas, the multiplication symbol × is not written. This is done to eliminate any confusion between the multiplication symbol × and the variable x, thus simplifying the notation. The expression $5.00 × t, is therefore written as $5.00t, where the multiplication is now implicitly understood.

During the sixteenth and seventeenth centuries, there was much debate between the primacy of algebra over geometry or vice versa.[1] The philosopher-mathematician Rene Descartes (1596-1650) is usually credited with merging the two fields together into analytic geometry, or as it is also called, coordinate geometry. Descartes, however, never actually made use of any coordinate system. His contribution

was in providing the idea for associating algebraic formulas with geometrical shapes. Those who followed him were responsible for the creation of what has become known as the Cartesian Coordinate System.[2]

The Cartesian Coordinate System

The Cartesian coordinate system is a useful tool for understanding many present day problems. An explanation of the Cartesian coordinate system begins with the number line.

The typical number line has zero placed at its center with negative numbers to the left and positive numbers to the right. The number line in Figure 7.1 is one-dimensional, meaning only one variable is used to describe a quantity. A thermometer is an example of a vertical

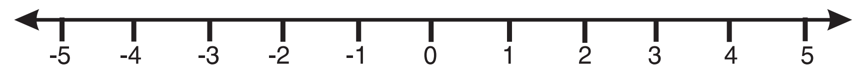

Figure 7.1
The number line

number line with temperature as the only variable.

Number lines offer a concise visual representation of numerical data. Figure 7.2 shows the distances between Santa Barbara,

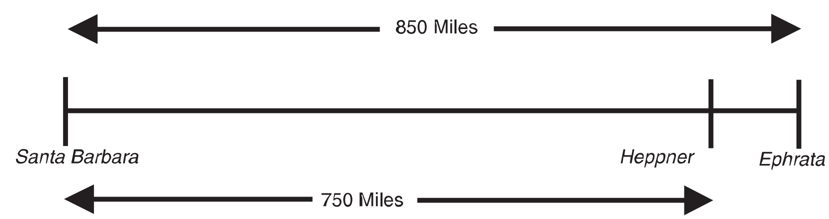

Figure 7.2
Using a number line as a visual aid

California and Heppner, Oregon, as well as between Santa Barbara and Ephrata, Washington. The lengths of the two lines gives an immediate "feel" of the comparative distances.

To investigate the relationship between two variables, two number lines, one for each variable, are needed. The lines, however, are arranged perpendicular to each other as in Figure 7.3. The negative parts of the number lines have been omitted for simplicity.

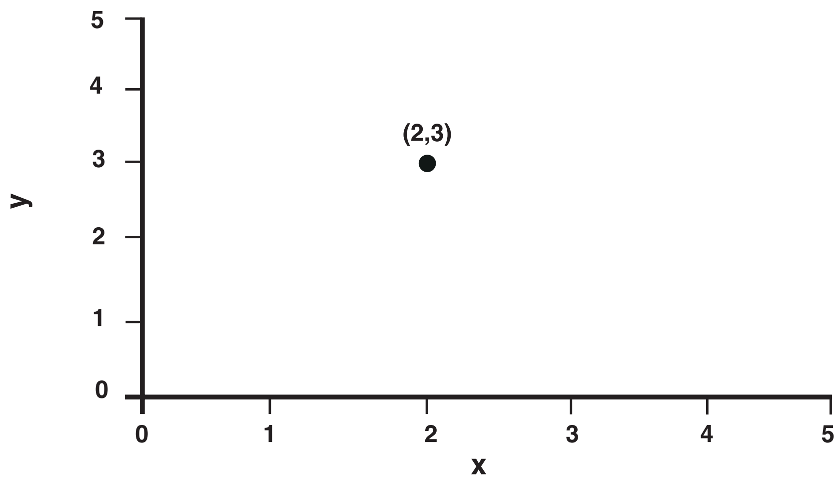

Figure 7.3
Two number lines at right angles forming a two dimensional coordiante plane with the ordered pair (2,3) plotted.

The two number lines form a coordinate plane, or a Cartesian Coordinate System. The horizontal number line is called the x-axis and the vertical number line is called the y-axis. To locate a position (a point) in the plane an "ordered pair" (or set of coordinates) is required. For example, in Figure 7.3 the ordered pair (2,3) represents the point in the plane that is located two units to the right of the point (0,0)—**called the origin**—and three units above the x-axis. **The horizontal distance from the origin is called the x-coordinate or abscissa, and the vertical distance from the x-axis is called the y-coordinate or ordinate**.

Had we wished to plot the point (3,2) rather than (2,3), we would have gone over 3 and up 2. The first coordinate in the parentheses, 2,

in the first example we chose, is associated with the horizontal direction or x-axis. Similarly, the second coordinate, 3, is associated with the vertical direction or y-axis. This is why the term "ordered pair" is used.

Ordered pairs can be generated by using a formula or by collecting data. More often than not, real-world problems are investigated by collecting data and then plotting points (ordered pairs) in the coordinate plane to generate a "scatter plot." Hopefully, a pattern emerges between the two variables that can be interpreted and extrapolated. The statistician, business person, and social and physical scientist, all try to find a "mathematical model" (an equation) that describes the relationship between the two variables. Once the correct relationship is found (that is, an equation is determined) the art of prediction making can be taken seriously. For instance, a graph that shows a correlation between the number of breast cancer deaths and fat consumption can be used to predict the optimum daily fat consumption, and the risks associated with varying levels of fat intake.

Classroom instruction in mathematics does not focus on data collection. The student is usually subjected to endless formulas without any real-world meaning. Only those few students who take courses in chemistry and physics begin to see how the mathematics they have studied is actually applied. Otherwise, students spend their time plugging the x-value into a given formula, which in turn produces a y-value, thereby obtaining ordered pairs which are then plotted on graph paper. This type of procedure always yields perfect looking relationships between variables, which is not the case in the real world. Even my gifted students were confounded when presented with real-world data for the first time. In a sense, they have been brainwashed; they do not see that in the real world most formulas represent tendencies, and not exact relationships. Students spend so much time seeing perfect curves emerge with formulas that they fail to see that nature, at best, only approximates these ideal shapes. Most students are not provided with an opportunity to plot real data for the purpose of **discovering** relationships.

Formulas and Data

An example of using a formula to generate ordered pairs can be shown with the algebraic expression $5.00t cited earlier.

Recall that "t" represented the time worked and the product $5.00t represented the money earned. In equation form, this is written as:

money earned = $5.00t, or as, y = $5.00t where y stands for the earnings and t represents time in hours and plays the role of x. It doesn't matter if we call the x-axis the t-axis as long as it's the horizontal axis.

To generate ordered pairs from this formula, we insert t values and compute the corresponding y values (e.g., if t = 5, y = $25).

Table 7.1

Values generated using the equation y = $5.00t

Time(t)	Earnings(y)
0	0
1	$ 5.00
2	$10.00
3	$15.00
4	$20.00
5	$25.00
6	$30.00
7	$35.00
8	$40.00

The values in Table 7.1 can be expressed as the ordered pairs: (0,0), (1,$5.00), (2,$10.00), (3,$15.00), (4,$20.00) (5,$25.00), (6,$30.00), (7,$35.00), (8,$40.00). These points are graphed in Figure 7.4.

Notice how perfectly straight the points line up—a result of plugging numbers into a formula. A straight line is drawn through the points in Figure 7.4, thus allowing one to read off earnings for other than the whole number of hours worked. No data collection was needed to generate this graph. The equation was determined by the pay rate, which was given. Now let's consider a real-world problem.

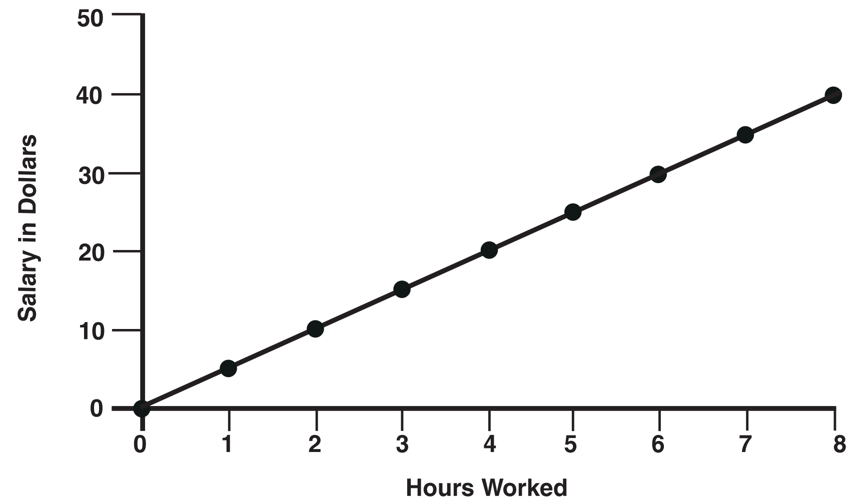

Figure 7.4
Graph showing money earned at a rate of $5.00 per hour

Data was collected from various countries for a 1975 study using fat consumption in grams per day and the number of breast cancer deaths per 100,000 females (Table 7.2).

Table 7.2

Fat Consumption and Breast Cancer Deaths

Country	Fat Consumption Grams/Day	Deaths/100,000
Thailand	24.9	0.75
Japan	35.3	3.56
El Salvador	39.7	1.13
Taiwan	42.4	4.43
Ceylon	46.0	2.44

Country	Fat Consumption Grams/Day	Deaths/100,000
Panama	56.5	7.44
Bulgaria	66.6	8.33
Portugal	67.2	12.74
Yugoslavia	70.8	6.43
Poland	86.6	10.59
Czechoslovakia	90.8	14.64
Hungary	96.5	13.21
Austria	116.8	16.54
Australia	128.4	18.57
Germany	134.4	16.67
Switzerland	135.0	21.36
Canada	140.1	23.18
United States	146.0	20.45
Netherlands	152.6	26.00

Source: Carroll,K., "Experimental Evidence of Dietary Factors in Hormone Dependent Cancers" Cancer Research, 35:3374, 1975

Data interpolated from graph: error ±2.0 grams/day, ±.20 deaths/100,000[3]

The second column—Fat Consumption in Grams/Day—provides the horizontal data and the third column—Deaths/100,000—provides the vertical data. An example of an ordered pair for this data set is (24.9,0.75) for Thailand. A scatter plot of Table 7.2 is given in Figure 7.5.

The first thing to look for in Figure 7.5 is a general trend. It appears that as fat consumption increases so does the number of deaths. This means there is a correlation between the two variables. But can the correlation be best represented with a straight line or a curved line? Since there is no obvious curvature to the data, a **linear-fit** (straight line) is the simplest and most reasonable approach.

The line in Figure 7.5 represents the general trend of the data. There are a number of different mathematical ways to determine

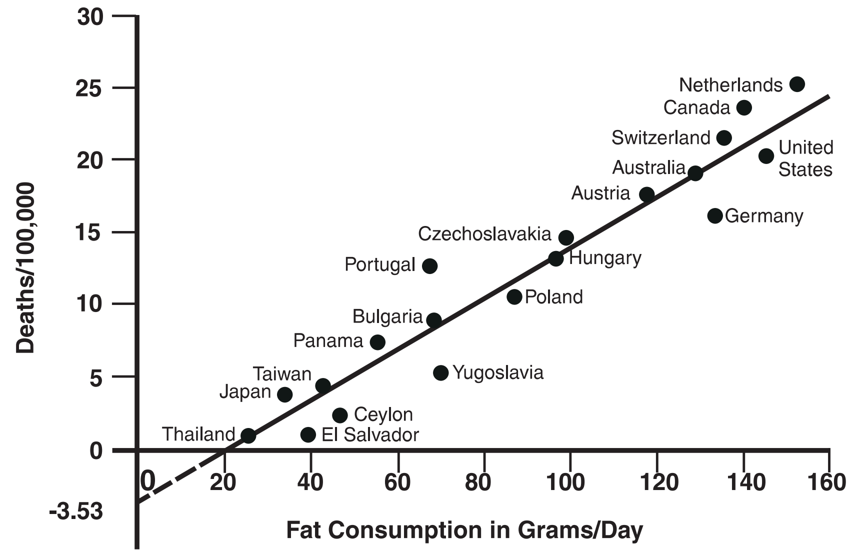

Figure 7.5

A scatter plot showing breast cancer deaths per 100,000 females verses fat consumed per day

Source: Adapted from a graph in John Robbin's *Diet for a New America*. Original source: Carroll, K., "Cancer Research," 35:3374, 1975

what is called a "**best-fit line**" for a scatter plot. None of these concerns us here. For our purposes it is enough to ask the reader to "**eyeball**" a line, i.e., sight along the data and draw a line that seems like a good representation of the trend.

Notice where the **trend-line** in figure 7.5 intersects the horizontal axis. The intersection is at approximately 20 grams of fat per day. This amount of fat represents about 10 percent of the calories $[(20 \times 9)/2000] \times 100 = 9\%$ on a 2000 calorie diet. Recall that the recommended percentage of calories derived from fat is 30 percent. Using the trend line we see that 67 grams of fat (which is equivalent to 30 percent fat consumption for a 2000 calorie diet) corresponds to about 8 deaths per 100,000—an improvement over the 1975 figure of 20 per 100,000 for the United States (see Figure 7.5 or Table 7.2).

Might we do better by reducing our fat consumption to 10 percent of our total caloric intake instead of the recommended 30 percent?

It is important to properly understand what a trend line means. The closer the data points cluster around the trend line the better the fit. In the ideal case, the line would go through every data point like the contrived salary example. The better the trend line is matched to the data, the greater the confidence we have in using the equation of the trend line to relate the two variables.

Scientifically, it is hoped that the equation used to describe the trend line represents the true relationship between the variables, admitting an accepted range of error. The more the data is scattered about the line in a random fashion without major deviations, the more apt we are to believe the fit is a true expression of reality. Then again, if the data were biased in one direction or if there was a distinctive curve away from the straight line fit, this would cause grave concern. Yet we might temporarily ignore such problems if the errors involved were not substantial. But under no condition would we believe such a trend line is the correct statement for relating the two variables. This brings up a somewhat philosophical point.

* * *

What made Kepler certain that the planets of our solar system travel in ellipses about the sun? He used Tycho's data for the planet Mars and found that the closest shape that fit the orbit was an ellipse. It should be pointed out that this conclusion was not at all obvious at the time, since the elliptical shape of Mars' orbit is nearly circular. But do the planets travel in perfect ellipses? Kepler's work shows (and Newton's universal theory of gravity confirms) that if only one force (due to the sun) acts on each planet, then each planet follows the path of a perfect ellipse. Only under this condition does the equation for an ellipse hold. But everything is connected. Some things are just connected more strongly than others.

The gravitational effects of the other planets of the solar system, and every star and particle of matter in the universe, are far too weak to perturb the overall elliptical motion of the planets as they orbit the sun. But regardless of how small these affects may be, they do still exist. In reality, the planets do not travel in orbits that repeat their motion over the same path. Their motion is therefore not exactly elliptical. We can easily account for the effects of the larger planets of the solar system, but not every particle of matter in the universe, nor is it necessary to do so. We ignore them because they will not influence our observations or calculations in any significant way. We can still predict where Mars will be at 10:00 P.M. on March 8, 2061. By assuming elliptical motion, we can send a spacecraft to orbit Mars or have Viking landers explore its surface. It matters little to the engineer if the geometric center of Mars is a fraction of an inch off of where the theoretical (elliptical) equation predicts.

From our vantage point, an ellipse expresses nature's structure for planetary orbits. But in reality, it is an incredibly good approximation—not truth. No equation **perfectly** describes any real-world phenomena. Ultimately, every equation that is written to explain a real-world situation is an idealized view of reality. So does mathematics describe objective reality or merely humankind's subjective "macroscopic" view? The best we can say is that the better the data fits our chosen mathematical model, the better we understand how the variables in question relate to each other. We do not lightly ignore other variables, but choose judiciously those that are considered to provide the best results and offer the greatest insights.

Finding the Equation of a Trend Line

It was shown previously that by plugging numbers into an equation, a series of points could be generated that formed a straight line. The reverse process is also possible; an equation can be determined from a line. Once a trend line is drawn, it represents the relationship between the variables in our experiment. Any process used to find the equation of a line can equally be applied to finding the equation of a trend line. This assumes we are merely taking our "best guess" at a representative line through the data—drawing it in by hand—rather than a mathematical approach.

Consider the trend line in figure 7.5, where Thailand appears to fall exactly on the line. If we had the equation for the trend line, we could insert the value for fat consumption (24.9 grams/day) and the equation would give the experimental value of .75 deaths/100,000. But if the fat consumption value for El Salvador (39.7 grams/day) were inserted, the equation would give a value closer to Japan's 3.56 deaths/100,000 rather than El Salvador's 1.13 deaths/100,000. One can, however, mathematically compute a reasonable error range for a scatter plot and its associated trend line. This is why it is not uncommon to see numbers such as 50.3 ± 2.5 arising from statistical results. Such numbers mean that there is a very high probability that if the experiment were performed, the answer would be some where between 47.8 and 52.8 for a given x value. It is unnecessary for our purposes to digress into the mechanics of error analysis. Just bear in mind that trend lines are not expected to yield exact answers, only reasonable approximations.

* * *

In the spirit of Descartes, we wish to find a relationship between an algebraic equation and a straight line. Odd as it may sound, right triangles are used to provide the needed insight.

A right triangle has two "legs" which meet at a 90° (right) angle and are referred to as the base and height of the triangle, and a third side connecting the two legs called the hypotenuse (see Figure 7.6). Any diagonal line can serve as the hypotenuse of an infinite number of right triangles.

In Figure 7.7 a line (ACE) is drawn in the coordinate plane with two triangles. The large triangle (ADE) and the small triangle (ABC) both share the line as their hypotenuse. The base AD has a length of 8 units and the height DE is 40 units. The ratio of the height to the

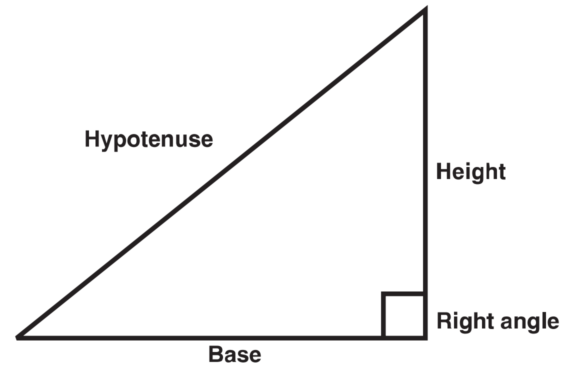

Figure 7.6
A right triangle

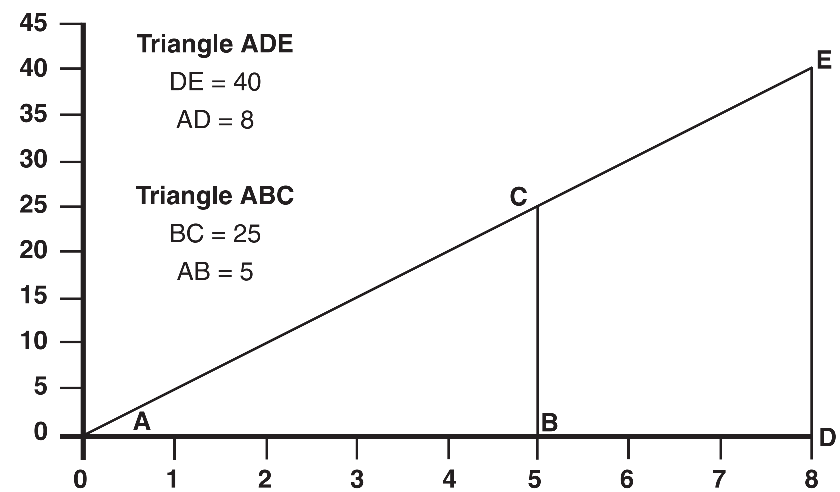

Figure 7.7
The ratio of height to base for any right triangle constructed in a like manner will be the same.

base (height/base) is 40/8. The same ratio holds for the smaller triangle, where the height/base is 25/5. For that matter, any right triangle that is formed by adjoining vertical and horizontal lines to the hypotenuse will give the same ratio of height/base. This is no coincidence. **All straight lines have what is known as a constant slope, defined by the ratio of height/base**. Since the height is a measure of the y value and the base is a measure of the x value, we can express this ratio as y/x = 5. **This is true only if the line contains the point (0,0)**. The relationship y/x = 5, is really the equation of the hypotenuse and can be rewritten as y = 5x. This is done by multiplying each side by x. A concrete example will help: 10/2 = 5, let the 10 play the role of y, and 2 play the role of x. If each side of the equation is multiplied by 2 we have: 10/2 × 2 = 5 × 2. The left side equals 10 (or y) and the right side equals 5 × 2 (or 5x). As long as the equation balances, each side still equals the other, nothing incorrect has been done. Any mathematical process is valid if the same mathematical operation is performed on each side of an equation. See the end of this chapter for more examples of this.

The equation y = 5x is identical to y = $5.00t (x and t represent unknowns), and therefore has the same graph. Since any straight line can have a horizontal and a vertical line adjoined to it, thus making a right triangle, the equation of the line can be found. **If the line contains the point (0,0) then the equation of the line can be found by finding the slope (height/base) of the line**. For example, if the ratio height/base for a line is 7, then y/x = 7; multiplying both sides by x gives y = 7x as the equation of the line.

The slope of a line is also referred to as the "**rise over the run**" as well as the ratio of height/base. If the line were nearly vertical then the triangle would have a small base (run) and large height (rise), meaning the slope is a large number. By the same reasoning, if the base (run) were large and the height (rise) small, the line would be closer to horizontal and its slope would be a small number. The ratio of the rise (height) to the run (base) is an indication of how the line is oriented in the coordinate plane. How would the line be oriented in the plane if the rise and run were equal? If the height were zero and the base any number?

When t = 0 (no time worked) in the equation y = $5.00t, the wage y is also 0. The graph of the line y = $5.00t, therefore passes through

the center of the coordinate system (0,0), **the origin**. Unfortunately, if a line does not go through the origin we cannot find its slope and consequently its equation by forming the simple ratio y/x as we did previously. To find the slope, and therefore the equation of any line in the plane, requires one additional step.

Finding the General Equation of a Line

To investigate the equation of a line not constrained to pass through the origin, we use the following example.

Many repair people charge an hourly rate and an additional fee for making a house call. Let's say a T.V. repair person charges a set fee of $35.00 just to walk through your door, and an hourly fee of $20.00. Therefore, at t = 0 you already owe $35.00. After the first hour of work, you owe $35.00 + $20.00/hour × 1 hour, after the second hour, $35.00 + $20.00/hour × 2 hours and so on. The general equa-

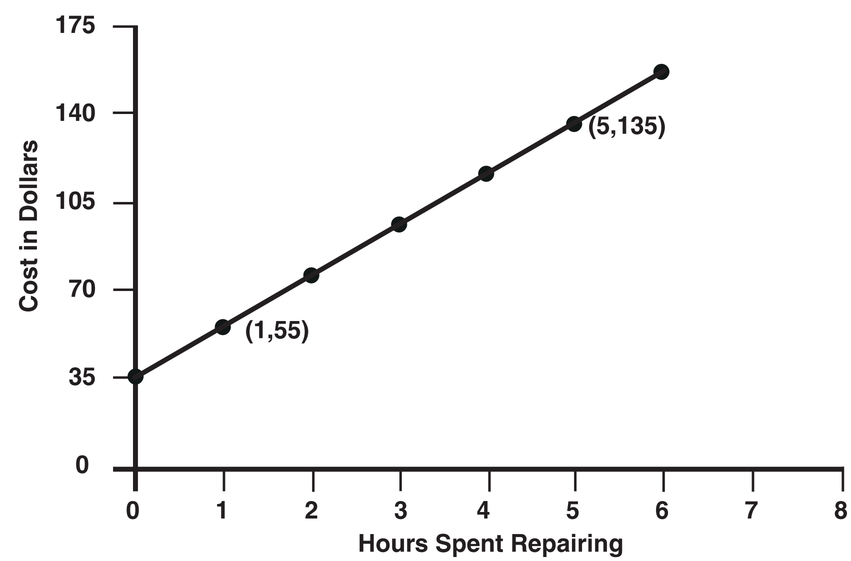

Figure 7.8
Graph showing cost to repair a T.V. over a six hour period

tion that represents this is: Cost = \$35.00 + \$20.00t or c = \$35.00 + \$20.00t. It is understood that c (the cost) is playing the role of y, and t (the time) means the same as x.

Several ordered pairs for this equation are:

(0 hours, \$35.00), (1 hour, \$55.00), (2 hours, \$75.00).

Figure 7.8 is a graph of the equation c = \$35.00 + \$20.00t. Notice that it is offset from the origin, and is oriented differently than the line y = \$5.00t.

Do the ratios of c/t formed in Figure 7.8 yield constant values? For example, does \$55.00/1 = \$75.00/2? No. But (\$55.00 − \$35.00)/1 equals (\$75.00 − \$35.00)/2. Subtracting \$35.00 removes the offset, and therefore preserves the rule for y/x.*

Any two points (x_1, y_1) and (x_2, y_2), where the subscripts, 1 and 2, stand for the first and second points picked, can be used to find a rise and a run. When the two y's are subtracted from each other ($y_2 - y_1$) the offset is canceled out. For instance, using the two points (1,55) and (5,135) gives \$135 − \$55 = \$80 for the rise. By expressing \$55.00 as \$35.00 + \$20.00, and \$135 as \$35.00 + \$100.00, it is easier to see how the \$35.00 cancels out. The calculation is: (\$35.00 + \$100.00) − (\$35.00 + \$20.00) = **\$35.00** − **\$35.00** + (\$100 − \$20) = \$80.00.

The general expression for the slope is: $(y_2 - y_1)/(x_2 - x_1)$

and the general equation for a line can be written as:

y = mx + b, where the m represents the slope, and **b** the offset, commonly called the **y-intercept** (that is, where the line intersects the y axis).

We can now verify that c = \$35 + \$20t (or c = 20t + 35, using the form y = mx + b) is the equation for the repair work.

The slope can be found using any two points on the line. For example, using the points (1,55) and (5,135) yields: m = (135 −

* Whenever a slope is stated as y/x there is an implied understanding that y/x is really (y-0)/(x-0). The differences in the y-values and x-values are often referred to as "delta-y" and "delta-x" values, written as Δy and Δx.

55)/(5 − 1) = 80/4 = \$20.00/h.

Similarly, for the points (2,75) and (4,115) we have:

m = (115 − 75)/(4 − 2) = 40/2 = \$20.00/h.

We already know that \$35.00 is equal to b, since it is the offset. Hence, y = mx + b is c = (\$20.00/h)t + \$35.00, where the roles of (x,y) are played by (t,c).

We are now in a position to determine the equation of the trend line in Figure 7.5. Ignoring the data, we choose any two points that we can accurately read on the line. (Remember, a line is composed of an infinite number of points.) Fortuitously, there are a number of data points that lie on the line, so determining the correct coordinates will be easy. It is usually best to choose two points that span as much of the data set as possible; for our case this means choosing Thailand (24.9,.75) and Australia (128.4,18.57).

The slope m equals: (18.57 − .75)/(128.4 − 24.9), and reduces to approximately .172 deaths per one hundred thousand women, per gram of fat consumed per day. Since we do not know "b" yet, all we can write is, y = .172x + b.

The unknown "b" can be found by extending the line and reading off where it crosses the y-axis, or mathematically by inserting any ordered pair (x,y) on the line into the equation y = .172x + b.

To better understand this last statement, consider the numerically simpler example: y = 2x + b and the point (3,12) known to be on the line. By plugging in for x and y we have: 12 = 2(3) + b. The unknown b can be found by subtracting 2(3) from each side of the equation [12 − 2(3) = 2(3) − 2(3) + b]. Recall that any operations are permissible as long as the same operation is performed on each side of the equation. This leaves 6 = b or b = 6, since the unknown is commonly written on the left. The final equation in our simplified example is: y = 2x + 6.

Returning to the original problem (y = .172x + b), it is already known that both (24.9,.75) and (128.4,18.57) are on the line, so either point can be used to find b. Since it does not matter which point we choose, we'll select the first point, (24.9,.75). Therefore: .75

= .172(24.9) + b, subtracting .172(24.9) from each side of the equation yields, b = −3.53. Now that m and b are known, the equation of the trend line can be written as: y = .172x − 3.53.

The above equation is the mathematical model relating breast cancer deaths to fat consumption. The model predicts that if 180 grams of fat are eaten per day, the incidence of breast cancer will be: y = .172(180) − 3.53, or 27.4 deaths/100,000. (Where we use only three significant figures.) It would be surprising if this answer were correct, because of the scatter of the real data about our model. Therefore, the y value should be stated with a plus or minus (±) range of values as previously discussed. Though it is not an accepted method for error analysis, finding the maximum difference between the model's value and the experimental value, for a given x, can give an upper bound on the error range. Remember, even with proper error analysis, predicted values represent only one possible answer out of a narrow range of possibilities.

As stated previously, the line crosses the x-axis at about 20 grams of fat, which means that the model predicts that close to 20 grams of fat consumption per day (remember there is error) is optimum for reducing breast cancer deaths. But what does the y-intercept, −3.53, tell us?

The y-intercept has no meaning here. It says that if zero fat is eaten there will be −3.53 deaths per hundred thousand females. Such "information" is meaningless, especially since without fat in our diets we would die. Like so many things, too little or too much can be hazardous.

For values less than 20 grams of fat, the mathematical model (y = .172x − 3.53) "breaks down" and is of no use. It is also unreasonable to assume that the model remains linear (straight) for ever increasing values of fat consumption. Students often memorize formulas pertaining to the real world, thinking them universal statements. Such formulas do not exist. All mathematical models (equations) have a limited domain over which they can yield meaningful results. A proper understanding of any physical process is necessary before

attempting to extrapolate and predict with a model.

The slope of the line also provides information. In the first example, a slope of $5.00/hour indicated the rate of pay was constant. In the breast cancer example, the slope of .172 gives the number of women per 100,000 dying for each gram of fat eaten over 20 grams per day.

On a graph with the y-axis representing distance traveled and the x-axis representing time, the slope has units of distance over time which is defined as speed. The slope of real-world data is a meaningful ratio. Most basic algebra books ignore this fact, and instead focus on the ratio of dimensionless quantities. (Dimensionless in this context means there are no units assigned to the numbers—no hours, miles, grams, etc.. Most dimensionless problems have no connection to the real world.) Even when units are given, mathematics text books and mathematics teachers are not prone to emphasize their importance.

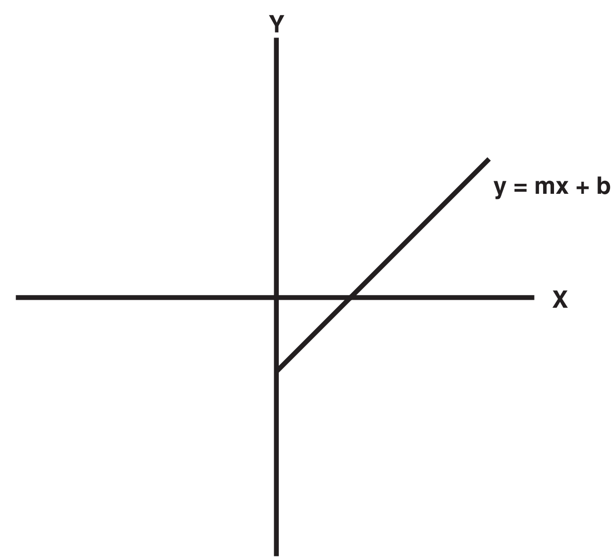

Figure 7.9a
A single-valued function

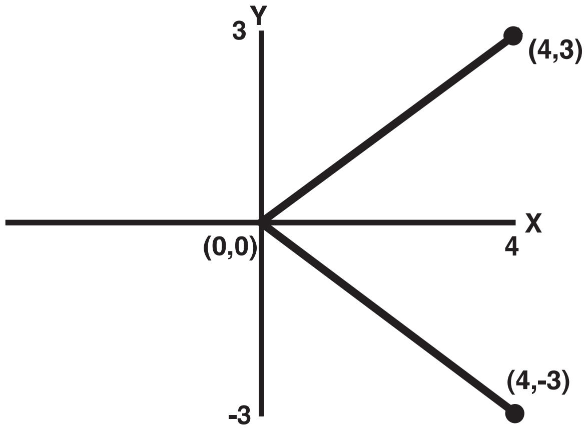

Figure 7.9b

An example of a graph that is not a function. The points (4,3) and (4,-3) are just two of an infinite number of points that disqualify the graph as a function.

Functions

A function is a special kind of relationship. Specifically, it is a unique relationship between the independent variable (the x-coordinate) and the dependent variable (the y-coordinate), where each independent variable must be associated with only one dependent variable. In the salary example, the worker is never paid two different sums for the same time worked; y = \$5.00t is therefore a function. Any equation that represents a straight line (all equations with the form y = mx + b) is a function—more correctly, a linear function. Study Figures 7.9a and 7.9b; the first is a function, the second one isn't.

In Figure 7.9a each ordered pair has a different x associated with a different y; it is thus a function—in particular, a **single-valued function**. Figure 7.9b, however, has two different y values for each x value, with the exception of (0,0); this disqualifies it from being a

function. **Specifically, if a vertical line intersects the graph in more than one place, the graph does not represent a functional relationship**. Furthermore, **if a horizontal line intersects the graph only once, the graph is a single-valued function**. **And if a horizontal line intersects the graph in more than one place, it is a many-to-one function. Often if a horizontal line intersects in just two places, the function is referred to as two-to-one.**

Figure 7.10 is an example of a two-to-one function. It shows how

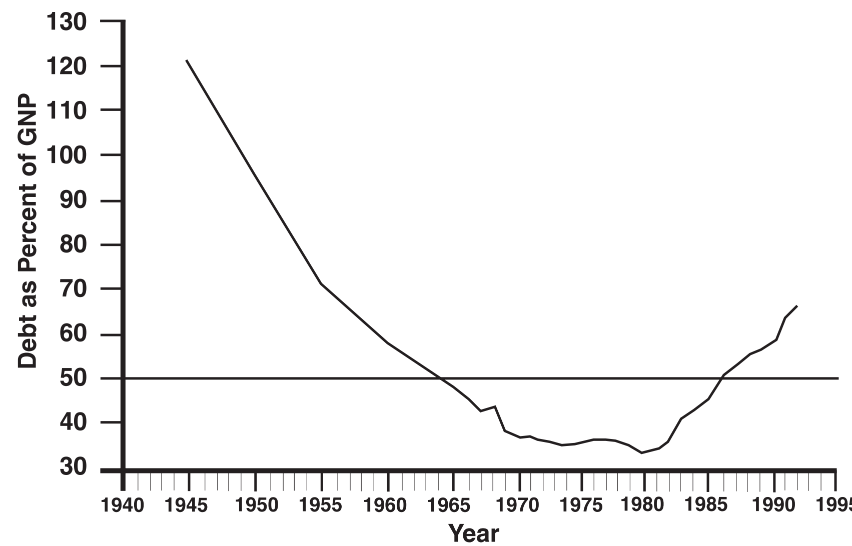

Figure 7.10

The percentage of GNP to national debt can be viewed as a multi-valued function.
Source: *Abstract of the United States 1992,*
U.S. Department of Commerce, Bureau of the Census

the percentage of national debt to GNP has changed over time.

A horizontal line drawn across the graph at the 50 percent level intersects at two different corresponding times (1964 and 1986). Other horizontal lines (below the 50% level) have more than two different x values for the same y value, hence the more proper designation as two-to-one.

(A certain liberty is being taken with this example, since the 50 percent mark does not correspond exactly to 1964 and 1986.)

Figure 7.10 is a two-to-one function when y equals 50 percent because two different independent variables are paired with the same dependent variable—(1964,50%) and (1986,50%). If the situation were reversed (that is, if the x values were the same and the y values were different, e.g., (1964,50%) and (1964,65%) then Figure 7.10 would not represent a function; nor would it make any sense within the given context.

Much of the physical world can be described with functions, which is why the study of mathematics takes on such importance. Everything from biological growth to the motion of the Earth through space can be written as a mathematical function.

Below are several exercises to help reinforce some of the concepts covered in this chapter. It is important that you at least read over problem 2, since we return to it in the next chapter.

Exercises

Find the value of "b" for the following problems.

1. y = 3x + b, (3,15) is on the given line. Answer: b = 6.
2. y = 2x + b, (6,8) is on the given line. Answer: b = −4.
3. 2y = 3.5x + b, (0,2) is on the given line. Answer: b = 4.
4. y/2 = × + b, (6,12) is on the given line. Answer: b = 0.

(A similar problem to the ones above was worked out on Page 139)

Problems:

There are four problems given below. Problems 1 and 2 are worked out, Problems 3 and 4 are not. Study Problems 1 and 2 and then do Problems 3 and 4 on your own.

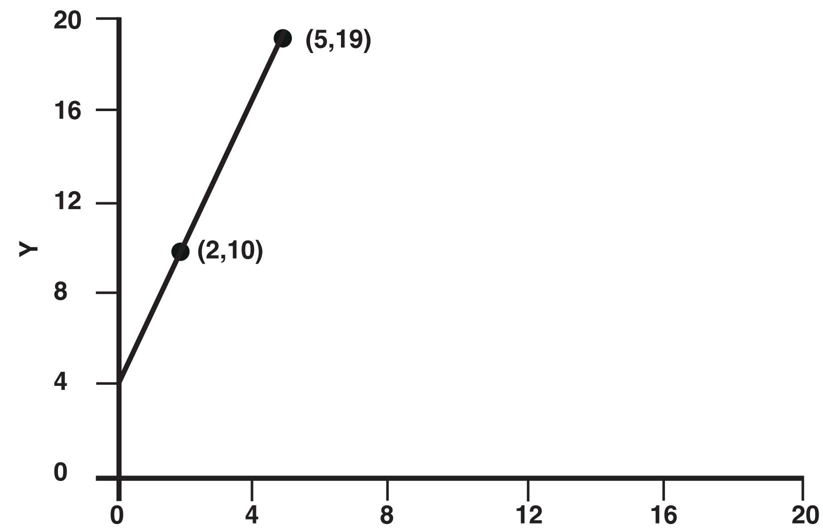

Figure 7.11
Problem 1

1. Two points determine a line. Given the points (2,10) and (5,19) determine the equation of the line that passes through them.

The objective is to find an equation with the form y = mx + b, where m and b are the unknown constants to be found.

First plot the two points and then draw the line they form (see Figure 7.11). It is not really necessary to draw the line, but a visual aid is always helpful.

The ratio of the height to the base for a right triangle constructed on this line equals the slope of the line m. The difference between the two given y values is (19 − 10). Similarly, the difference between the two given x values is (5 − 2). Therefore, the slope is m = (19 − 10)/(5 − 2) = 3. Knowing the slope allows us to write: y = 3x + b. Since the graph intersects the y-axis at 4, this must be b.

Often the intercept is found mathematically, because it cannot always be read off the graph so simply. This is done by replacing the x and y (in y = 3x + b) with any coordinates that are on the line. Since

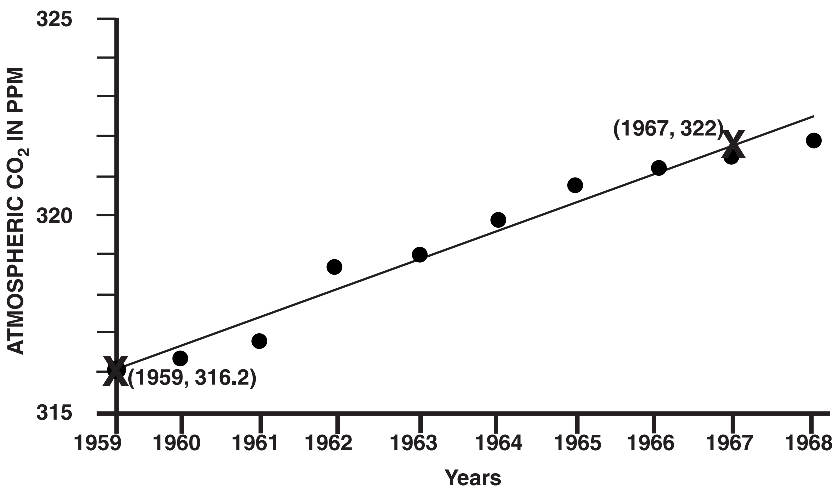

Figure 7.12

CO_2 in parts per million from 1959-1968.

Source: Adapted from, *Vital Signs 1992*, Worldwatch Institute

both (2,10) and (5,19) are on the line, either one can be substituted. Choosing (2,10) yields: 10 = 3(2) + b, or 10 = 6 + b. Subtracting 6 from each side of the equation gives, b = 4. Having found m = 3 and b = 4, the equation of the line can now properly be written as, y = 3x + 4.

Any value of x the reader chooses for the equation y = 3x + 4 will produce a corresponding value of y such that the ordered pair (x,y) must be on the line in Figure 7.11. Try inserting 5 for x to convince yourself that (5,19) is indeed on the line.

2. Figure 7.12 is a scatter plot of the atmospheric concentration of carbon dioxide in ppm from 1959 to 1968. The data is given as (year,ppm): (1959,316.2), (1960,316.7), (1961,317.2), (1962,318.9), (1963,319.2), (1964,320), (1965,320.9), (1966,321.5), (1967,321.8), (1968,322.5).
a) Assuming this data to be linear, "eye-ball" a best-fit line and determine its equation.

b) Predict the concentration of carbon dioxide in the year 2000, for this model. (We will discuss this problem more thoroughly in Chapter Eight.)
c) How reliable is the answer for part b?

Answers

a) The line passes through the data point (1959,316.2) and through the point (1967,322), which is not a data point. The slope of the line is: m = (322 − 316.2)/(1967 − 1959), which gives .725 ppm/year. Thus, so far we have: y = .725x + b. To find b, the temptation is to use 316.2 since this is where the line intersects the y-axis. However, **the y-intercept b is defined where x is zero**. Since we do not know the y-value when x equals zero (it is far off the graph), we must plug in one of the given points and solve for b. Using the first point (1959,316.2) gives:

316.2 =.725(1959) + b, b = 316.2 − .725(1959) or −1104.

The equation linking the variables of time and atmospheric carbon dioxide can then be written as: y = .725x − 1104.

There is an option that we could have used to simplify the numbers in this problem. It is perfectly correct to define 1959 as the zero year, and express the x-axis from 0 to 10. The equation would then simply be: y = .725x + 316.2

For instance, if we wish to calculate the ppm value in 1949 with this model, the first equation gives:

y = .725(1949) − 1104 which works out to 309 ppm. Using the second equation, we insert −10 for 1949, since 1959 = 0, thus giving us: y = .725(−10) + 316.2 which also works out to 309 ppm.

b) If "year 0" corresponds to 1959 then "year 41" corresponds to 2000. Using the second equation above we get:

y = .725(41) + 316.2 which gives 345.9 ppm.

c) The year 2000 is quite far beyond our last data point in 1968, so caution needs to be exercised here—unless we are sure that the rate of increase remains constant for the rest of the century. Had the question read 1975 instead of 2000, there would be less cause for concern, since we would not have to extrapolate so far into the "future." We will look more deeply into the implication of a limited data set (as in this problem) in the next chapter.

3. Find the equation of the line given the points (3,17) and (6,29). Don't forget to draw the line as a visual aid.

Answer: $y = 4x + 5$

4. Table 7.3 is a summary of the results of a study done across various countries in an attempt to find an association between the amount of cow's milk consumed in liters per year and the annual incidence of diabetes for children of ages 0 to 14.

Table 7.3

Table showing milk consumption rate for various countries and the corresponding incidence of childhood diabetes for children ages of 0-14.

Country	Liters Consumed Annually	Diabetes/100,000
Japan	38	1.4
France	79	4.4
Israel	90	3.9
Canada	107	8.9
United States	107	13.3
Netherlands	114	9.5
Great Britain	134	14.7
New Zealand	138	11.4
Denmark	138	13.7
Sweden	169	22.8
Norway	183	20.0
Finland	231	30.0

Source: "Diabetes Care" 1991.

Data interpolated from October 1992 "Scientific American" graph: error ±4 liters, error ±0.9 diabetes/100,000.

a) Graph a scatter plot with consumption as the independent variable (horizontal axis), and incidence number as the dependent variable (vertical axis).
b) Assume a linear model (a straight line) and "eye-ball" a best-fit line for the data.
c) According to your equation, how many liters of milk per year can safely be consumed? This can be read directly off the graph, or calculated from the answer in part b by setting the incidence number to zero.
d) What is the maximum number of ounces of milk that can be consumed per day based on your answer in part c? (One liter equals 1.0567 quarts)
e) What is the annual incidence for diabetes per 100,000 for a consumption rate of 250 liters per year according to your model?
f) How reliable is this value?
g) Does the data really represent a function since Canada and the United States both have the same independent variable paired with two different dependent variables, as do New Zealand and Denmark?
(Answers will vary depending on the best-fit estimate.)

Answers:

b) $y = .154x - 6.3$

c) 41 liters per year

d) 3.8 ounces

e) 32.2 diabetes/100,000

f) Since 250 liters per year is not too far removed from the Finland value, the resulting 32.2 diabetes/100,000 would not be an unreasonable extrapolation. Still, there is no way of really knowing (short of finding a country where the intake is 250 liters per year) without a better understanding of the human body. Additionally, common sense demands that the relationship cannot remain linear, since human beings do not have the capacity to drink endlessly.

g) Yes. If the association between the two variables was perfect (a straight line) it would mean there were no other effects influencing a child's disposition to developing diabetes. This is not the case. Other factors, such as genetics, will also play a role, but milk in the diet of children is still a strong enough factor to show an obvious trend. And it is with this trend line that the function is defined.