1

Introduction

The economic world is a misty region. The first explorers used unaided vision. Mathematics is the lantern by which what was before dimly visible now looms up in firm, bold outlines. The old phantasmagoria[1] *disappear. We see better. We also see further.*
—*Irving Fisher* (1892)

1.1 Why Economists Use Mathematics

Economic activity has been part of human life for thousands of years. The word "economics" itself originates from a classical Greek word meaning "household management." Even before the Greeks there were merchants and traders who exhibited an understanding of some economic phenomena; they knew, for instance, that a poor harvest would increase the price of corn, but that a shortage of gold might result in a decrease in the price of corn. For many centuries, the most basic economic concepts were expressed in simple terms requiring only the use of rudimentary mathematics. Concepts like integers and fractions, together with the operations of addition, subtraction, multiplication, and division, were sufficient to allow traders, merchants, farmers and other economic agents to discuss and debate the economic activities and events that affected their daily lives. These tools were enough to enable merchants to keep accounts and to work out what prices to charge.

[1]"Phantasmagoria" is a term invented in 1802 to describe an exhibition of optical illusions produced by means of a magic lantern.

Even calculations of interest on loans were not very complicated. Arithmetic could perform the tasks that merchants required of it even without the concept of zero and the decimal system of notation. Where a calculating device was required, the abacus was powerful enough.

The science of economics reached a turning point in the eighteenth century with the publication of works such as David Hume's *Political Discourses* (1752), François Quesnay's *Tableau Economique* (1758–1759), and Adam Smith's *The Wealth of Nations* (1776). Economic arguments began to be formalized and developed into theories. This created the need to express increasingly complex ideas and interrelationships in a straightforward manner. By the mid-1800s, some writers were beginning to use mathematics to communicate their theories. Some of the first to do this were economists such as Antoine Cournot (the first writer to define and draw an explicit demand curve, and to use calculus in solving a maximization problem in economics) and Léon Walras (who distinguished himself by writing down and solving the first multiequation model of general equilibrium of supply and demand in all markets simultaneously). They found that many of their ideas could be formulated most effectively by means of mathematical language, including algebraic symbols, simple diagrams, and graphs. Indeed, much more sophisticated economic concepts and increasingly complex economic theories have become possible as mathematical language has been used to express them.

Today, a firm understanding of mathematics is essential for any serious student of economics. Although simple economic arguments relying on only two or three variables can sometimes be made in a clear and convincing fashion without mathematics, if we want to consider many variables and the way they interact, it becomes essential to resort to a mathematical model.

As an example, suppose that some government agency is planning to allow a large amount of new housing to be constructed on some land it controls. What consequences will this have for employment? Initially, new jobs will be created in the construction sector as laborers are hired for the project. Moreover, the construction of new houses requires bricks, cement, reinforcing steel, timber, glass, and other building materials. Employment must also grow in firms that manufacture these materials. These producers in turn require materials from other producers, and so on. In addition to all these production effects, increased employment leads to increased incomes. If these income gains are not entirely neutralized by taxes, then a greater demand for consumer goods results. This in turn leads to an increased need for employment among producers of consumer goods, and again the flow of input requirements expands. At the same time, there are feedbacks in the system; for example, increased incomes also generate more demand for housing. In this manner, both positive and negative changes in one sector are transmitted to other sectors of the economy.

The point of this example is that the economic system is so complex that the final effects are difficult to determine without resorting to more formal mathematical devices such as a "circular-flow model" of the entire economy. An example will be the input–output model presented in Section 12.1.

Mathematical Analysis

The principal topic of this book is an important branch of mathematics called **mathematical analysis**. This includes differential and integral calculus and their extensions. Calculus was developed at the end of the seventeenth century by Newton and Leibniz. Their discoveries completely transformed mathematics, physics, and the engineering sciences, giving them all new life. In similar fashion, the introduction of calculus into economics has radically changed the way in which economists analyze the world around them. Calculus is now employed in many different areas of economics: for example, it is used to study the effects of relative price changes on demand, the effects of a change in the price or availability of an essential input such as oil on the production process, the consequences of population growth for the economy, and the extent to which a tax on energy use might reduce carbon dioxide emissions.

The following episode illustrates how economists can use mathematical analysis to solve practical problems. In February 1953, the Netherlands was struck by a catastrophic flood far more extensive than any previously recorded. The dikes protecting the country were washed away and over 1800 people died. Total damages were estimated at about 7% of national income for that year. A commission was established to determine how to prevent similar disasters in the future. Rebuilding the dikes to ensure 100% security would have cost an astronomical amount, even if it were possible at all. The real problem therefore involved a trade-off between cost and security: higher dikes would obviously cost more, but would reduce both the probability and likely severity of future flooding. So the commission had to try to select the optimal height for the dikes. Some economists on the commission applied *cost-benefit analysis*, a branch of economics that involves the use of mathematical analysis, in order to weigh the relative costs and benefits of different alternatives for rebuilding the dikes. This problem is discussed in more detail in Problem 7 in Section 8.4.

Such trade-offs are central to economics. They lead to optimization problems of a type that is naturally handled by mathematical analysis.

1.2 Scientific Method in the Empirical Sciences

Economics is now generally considered to be one of the *empirical sciences*. These sciences share a common methodology that includes the following as its most important elements:

1. Qualitative and quantitative observations of phenomena, either directly or by carefully designed experiment.
2. Numerical and statistical processing of the observed data.
3. Constructing theoretical models that describe the observed phenomena and explain the relationships between them.

4. Using these theoretical models in order to derive predictions.
5. Correcting and improving models so that they predict better.

Empirical sciences thus rely on processes of *observation*, *modeling*, and *verification*. If an activity is to qualify fully as an empirical science, each of the foregoing points is important. Observations without theory can only give purely descriptive pictures of reality that lack explanatory power. But theory without observation risks losing contact with the reality that it is trying to explain.

Many episodes in the history of science show the danger of error when "pure theory" lacks any foundation in reality. For example, around 350 B.C., Aristotle developed a theory that concluded that a freely falling object travels at a constant speed, and that a heavier object falls more quickly than a lighter one. This was convincingly refuted by Galileo Galilei in the sixteenth century when he demonstrated (partly by dropping objects from the Leaning Tower of Pisa) that, excluding the effects of air friction, the speed at which any object falls is proportional to the time it has fallen, and that the constant of proportionality is the same for all objects, regardless of their weight. Thus, Aristotle's theory was eventually disproved by empirical observation.

A second example comes from the science of astronomy. In the year 1800, Hegel advanced a philosophical argument to show that there could only be seven planets in the solar system. Hegel notwithstanding, an eighth planetary body, the asteroid Ceres, was discovered in January 1801. The eighth principal planet, Neptune, was discovered in 1846, and by 1930 the existence of Pluto was known.[2]

With hindsight, the falseness of these assertions by Aristotle and Hegel appears elementary. In all sciences, however, false assertions are being put forth repeatedly, only to be refuted later. Correcting inaccurate theories is an important part of scientific activity, and the previous examples demonstrate the need to ensure that theoretical models are supported by empirical evidence.

In economics, hypotheses are usually less precise than in the physical sciences, and so less obviously wrong than Aristotle's and Hegel's assertions just discussed. But there are a few old theories that have since become so discredited that few economists now take them seriously. One example is the "Phillips curve," that purported to show how an economy could trade off unemployment against inflation. The idea was that employment might be created through tax cuts and/or increased public expenditure, but at the cost of increased inflation. Conversely, inflation could be reduced by tax increases or expenditure cuts, but at the cost of higher unemployment.

[2]The process of discovery relied on looking at how the motion of other known planets deviated from the orbits predicted by Newton's theory of gravitation. These deviations even suggested where to look for an additional planet that could, according to Newton's theory, account for them. Until recently, scientists were still using Newton's theory to search for a tenth planetary body whose existence they suspected. However, more accurate estimates of the masses of the outer planets now suggest that there are no further planets to find after all.

Unlike Hegel, who could never hope to count all the planets, or Aristotle, who presumably never watched with any care the fall of an object that was dropped from rest, the Phillips curve was in fact based on rather careful empirical observation. In an article published in 1958, A. W. Phillips examined the average yearly rates of wage increases and unemployment for the economy of the United Kingdom over the long period from 1861 to 1957. The plot of those observations formed the Phillips curve, and the inflation–unemployment trade-off was part of conventional economic thinking until the 1970s. Then, however, the decade of simultaneous high inflation and high unemployment (stagnation and inflation, generally abbreviated "stagflation") that many Western economies experienced during the period 1973–1982 produced observations that obviously lay well above the usual Phillips curve. The alleged trade-off became hard to discern.

Just as Aristotle's and Hegel's assertions were revised in the light of suitable evidence, this stagflationary episode caused the theory behind the Phillips curve to be modified. It was suggested that as people learn to live with inflation, they adjust wage and loan contracts to reflect expected rates of inflation. Then the trade-off between unemployment and inflation that seemed to be described by the Phillips curve becomes replaced by a new trade-off between unemployment and the deviation in inflation from its expected rate. Moreover, this expected rate increases as the current rate of inflation rises. So lowering unemployment was thought to lead not simply to increased inflation, but to accelerating inflation that increased each period by more than was expected previously. On the other hand, when high inflation came to be expected, combating it with policies leading to painfully high unemployment would lead only to gradual decreases in inflation, as people's expectations of inflation fall rather slowly. Thus, the original Phillips curve theory has been significantly revised and extended in the light of more recent evidence.

Models and Reality

In the eighteenth century, the philosopher Immanuel Kant considered Euclidean geometry to be an absolutely true description of the physical space we observe through our senses. This conception seemed self-evident and was shared by all those who had reflected upon it. The reason for this agreement was undoubtedly that all the results of this geometry could be derived by way of irrefutable logic from only a few axioms, and that these axioms were regarded as self-evident truths about physical space. The first person to question this point of view was the German mathematician Gauss at the beginning of the 1800s. He insisted that the relationship between physical space and Euclid's model could only be made clear by empirical methods. During the 1820s, the first non-Euclidean geometry was developed—that is, a geometry built upon axioms other than Euclid's. Since that time it has been accepted that only observations can decide which geometric model gives the best description of physical space.

This shows how there can be an important difference between a mathematical model and its possible interpretations in reality. Moreover, it may happen that more than one model is capable of describing a certain phenomenon, such as the relationship between money supply and inflation in the United States or Germany. Indeed, this often seems to be the case in economics. As long as all the models to be considered are internally consistent, the best way to select among competing explanations is usually to see which one gives the best description of reality. But this is often surprisingly difficult, especially in economics.

In addition, we must recognize that a model intended to explain a phenomenon like inflation can never be considered as absolutely true; it is at best only an approximate representation of reality. We can never consider all the factors that influence such a complex phenomenon. If we tried to do so, we would obtain a hopelessly complicated theory. This is true not only for models of physical phenomena, but for all models within the empirical sciences.

These comments are particularly relevant for economic research. Consider once again the effects of allowing new housing to be built. In order to understand the full implications of this, an economist would require an incredible amount of data on millions of consumers, businesses, goods and services, etc. Even if it were available in this kind of detail, the amount of data would swamp the capacities of even the most modern computers. In their attempts to understand the underlying relationships in the economy, economists are therefore forced to use various kinds of aggregate data, among other simplifications. Thus, we must always remember that a model is only able to give an approximate description of reality; the goal of empirical researchers should be to make their models reflect reality as closely and accurately as possible.

1.3 The Use of Symbols in Mathematics

Before beginning to study any subject, it is important that everyone agrees on a common "language" in which to discuss it. Similarly, in the study of mathematics, which is in a sense a "language" of its own, it is important to ensure that we all understand exactly the same thing when we see a given symbol. Some symbols in mathematics nearly always signify the same definite mathematical object. Examples are 3, $\sqrt{2}$, π, and $[0, 1]$, which respectively signify three special numbers and a closed interval. Symbols of this type are called *logical constants*. We also frequently need symbols that can represent **variables.** The objects that a variable is meant to represent are said to make up the **domain of variation**. For example, we use the letter x as a symbol for an arbitrary number when we write

$$x^2 - 16 = (x + 4)(x - 4)$$

In words the expression reads as follows:

> The difference between the square of the number (hereby called x) and 16 is always equal to the product of the two numbers obtained by adding 4 to the number and subtracting 4 from the number x.

The equality $x^2 - 16 = (x + 4)(x - 4)$ is called an *identity* because it is valid identically for all x. In such cases, we sometimes write $x^2 - 16 \equiv (x + 4)(x - 4)$, where $\equiv$ is the symbol for an identity.

The equality sign (=) is also used in other ways. For example, we write $A = \pi r^2$ as the formula for the area A of a circle with radius r. In addition, the equality sign is used in equations such as

$$x^2 + x - 12 = 0$$

where x stands as a symbol for the unknown number. If we substitute various numbers for x, we discover that the equality sign is often invalid. In fact, the equation is only true for $x = 3$ and for $x = -4$, and these numbers are therefore called its *solutions*.

Example 1.1

A farmer has 1000 meters of fence wire with which to enclose a rectangle. If one side of the rectangle is x (measured in meters), find the area enclosed when x is chosen to be 150, 250, 350, and for general x. Which value of x do you believe gives the greatest possible area?

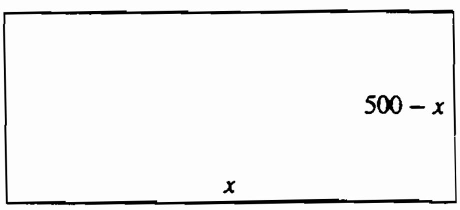
Solution If the other side of the rectangle is y, then $2x + 2y = 1000$. Hence, $x + y = 500$, so that $y = 500 - x$. (See Fig. 1.1.) The area A (in m^2) of this rectangle is, therefore,

$$A = x(500 - x) = 500x - x^2$$

Because both sides must be positive, x must be positive and $500 - x$ must be positive. This means that x must be between 0 and 500 m. The areas when $x = 150$, 250, and 350 are $150 \cdot 350 = 52{,}500$, $250 \cdot 250 = 62{,}500$, and $350 \cdot 150 = 52{,}500$, respectively. Of these, $x = 250$ gives the greatest value. In Problem 7 of Section 3.1 you will be asked to show that $x = 250$ really does give the greatest possible area.

When studying problems requiring several (but not too many) variables, we usually denote these with different letters such as a, b, c, x, y, z, A, B, and so on. Often, we supplement the letters of the Latin alphabet with lowercase and capital

FIGURE 1.1



Greek letters such as α, β, γ, Γ, and Ω. If the number of variables becomes large, we can use subscripts or superscripts to distinguish variables from each other. For example, suppose that we are studying employment in a country that is divided into 100 regions, numbered from 1 to 100. We can then denote employment in region 1 by N_1, employment in region 2 by N_2, and so on. In general, we can define

$$N_i = \text{total employment in region } i, \qquad i = 1, 2, \ldots, 100$$

The suffixes $i = 1, 2, \ldots, 100$ suggest that the index i can be an arbitrary number in the range from 1 to 100. If $N_{59} = 2690$, this means that 2690 people are employed in region 59. If we want to go further and divide the employed into men and women, we could denote the number of women (men) employed in region i by $N_i^{(W)}$ ($N_i^{(M)}$). Then, we would have $N_i^{(W)} + N_i^{(M)} = N_i$, for $i = 1, 2, \ldots, 100$. Note that this notation is actually much clearer than if we were to use 100 different letters to represent the variables N_i—even if we could find 100 different letters from some combination of the Latin, Greek, Cyrillic, and Sanskrit alphabets!

Many students who are used to dealing with algebraic expressions involving only *one* variable (usually x) have difficulties at first in handling expressions involving several variables. For economists, however, the previous example shows how important it is to be able to handle algebraic expressions and equations with many different variables. Here is another example.

Example 1.2

Consider the simple macroeconomic model

$$Y = C + \bar{I}, \qquad C = a + bY \tag{1}$$

where Y is the net national product, C is consumption, and $\bar{I}$ is the total investment, which is treated as fixed.[3] The three letters, $\bar{I}$, a, and b, denote positive numerical constants—for example, $\bar{I} = 100$, $a = 500$, and $b = 0.8$ are possible values of these constants. Rather than thinking of the two models with $\bar{I} = 100$, $C = 500 + 0.8Y$ and with $\bar{I} = 150$, $C = 600 + 0.9Y$ as entirely different, however, it is often more sensible to regard them as two particular instances of the general model [1], where $\bar{I}$, a, and b are unknown, and can vary; they are often called **parameters**. But they should be distinguished from the **variables** C and Y of the model.

After this discussion of constants as parameters of the model, solve [1] for Y.

Solution Substituting $C = a + bY$ from the second equation of [1] for C into the first equation gives

$$Y = a + bY + \bar{I}$$

[3] In economics, we often use a bar over a symbol to indicate that it is fixed.

Now rearrange this equation so that all the terms containing Y are on the left-hand side. This can be done by adding $-bY$ to both sides, thus canceling the bY term on the right-hand side to give

$$Y - bY = a + \bar{I}$$

Notice that the left-hand side is equal to $(1 - b)Y$, so $(1 - b)Y = a + \bar{I}$. Dividing both sides by $1 - b$, so that the coefficient of Y becomes 1, then gives the answer, which is

$$Y = \frac{a}{1-b} + \frac{1}{1-b}\bar{I}$$

This solution is a formula expressing Y in terms of the three parameters $\bar{I}$, a, and b. The formula can be applied to particular values of the constants, such as $\bar{I} = 100$, $a = 500$, $b = 0.8$, to give the right answer in every case. Note the power of this approach: The model is solved only once, and then numerical answers are found simply by substituting appropriate numerical values for the parameters of the model.

Problems

1. **a.** A person buys x_1, x_2, and x_3 units of three goods whose prices per unit are, respectively, p_1, p_2, and p_3. What is the total expenditure?
   b. A rental car costs F dollars per day in fixed charges and b dollars per kilometer. How much must a customer pay to drive x kilometers in 1 day?
   c. A company has fixed costs of F dollars per year and variable costs of c dollars per unit produced. Find an expression for the total cost per unit (total average cost) incurred by the company if it produces x units in one year.
   d. A person has an annual salary of $\$L$ and then receives a raise of $p\%$ followed by a further increase of $q\%$. What is the person's new yearly salary?
   e. A square tin plate 18 cm wide is to be made into an open box by cutting out equally sized squares of width x in each corner and then folding over the edges. Find the volume of the resulting box. (Draw a figure.)
2. **a.** Prove that

$$a + \frac{a \cdot p}{100} - \frac{\left(a + \frac{a \cdot p}{100}\right) \cdot p}{100}$$

can be written as

$$a\left[1 - \left(\frac{p}{100}\right)^2\right]$$

b. An item initially costs \$2000 and then its price is increased by 5%. Afterwards the price is lowered by 5%. What is the final price?

c. An item initially costs a dollars and then its price is increased by p%. Afterwards the (new) price is lowered by p%. What is the final price of the item? (After considering this problem, look at the expression in part (a).)

d. What is the result if one first *lowers* a price by p% and then *increases* it by p%?

3. Solve the following equations for the variables specified:

a. $x = \frac{2}{3}(y-3)+y$ for y **b.** $ax - b = cx + d$ for x

c. $AK\sqrt{L} = Y_0$ for L **d.** $px + qy = m$ for y

e. $\dfrac{\dfrac{1}{1+r} - a}{\dfrac{1}{1+r} + b} = c$ for r **f.** $Y = a(Y - tY - k) + b + I_p + G$ for Y

4. The relationship between a temperature measured in degrees Celsius (or Centigrade) (C) and in Fahrenheit (F) is given by $C = \frac{5}{9}(F - 32)$.

a. Find C when F is 32; find F when $C = 100$.

b. Find a general expression for F in terms of C.

c. One day the temperature in Oslo was $40° F$, while in Los Angeles it was $80° F$. How would you respond to the assertion that it was twice as warm in Los Angeles as in Oslo? (*Hint:* Find the two temperatures in degrees Celsius.)

5. If a rope could be wrapped around the earth's surface at the equator, it would be approximately circular and about 40 million meters long. Suppose we wanted to extend the rope to make it 1 meter above the equator at every point. How many more meters of rope would be needed? (Guess first, and then find the answer by precise calculation. For the formula for the circumference of the circle, see Appendix D.)

Harder Problems

6. Solve the following pair of simultaneous equations for x and y:

$$px + (1-q)y = R \qquad \text{and} \qquad qx + (1-p)y = S$$

7. Consider an equilateral triangle, and let P be an arbitrary point within the triangle. Let h_1, h_2, and h_3 be the shortest distances from P to each of the three sides. Show that the sum $h_1 + h_2 + h_3$ is independent of where point P is placed in the triangle. (*Hint:* Compute the area of the triangle as the sum of three triangles.)

1.4 The Real Number System

God created the integers;
everything else is the work of man.
—L. Kronecker

Real numbers were originally developed in order to measure physical characteristics such as length, temperature, and time. Economists also use real numbers to measure prices, quantities, incomes, tax rates, interest rates, and average costs, among other things. We assume that you have some knowledge of the real number system, but because of its fundamental role, we shall restate its basic properties.

Natural Numbers, Integers, and Rational Numbers

The everyday numbers we use for counting are 1, 2, 3, ... These are called **natural numbers.** Though familiar, such numbers are in reality rather abstract and advanced concepts. Civilization crossed a significant threshold when it grasped the idea that a flock of four sheep and a collection of four stones have something in common, namely "fourness." This idea came to be represented by symbols such as the primitive :: (still used on dominoes or playing cards), the modern 4, and the Roman numeral IV. This notion is grasped again and again as young children develop their mathematical skills.

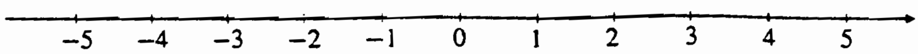
During the early stages of many cultures, day-to-day problems motivated the four basic arithmetic operations of addition, subtraction, multiplication, and division. If we add or multiply two natural numbers, we always obtain another natural number. Moreover, the operations of subtraction and division suggest the desirability of having a number zero $(4-4=0)$, negative numbers $(3-5=-2)$, and fractions $(3 \div 5 = 3/5)$. The numbers $0, \pm 1, \pm 2, \pm 3, \ldots$ are called the **integers.** They can be represented on a **number line** like the one shown in Fig. 1.2.

The **rational numbers** are those like 3/5, that can be written in the form a/b, where a and b are both integers. An integer n is also a rational number, because $n = n/1$. Examples of rational numbers are

$$\frac{1}{2}, \qquad \frac{11}{70}, \qquad \frac{125}{7}, \qquad -\frac{10}{11}, \qquad 0=\frac{0}{1}, \qquad -19, \qquad -1.26=-\frac{126}{100}$$

The rational numbers can also be represented on the number line. Imagine that we first mark 1/2 and all the multiples of 1/2. Then we mark 1/3 and all the

FIGURE 1.2 The number line.



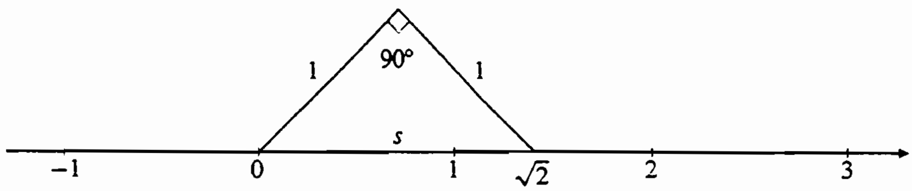


FIGURE 1.3

multiples of 1/3, and so forth. You can be excused for thinking that "finally" there will be no more places left for putting more points on the line. But in fact this is quite wrong. The ancient Greeks already understood that "holes" would remain in the number line even after all the rational numbers had been marked off. This is demonstrated in the construction in Fig. 1.3.

Pythagoras' theorem tells us that $s^2 = 1^2 + 1^2 = 2$, so $s = \sqrt{2}$. It can be shown, however, that there are no integers p and q such that $\sqrt{2} = p/q$. Hence, $\sqrt{2}$ is not a rational number. (Euclid proved this fact in about 300 B.C. See Problem 3 in Section 1.6.)

The rational numbers are therefore insufficient for measuring all possible lengths, let alone areas and volumes. This deficiency can be remedied by extending the concept of numbers to allow for the so-called **irrational numbers**. This extension can be carried out rather naturally by using decimal notation for numbers.

The Decimal System

The way most people write numbers today is called the **decimal system**, or the **base 10 system**. It is a positional system with 10 as the base number. Every natural number can be written using only the symbols, 0, 1, 2, ..., 9, that are called **digits**. You will note that "digit" also means "finger," or "thumb," and that most humans have 10 digits. The positional system defines each combination of digits as a sum of exponents of 10. For example,

$$1984 = 1 \cdot 10^3 + 9 \cdot 10^2 + 8 \cdot 10^1 + 4 \cdot 10^0$$

Each natural number can be uniquely expressed in this manner. With the use of the signs + and −, all integers, positive or negative, can be written in the same way. Decimal points also enable us to express rational numbers other than natural numbers. For example,

$$3.1415 = 3 + 1/10^1 + 4/10^2 + 1/10^3 + 5/10^4$$

Rational numbers that can be written exactly using only a finite number of decimal places are called **finite decimal fractions**.

Each finite decimal fraction is a rational number, but not every rational number can be written as a finite decimal fraction. We also need to allow for **infinite**

decimal fractions such as

$$100/3 = 33.333\ldots$$

where the three dots indicate that the decimal 3 recurs indefinitely.

If the decimal fraction is a rational number, then it will always be **periodic**—that is, after a certain place in the decimal expansion, it either stops or continues to repeat a finite sequence of digits. For example, $11/70 = 0.1\underbrace{571428}\underbrace{571428}5\ldots.$

Real Numbers

The definition of a real number follows from the previous discussion. We define a **real number** as an arbitrary infinite decimal fraction. Hence, a real number is of the form $x = \pm m.\alpha_1\alpha_2\alpha_3\ldots$, where m is an integer, and α_n $(n = 1, 2\ldots)$ is an infinite series of digits, each in the range 0 to 9. We have already identified the periodic decimal fractions with the rational numbers. In addition, there are infinitely many new numbers given by the nonperiodic decimal fractions. These are called **irrational numbers.** Examples include $\sqrt{2}$, $-\sqrt{5}$, π, $2^{\sqrt{2}}$, and $0.12112111211112\ldots.$

It turns out that, in general, it is very difficult to decide whether a given number is rational or irrational. It has been known since the year 1776 that π is irrational and since 1927 that $2^{\sqrt{2}}$ is irrational. However, we still do not know as of 1993 whether $2^{\sqrt{2}} + 3^{\sqrt{3}}$ is irrational or not. One might gain the impression that there are relatively few irrational numbers. In fact, there are (in a certain precise sense) infinitely more irrational numbers than there are rational numbers.

We mentioned earlier that each rational number can be represented by a point on the number line. But not all points on the number line represent rational numbers. It is as if the irrational numbers "close up" the remaining holes on the number line after all the rational numbers have been positioned. Hence, an unbroken and endless straight line with an origin and a positive unit of length is a satisfactory model for the real numbers. We frequently state that there is a *one-to-one correspondence* between the real numbers and the points on a number line.

The rational and irrational numbers are said to be "dense" on the number line. This means that between any two different real numbers, irrespective of how close they are to each other, we can always find both a rational and an irrational number—in fact, we can always find infinitely many of each.

When applied to the real numbers, the four basic arithmetic operations always result in a real number. The only exception is that we cannot divide by 0.

$\frac{a}{0}$ is not defined for any real number a

This is very important and should not be confused with $0/a = 0$, for all $a \neq 0$. Notice especially that 0/0 is not defined as any real number. For example, if a car requires 60 liters of fuel to go 600 kilometers, then its fuel consumption is 60/600 = 10 liters per 100 kilometers. However, if told that a car uses 0 liters of fuel to go 0 kilometers, we know nothing about its fuel consumption; 0/0 is completely undefined.

Inequalities

In mathematics and especially in economics, inequalities are encountered almost as often as equalities. It is important, therefore, to know and understand the rules for carring out calculations involving inequalities. These are presented in Section A.7 in Appendix A. The following example is of interest in statistics.

Example 1.3

Show that if $a \geq 0$ and $b \geq 0$, then

$$\sqrt{ab} \leq \frac{a+b}{2} \qquad [1.1]$$

Solution (You should first test this inequality by choosing some specific numbers, using a calculator if you wish.) To show the given inequality, it is enough to verify that $ab \leq (a+b)^2/4$ because then the square root of the left-hand side cannot exceed the square root of the right-hand side—that is, $\sqrt{ab} \leq \frac{1}{2}(a+b)$. To verify this, it is enough to check that the right-hand side minus the left-hand side is nonnegative. But indeed

$$\frac{(a+b)^2}{4} - ab = \frac{a^2+2ab+b^2-4ab}{4} = \frac{a^2-2ab+b^2}{4} = \frac{(a-b)^2}{4} \geq 0$$

In fact, essentially the same proof can be used to show that $\sqrt{ab} < \frac{1}{2}(a+b)$ unless $a = b$.

The number $\frac{1}{2}(a + b)$ is called the **arithmetic mean** of a and b, and $\sqrt{ab}$ is called the **geometric mean**. What does the inequality in [1.1] state about the different means?

Intervals

If a and b are two numbers on the number line, then we call the set of all numbers that lie between a and b an **interval**. In many situations, it is important to distinguish between the intervals that include their endpoints and the intervals that do not. When $a < b$, there are four different intervals that all have a and b as endpoints, as shown in Table 1.1. Note that the names in the table do not distinguish $[a, b)$ from $(a, b]$. To do so, one could speak of "closed on the left," "open on the right," and so on. Note, too, that an open interval includes neither of its endpoints, but a closed interval includes both of its endpoints. All four intervals, however, have the same length, $b - a$.

We usually illustrate intervals on the number line as in Fig. 1.4, with included endpoints represented by dots, and excluded endpoints at the tips of arrows. The intervals mentioned so far are all *bounded intervals*. We also use the word "interval" to signify certain unbounded sets of numbers. For example, we have

$$[a, \infty) = \text{all numbers } x, \text{ with } x \geq a$$

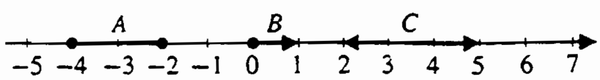
$$(-\infty, b) = \text{all numbers } x, \text{ with } x < b$$

with ∞ as the common symbol for infinity. Note that the symbol ∞ is not a number at all, and therefore the usual arithmetic rules do not apply to it. In $[a, \infty)$, the symbol ∞ is only a handy notation indicating that we are are considering the collection of *all* numbers larger than or equal to a, without any upper limit to the size of the number. From the preceding, it should be readily apparent what we mean

TABLE 1.1

Notation	*Name*	*The interval consists of all x satisfying:*
(a,b)	The **open** interval from a to b.	$a < x < b$
$[a,b]$	The **closed** interval from a to b.	$a \leq x \leq b$
$(a,b]$	The **half-open** interval from a to b.	$a < x \leq b$
$[a,b)$	The **half-open** interval from a to b.	$a \leq x < b$

FIGURE 1.4 $A = [-4, -2]$, $B = [0, 1)$, **and** $C = (2, 5)$.



by (a, ∞) and $(-\infty, b]$. The collection of all real numbers is sometimes denoted by the symbol $(-\infty, \infty)$.

Absolute Value

Let a be a real number and imagine its position on the number line. The distance between a and 0 is called the **absolute value** of a. If a is positive or 0, then the absolute value is the number a itself; if a is negative, then because distance must be positive, the absolute value is equal to the positive number $-a$.

The **absolute value** of a is denoted by $|a|$, and

$$|a| = \begin{cases} a, & \text{if } a \geq 0 \\ -a, & \text{if } a < 0 \end{cases} \qquad [1.2]$$

For example, $|13| = 13$, $|-5| = -(-5) = 5$, $|-1/2| = 1/2$, and $|0| = 0$.
Note: It is a common fallacy to assume that a must denote a positive number, even if this is not explicitly stated. Similarly, on seeing $-a$, many students are led to believe that this expression is always negative. Observe, however, that the number $-a$ is positive when a itself is negative. For example, if $a = -5$, then $-a = -(-5) = 5$. Nevertheless, it is often a useful convention in economics to define variables so that, as far as possible, their values are positive rather than negative. Where a variable has a definite sign, we shall try to follow this convention.

Example 1.4

(a) Compute $|x - 2|$ for $x = -3$, $x = 0$, and $x = 4$.
(b) Rewrite $|x - 2|$ using (1.2).

Solution

(a) For $x = -3$,

$$|x - 2| = |-3 - 2| = |-5| = 5$$

For $x = 0$,

$$|x - 2| = |0 - 2| = |-2| = 2$$

For $x = 4$,

$$|x - 2| = |4 - 2| = |2| = 2$$

(b) According to [1.2], $|x - 2| = x - 2$ if $x - 2 \geq 0$, that is, $x \geq 2$. However, $|x - 2| = -(x - 2) = 2 - x$ if $x - 2 < 0$, that is, $x < 2$.

Hence,

$$|x-2| = \begin{cases} x-2, & \text{if } x \geq 2 \\ 2-x, & \text{if } x < 2 \end{cases}$$

(Check this answer by trying the values of x tested in part (a).)

Let x_1 and x_2 be two arbitrary numbers. The **distance** between x_1 and x_2 on the number line is equal to $x_1 - x_2$ if $x_1 \geq x_2$, and equal to $-(x_1 - x_2)$ if $x_1 < x_2$. Therefore, we have

$$|x_1 - x_2| = \text{distance between } x_1 \text{ and } x_2 \text{ on the number line} \qquad [1.3]$$

In Fig. 1.5, we have indicated geometrically that the distance between 7 and 2 is 5, whereas the distance between -3 and -5 is equal to 2, because $|-3-(-5)| = |-3+5| = |2| = 2$.

Suppose $|x| = 5$. What values can x have? There are only two possibilities: either $x = 5$ or $x = -5$, because no other numbers have absolute values equal to 5. Generally, if a is greater than or equal to 0, then $|x| = a$ means that $x = a$ or $x = -a$. Because $|x| \geq 0$ for all x, the equation $|x| = a$ has no solution when $a < 0$.

If a is a positive number and $|x| < a$, then the distance from x to 0 is less than a, and so

$$|x| < a \text{ means that } -a < x < a \qquad [1.4]$$

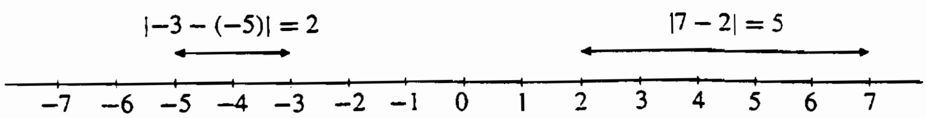
Furthermore, when a is nonnegative, it is clear that

$$|x| \leq a \text{ means that } -a \leq x \leq a \qquad [1.5]$$

Example 1.5

Find all the x such that $|3x-2| \leq 5$. Check first to see if this inequality is fulfilled for $x = -3$, $x = 0$, $x = 7/3$, and $x = 10$.

FIGURE 1.5 The distance between 7 and 2, and between -3 and -5.



Solution For $x = -3$, $|3x - 2| = |-9 - 2| = 11$; for $x = 0$, we have $|3x - 2| = |-2| = 2$; for $x = 7/3$, $|3x - 2| = |7 - 2| = 5$; and for $x = 10$, $|3x - 2| = |30 - 2| = 28$. Hence, we see that the given inequality is satisfied for $x = 0$ and $x = 7/3$, but not for $x = -3$ or $x = 10$.

From [1.5] we see that $|3x - 2| \leq 5$ means $-5 \leq 3x - 2 \leq 5$. Adding 2 to all three expressions gives $-5 + 2 \leq 3x - 2 + 2 \leq 5 + 2$, or $-3 \leq 3x \leq 7$. Dividing by 3 gives $-1 \leq x \leq 7/3$.

Problems

1. Which of the following numbers is a natural number, an integer, or a rational number?

a. 3.1415926 **b.** $\sqrt{\frac{9}{2} - \frac{1}{2}}$ **c.** $(\sqrt{3} + \sqrt{2})(\sqrt{3} - \sqrt{2})$ **d.** $3\pi - \frac{1}{4}$

2. Which of the following statements are correct?

a. 1984 is a natural number.
b. -5 is to the right of -3 on the number line.
c. -13 is a natural number.
d. There is no natural number that is not rational.
e. 3.1415 is not rational.
f. The sum of two irrational numbers is irrational.

3. For what real numbers x is each of the following expressions defined?

a. $\dfrac{3}{x - 4}$ **b.** $\dfrac{x - 1}{x(x + 2)}$ **c.** $\dfrac{3x}{x^2 + 4x - 5}$ **d.** $\dfrac{1/4}{x^2 + 4x + 4}$

4. Solve the following inequalities for y in terms of the other variables:

a. $3x + 4y \leq 12$ **b.** $-x + 3y - z > y - (x - y) + \frac{1}{2}z$

c. $px + qy \leq m \quad (q > 0)$

5. Consider Problem 1(c) in Section 1.3. Set up an inequality that determines how many units x the company must produce before the average cost falls below \$$q$. Solve the inequality for x. Put $F = 100{,}000$, $c = 120$, $q = 160$, and solve the problem for this case.

6. Calculate $|2x - 3|$, for $x = 0$, $1/2$, and $7/2$.

7. **a.** Calculate $|5 - 3x|$, for $x = -1$, 2, and 4.
b. Solve the equation $|5 - 3x| = 0$.
c. Rewrite $|5 - 3x|$ by using [1.2].

8. Determine x such that

a. $|3 - 2x| = 5$ **b.** $|x| \leq 2$ **c.** $|x - 2| \leq 1$

d. $|3 - 8x| \leq 5$ **e.** $|x| > \sqrt{2}$ **f.** $|x^2 - 2| \leq 1$

9. A 5-meter iron bar is to be produced. It is necessary that the length does not deviate more than 1 mm from its stated size. Write a specification for the rod's length x in meters: (a) by using a double inequality and (b) with the aid of an absolute-value sign.

1.5 A Few Aspects of Logic

An astronomer, a physicist, and a mathematician were travelling on a train in Scotland. Through the window they saw a flock of sheep grazing in a meadow. The astronomer remarked, "In Scotland all sheep are black." The physicist protested, "Some Scottish sheep are black." The mathematician declared, "In Scotland there exists a flock of sheep all of which are black on at least one side."

So far we have emphasized the role of mathematical models in the empirical sciences, especially in economics. The more complicated the phenomena to be described, the more important it is to be exact. Errors in models applied to practical situations can have catastrophic consequences. For example, in the early stages of the U.S. space program, a rocket costing millions of dollars to develop and build had to be destroyed only seconds after launch because a semicolon had been left out of the computer program intended to control the guidance system.

Although the consequences may be less dramatic, errors in mathematical reasoning also occur rather easily. In what follows, we offer a typical example of how a student (or professor) might use faulty logic and thus end up with an incorrect answer to a problem.

Example 1.6

Find a possible solution for the equation $x + 2 = \sqrt{4 - x}$.

"Solution" Squaring each side of the equation gives $(x + 2)^2 = (\sqrt{4-x})^2$, and thus $x^2 + 4x + 4 = 4 - x$. Rearranging this last equation gives $x^2 + 5x = 0$. Canceling x results in $x + 5 = 0$, and therefore $x = -5$.

According to this reasoning, the answer should be $x = -5$. Let us check this. For $x = -5$, we have $x + 2 = -3$. Yet $\sqrt{4-x} = \sqrt{9} = 3$, so this answer is incorrect. In Example 1.9, we explain how the error arose. (Note the wisdom of checking your answer whenever you think you have solved an equation.)

This example highlights the dangers of routine calculation without adequate thought. It may be easier to avoid similar mistakes after studying more closely the structure of logical reasoning.

Propositions

Assertions that are either true or false are called statements, or **propositions.** Most of the propositions in this book are mathematical ones, but others may arise in daily life. "All individuals who breathe are alive" is an example of a true proposition, whereas the assertion "all individuals who breathe are healthy" is an example of a false proposition. It should be noted that if the words used to express such assertions lack a precise meaning, it will often be difficult to distinguish between a true and a false proposition.

Suppose an assertion such as "$x^2 - 1 = 0$" includes one or more variables. By substituting various real numbers for the variable x, we can generate many different propositions, some true and some false. For this reason we say that the assertion is an **open proposition.** In fact, the proposition $x^2 - 1 = 0$ happens to be true if $x = 1$ or -1, but not otherwise. Thus, an open proposition is not simply true or false. It is neither true nor false until we choose a particular value for the variable. In practice we are somewhat careless about this distinction between propositions and open propositions; instead, we simply call both types propositions.

Implications

In order to keep track of each step in a chain of logical reasoning, it often helps to use implication arrows.

Suppose P and Q are two propositions such that whenever P is true, then Q is necessarily true. In this case, we usually write

$$P \implies Q \qquad [*]$$

This is read as "P implies Q," or "if P, then Q," or "Q is a consequence of P." The symbol $\implies$ is an **implication arrow,** and it points in the direction of the logical implication. Here are some examples of correct implications.

Example 1.7

(a) $x > 2 \implies x^2 > 4$.

(b) $xy = 0 \implies x = 0$ or $y = 0$.

(c) x is a square $\implies$ x is a rectangle.

(d) x is a healthy person $\implies$ x is breathing.

Notice that the word "or" in mathematics means the "inclusive or," signifying that "P or Q" means "either P or Q or both."

All the propositions in Example 1.7 are open propositions, just as are most propositions encountered in mathematics. An implication $P \implies Q$ means that for each value of some variable for which P is true, Q is also true.

In certain cases where the implication [*] is valid, it may also be possible to draw a logical conclusion in the other direction:

$$Q \implies P$$

In such cases, we can write both implications together in a single **logical equivalence**:

$$P \iff Q$$

We then say that "P is equivalent to Q," or "P if and only if Q," or just "P iff Q." Note that the statement "P only if Q" expresses the implication $P \Longrightarrow Q$, whereas "P if Q" expresses the implication $Q \Longrightarrow P$.

The symbol $\iff$ is an **equivalence arrow.** In previous Example 1.7, we see that the implication arrow in (b) could be replaced with the equivalence arrow, because it is also true that $x = 0$ or $y = 0$ implies $xy = 0$. Note, however, that no other implication in Example 1.7 can be replaced by the equivalence arrow. For even if x^2 is larger than 4, it is not necessarily true that x is larger than 2 (for instance, x might be -3); also, a rectangle is not necessarily a square; and, finally, just because person x is breathing does not mean that he or she is healthy.

Necessary and Sufficient Conditions

There are other commonly used ways of expressing that proposition P implies proposition Q, or that P is equivalent to Q. Thus, if proposition P implies proposition Q, we state that P is a "sufficient condition" for Q. After all, for Q to be true, it is sufficient that P is true. Accordingly, we know that if P is satisfied, then it is certain that Q is also satisfied. In this case, we say that Q is a "necessary condition" for P. For Q must necessarily be true if P is true. Hence,

> P is a **sufficient condition** for Q means: $P \Longrightarrow Q$
>
> Q is a **necessary condition** for P means: $P \Longrightarrow Q$

For example, if we formulate the implication in Example 1.7(c) in this way, it would read:

A necessary condition for x to be a square is that x be a rectangle.

or

A sufficient condition for x to be a rectangle is that x be a square.

The corresponding verbal expression for $P \iff Q$ is simply: *P is a necessary and sufficient condition for Q, or P if and only if Q, or P iff Q.* It is evident from this that it is very important to distinguish between the propositions "P is a necessary condition for Q" (meaning $Q \Longrightarrow P$) and "P is a sufficient condition for Q" (meaning $P \Longrightarrow Q$). To emphasize the point, consider two propositions:

1. Breathing is a necessary condition for a person to be healthy.
2. Breathing is a sufficient condition for a person to be healthy.

Evidently proposition 1 is true. But proposition 2 is false, because sick (living) people are still breathing. In the following pages, we shall repeatedly refer to necessary and sufficient conditions. Understanding them and the difference between them is a necessary condition for understanding much economic analysis. It is not a sufficient condition, alas!

Solving Equations

We shall now give examples showing how using implication and equivalence arrows can help avoid mistakes in solving equations like that in Example 1.6.

Example 1.8
Find all x such that $(2x-1)^2 - 3x^2 = 2\left(\frac{1}{2} - 4x\right)$.

Solution By expanding $(2x-1)^2$ and also multiplying out the right-hand side, we obtain a new equation that obviously has the same solutions as the original one:

$$(2x-1)^2 - 3x^2 = 2\left(\tfrac{1}{2} - 4x\right) \iff 4x^2 - 4x + 1 - 3x^2 = 1 - 8x$$

Adding $8x - 1$ to each side of the second equality and then gathering terms gives the equivalent expression

$$4x^2 - 4x + 1 - 3x^2 = 1 - 8x \iff x^2 + 4x = 0$$

Now $x^2 + 4x = x(x+4)$, and the latter expression is 0 if and only if $x = 0$ or $x + 4 = 0$. That is,

$$\begin{aligned} x^2 + 4x = 0 \iff x(x+4) = 0 &\iff x = 0 \quad \text{or} \quad x + 4 = 0 \\ &\iff x = 0 \quad \text{or} \quad x = -4 \end{aligned}$$

Putting everything together, we have derived a chain of equivalence arrows showing that the given equation is fulfilled for the two values $x = 0$ and $x = -4$, and for no other values of x. That is,

$$(2x-1)^2 - 3x^2 = 2\left(\tfrac{1}{2} - 4x\right) \iff x = 0 \quad \text{or} \quad x = -4$$

Example 1.9
Find all x such that $x + 2 = \sqrt{4 - x}$. (Recall Example 1.6.)

Solution Squaring both sides of the given equation yields

$$(x+2)^2 = \left(\sqrt{4-x}\right)^2$$

Consequently, $x^2 + 4x + 4 = 4 - x$, that is, $x^2 + 5x = 0$. From the latter equation it follows that

$$x(x+5) = 0$$

which implies $x = 0$ or $x = -5$. Thus, a necessary condition for x to solve $x+2 = \sqrt{4-x}$ is that $x = 0$ or $x = -5$. Inserting these two possible values of x into the original equation shows that only $x = 0$ satisfies the equation. The unique solution to the equation is, therefore, $x = 0$.

In finding the solution to Example 1.9, why was it necessary to test whether the values we found were actually solutions, whereas this step was unnecessary in Example 1.8? To answer this, we must analyze the logical structure of our solution to Example 1.9. With the aid of numbered implication and equivalence arrows, we can express the previous solution as

$$x+2 = \sqrt{4-x} \overset{(1)}{\Longrightarrow} (x+2)^2 = 4-x \overset{(2)}{\Longrightarrow} x^2+4x+4 = 4-x$$
$$\overset{(3)}{\Longrightarrow} x^2+5x = 0 \overset{(4)}{\Longrightarrow} x(x+5) = 0 \overset{(5)}{\Longrightarrow} x = 0 \text{ or } x = -5$$

Implication (1) is true (because $a = b \Longrightarrow a^2 = b^2$ and $\left(\sqrt{a}\right)^2 = a$). *It is important to note, however, that the implication cannot be replaced by an equivalence.* If $a^2 = b^2$, then either $a = b$ or $a = -b$; it need not be true that $a = b$. Implications (2), (3), (4), and (5) are also all true; moreover, all could have been written as equivalences, though this is not necessary in order to find the solution. Therefore, a chain of implications has been obtained that leads from the equation $x + 2 = \sqrt{4-x}$ to the proposition "$x = 0$ or $x = -5$." Because the implication (1) cannot be reversed, there is no corresponding chain of implications going in the opposite direction. We have verified that if the number x satisfies $x+2 = \sqrt{4-x}$, then x must be either 0 or -5; no other value can satisfy the given equation. However, we have not yet shown that either 0 or -5 really satisfies the equation. Until we try inserting 0 and -5 into the equation, we cannot see that only $x = 0$ is a solution. *Note that in this case, the test we have suggested not only serves to check our calculations, but is also a logical necessity.*

Looking back at Example 1.6, we now realize that two errors were committed. Firstly, the implication $x^2 + 5x = 0 \Rightarrow x + 5 = 0$ is wrong, because $x = 0$ is also a solution of $x^2 + 5x = 0$. Secondly, it is logically necessary to check if 0 or -5 really satisfies the equation.

The method used to solve Example 1.9 is the most common. It involves setting up a chain of implications that starts from the given equation and ends with a set of its possible solutions. By testing each of these trial solutions in turn, we find which of them really do satisfy the equation. Even if the chain of implications is also a chain of equivalences (as it was in Example 1.8), such a test is always a useful check of both logic and calculations.

Problems

1. Implications and equivalences can be expressed in ways that differ from those already mentioned. Use the implication or equivalence arrows to mark in which direction you believe the logical conclusions proceed in the following propositions:
   a. The equation $2x - 4 = 2$ is fulfilled only when $x = 3$.
   b. If $x = 3$, then $2x - 4 = 2$.
   c. The equation $x^2 - 2x + 1 = 0$ is satisfied if $x = 1$.
   d. If $x^2 > 4$, then $x > 2$ or $x < -2$, and conversely.
2. Consider the following six implications and decide in each case: (i) if the implication is true, and (ii) if the converse implication is true. (x, y, and z are real numbers.)
   a. $x = 2$ and $y = 5 \Longrightarrow x + y = 7$
   b. $(x-1)(x-2)(x-3) = 0 \Longrightarrow x = 1$
   c. $x^2 + y^2 = 0 \Longrightarrow x = 0$ or $y = 0$
   d. $x = 0$ and $y = 0 \Longrightarrow x^2 + y^2 = 0$
   e. $xy = xz \Longrightarrow y = z$
   f. $x > y^2 \Longrightarrow x > 0$
3. Consider the proposition $2x + 5 \geq 13$.
   a. Is the condition $x \geq 0$ necessary, sufficient, or both necessary and sufficient for the proposition to be satisfied?
   b. Answer the same question when $x \geq 0$ is replaced by $x \geq 50$.
   c. Answer the same question when $x \geq 0$ is replaced by $x \geq 4$.
4. Solve the equation

$$\frac{(x+1)^2}{x(x-1)} + \frac{(x-1)^2}{x(x+1)} - 2\frac{3x+1}{x^2-1} = 0$$

5. Solve the following equations:
   a. $x + 2 = \sqrt{4x + 13}$ **b.** $|x + 2| = \sqrt{4 - x}$ **c.** $x^2 - 2|x| - 3 = 0$
6. Solve the following equations:
   a. $\sqrt{x-4} = \sqrt{x+5} - 9$ **b.** $\sqrt{x-4} = 9 - \sqrt{x+5}$
7. Fill in the blank rectangles with "iff" (if and only if) when this results in a true statement, or alternatively with "if" or "only if."
   a. $x = \sqrt{4}$ $\boxed{}$ $x = 2$

b. $x^2 > 0$ $\boxed{}$ $x > 0$

c. $x^2 < 9$ $\boxed{}$ $x < 3$

d. $x(x^2 + 1) = 0$ $\boxed{}$ $x = 0$

e. $x(x + 3) < 0$ $\boxed{}$ $x > -3$

8. Consider the following attempt to solve the equation $x + \sqrt{x+4} = 2$: "From the given equation, it follows that $\sqrt{x+4} = 2 - x$. Squaring both sides gives $x + 4 = 4 - 4x + x^2$. After rearranging the terms, it is seen that this equation implies $x^2 - 5x = 0$. Canceling x, we obtain $x - 5 = 0$ and this equation is satisfied when $x = 5$."

a. Mark with arrows the implications or equivalences expressed in the text. Which ones are correct?

b. Give a correct solution to the equation.

9. For each of the following 6 propositions, state the negation as simply as possible.

a. $x \geq 0$ and $y \geq 0$.

b. All x satisfy $x \geq a$.

c. Neither x nor y is less than 5.

d. For each $\varepsilon > 0$, there exists a $\delta > 0$ such that B is satisfied.

e. No one can avoid liking cats.

f. Everyone loves someone at certain times.

10. "Supreme Court refuses to hear challenge to lower court's decision approving a trial judge's refusal to allow a defendant to refuse to speak." Has the defendant the right not to speak?

1.6 Mathematical Proof

In science, what can be proved should not be believed without proof.[4]
—R. Dedekind (1887)

In every branch of mathematics, the most important results are called **theorems.** Constructing logically valid proofs for these results often can be rather complicated. For example, the "four-color theorem" states that any map in the plane needs at most four colors in order that all contiguous regions should have different colors. Proving this involved checking hundreds of thousands of different cases, a task that was impossible without a sophisticated computer program.

In this book, we often omit formal proofs of theorems. Instead, the emphasis is on providing a good intuitive grasp of what the theorems tell us. However,

[4]Here is the German original: "Was beweisbar ist, soll in der Wissenschaft nicht ohne Beweis geglaubt werden."

although proofs do not form a major part of this book, it is still useful to understand something about the different types of proof that are used in mathematics. In fact, a proof that is actually readable is likely to some extent to rely on the reader's intuition. Although many mathematical logicians do take care to present every step and every argument, and this may indeed be a necessary step in enabling computers to check a proof, the overall result is usually unreadable by most people.

Every mathematical theorem can be formulated as an implication

$$P \Longrightarrow Q \qquad [*]$$

where P represents a proposition or a series of propositions called *premises* ("what we know"), and Q represents a proposition or a series of propositions that are called the *conclusions* ("what we want to know"). A statement of the form $P \iff Q$ can be regarded as two theorems.

Usually, it is most natural to prove a result of the type [*] by starting with the premises P and successively working forward to the conclusion Q; we call this a **direct proof**. Sometimes, however, it is more convenient to prove the implication $P \Longrightarrow Q$ by an **indirect proof**. In this case, we begin by supposing that Q is not true, and on that basis demonstrate that neither can P be true. This is completely legitimate, because we have the following equivalence:

$$P \Longrightarrow Q \quad \text{is equivalent to} \quad \textit{not } Q \Longrightarrow \textit{not } P \qquad [1.6]$$

It is helpful to consider how this rule of logic applies to some concrete examples:

If it is raining, the grass is getting wet

asserts precisely the same thing as

If the grass is not getting wet, then it is not raining.

If T denotes a triangle, then

The base angles of T are equal implies that T is isosceles asserts the same as *If T is not isosceles, then its base angles are not equal.*

There is a third method of proof that is also sometimes useful. It is called **proof by contradiction**. The method is based upon a fundamental logical principle: that it is impossible for a chain of valid inferences to proceed from a true proposition to a false one. Therefore, if we have a proposition R and we can derive a contradiction on the basis of supposing that R is false, then it follows that R must be true.

Example 1.10

Use three different methods to prove that

$$-x^2 + 5x - 4 > 0 \Longrightarrow x > 0$$

Solution

(a) *Direct proof:* Suppose $-x^2 + 5x - 4 > 0$. Adding $x^2 + 4$ to each side of the inequality gives $5x > x^2 + 4$. Because $x^2 + 4 \geq 4$, for all x, we have $5x > 4$, and so $x > 4/5$. In particular, $x > 0$.

(b) *Indirect proof:* Suppose $x \leq 0$. Then $5x \leq 0$ and so $-x^2 + 5x - 4$, as a sum of three nonpositive terms, is ≤ 0.

(c) *Proof by contradiction:* Suppose that the statement is not true. Then there has to exist an x such that $-x^2 + 5x - 4 > 0$ and $x \leq 0$. But if $x \leq 0$, then $-x^2 + 5x - 4 \leq -x^2 - 4 \leq -4$, and we have arrived at a contradiction.

Deductive vs. Inductive Reasoning

The three methods of proof just outlined are all examples of *deductive reasoning*, that is, reasoning based on consistent rules of logic. In contrast, many branches of science use *inductive reasoning*. This process draws general conclusions based only on a few (or even many) observations. For example, the statement that "the price level has increased every year for the last n years; therefore, it will surely increase next year too," demonstrates inductive reasoning. Owners of houses in California know how dangerous such reasoning can be in economics. This inductive approach is nevertheless of fundamental importance in the experimental and empirical sciences, despite the fact that conclusions based upon it never can be absolutely certain.

In mathematics, inductive reasoning is not recognized as a form of proof. Suppose, for instance, that the students taking a course in geometry are asked to show that the sum of the angles of a triangle is always 180 degrees. If they painstakingly measure as accurately as possible 1000 or even 1 million different triangles, demonstrating that in each case the sum of the angles is 180, would this not serve as proof for the assertion? No; although it would represent a very good indication that the proposition is true, it is not a mathematical proof. Similarly, in business economics, the fact that a particular company's profits have risen for each of the past 20 years is no guarantee that they will rise once again this year.

Nevertheless, there is a *mathematical* form of induction that is much used in valid proofs. This is discussed in Section B.5 in Appendix B.

Problems

1. Consider the following (dubious) statement: "If inflation increases, then unemployment decreases." Which of the following statements are equivalent?

a. For unemployment to decrease, inflation must increase.
b. A sufficient condition for unemployment to decrease is that inflation increases.
c. Unemployment can only decrease if inflation increases.
d. If unemployment does not decrease, then inflation does not increase.
e. A necessary condition for inflation to increase is that unemployment decreases.

2. Analyze the following epitaph: (a) using logic and (b) from a poetic viewpoint.

   Those who knew him, loved him.
   Those who loved him not, knew him not.

3. Fill in the details of the following proof that $\sqrt{2}$ is irrational. Suppose it were true that $\sqrt{2} = p/q$, where p and q are integers with no common factor. Then $p^2 = 2q^2$, which would mean that p^2, and hence p, would have 2 as a factor. Therefore, $p = 2s$ for some integer s, and so $4s^2 = 2q^2$. Thus, $q^2 = 2s^2$. It follows that q would also have 2 as a factor, a contradiction of the hypothesis that p and q have no common factor.

1.7 Set Theory

If you know set theory up to the hilt, and no other mathematics, you would be of no use to anybody. If you knew a lot of mathematics, but no set theory, you might achieve a great deal. But if you knew just some set theory, you would have a far better understanding of the language of mathematics.
—*I. Stewart* (1975)

In daily life, we constantly group together objects of the same kind. For instance, we refer to the university faculty to signify all the members of the academic staff at the university. A garden refers to all the plants that are growing in it. We talk about all firms with more than 1000 employees, all taxpayers in Los Angeles who earned between \$50,000 and \$100,000 in 1992, and so on. In all these cases, we have a collection of objects viewed as a whole. In mathematics, such a collection is called a **set**, and the objects are called the **elements** of, or the **members** of, the set.

How is a set specified? The simplest way is to list its members, in any order, between the two braces { and }. An example is the set

$$S = \{a, b, c\}$$

whose members are the first three letters in the alphabet of most languages of European origin, including English. Or it might be a set consisting of three members represented by the letters a, b, and c. For example, if $a = 0$, $b = 1$, and $c = 2$, then $S = \{0, 1, 2\}$. Also S denotes the set of roots of the cubic equation

$$(x - a)(x - b)(x - c) = 0$$

in the unknown x, where a, b, and c are any three real numbers.

Alternatively, suppose that you are to eat a meal at a restaurant that offers a choice of several main dishes. Four choices might be feasible—fish, pasta, omelette, and chicken. Then the *feasible set* F has these four members, and is fully specified as

$$F = \{\text{fish, pasta, omelette, chicken}\}$$

Notice that the order in which the dishes are listed does not matter. The feasible set remains the same even if the order of the items on the menu is changed.

Two sets A and B are considered **equal** if each element of A is an element of B and each element of B is an element of A. In this case, we write $A = B$. This means that the two sets consist of exactly the same elements. Consequently, $\{1, 2, 3\} = \{3, 2, 1\}$, because the order in which the elements are listed has no significance; and $\{1, 1, 2, 3\} = \{1, 2, 3\}$, because a set is not changed if some elements are listed more than once.

Specifying a Property

Not every set can be defined by listing all its members, however. Some sets can be infinite, that is, they contain an infinite number of members.

Actually, such infinite sets are rather common in economics. Take, for instance, the *budget set* that arises in consumer theory. Suppose there are two goods with quantities denoted by x and y that can be bought at prices p and q, respectively. A consumption bundle (x, y) is a pair of quantities of the two goods. Its value at prices p and q is $px + qy$. Suppose that a consumer has an amount m to spend on the two goods. Then the *budget constraint* is $px + qy \leq m$ (assuming that the consumer is free to underspend). If one also accepts that the quantity consumed of each good must be nonnegative, then the *budget set*, that will be denoted by B, consists of those consumption bundles (x, y) satisfying the three inequalities $px + qy \leq m$, $x \geq 0$, and $y \geq 0$. (The set B is shown in Fig. 2.41.) Standard notation for such a set is

$$B = \{(x, y) : px + qy \leq m, \ x \geq 0, \ y \geq 0\} \qquad [1.7]$$

The braces { } are still used to denote "the set consisting of." However, instead of listing all the members, which is impossible for the infinite set of points in the triangular budget set B, the set is specified in two parts. To the left of the colon, (x, y) is used to denote the form of the typical member of B, here a consumption bundle that is specified by listing the respective quantities of the two goods. To the right of the colon, the three properties that these typical members must satisfy are all listed, and the set thereby specified. This is an example of the general specification:

$$S = \{\text{typical member : defining properties}\}$$

Note that it is not just infinite sets that can be specified by properties—finite sets can also be specified in this way. Indeed, even some finite sets almost *have* to be specified in this way, such as the set of all human beings currently alive, or even (we hope!), the set of all readers of this book.

Mathematics makes frequent use of infinite sets. For example, in Section 1.4, we studied the set of positive integers, which is often denoted by N, as well as the set of rational numbers, denoted by Q, and the set of real numbers, denoted by R. All these sets are infinite.

Set Membership

As we stated earlier, sets contain members or elements. There is some convenient standard notation that denotes the relation between a set and its members. First,

$$x \in S$$

indicates that x *is an element of* S. Note the special symbol $\in$ (which is a variant of the Greek letter ϵ, or "epsilon"). Occasionally, one sees $S \ni x$ being used to express exactly the same relationship as $x \in S$. The symbol "$\ni$" is generally read as "owns," but is not used very often. To express the fact that x is *not* a member of S, we write $x \notin S$. For example, $d \notin \{a, b, c\}$ says that d is not an element of the set $\{a, b, c\}$.

For additional illustrations of set membership notation, let us return to our earlier examples. Given the budget set B in [1.7], let (x^*, y^*) denote the consumer's actual purchases. Then it must be true that $(x^*, y^*) \in B$. Confronted with the choice from the set of feasible main courses $F = \{\text{fish, pasta, omelette, chicken}\}$, let s denote your actual selection. Then, of course, $s \in F$. This is what we mean by "feasible set"—it is possible only to choose some member of that set but nothing outside it.

In the example of choice from four main courses, it may be argued that if none is to the customer's liking, then she cannot be prevented from ordering nothing at all from the menu. She can eat somewhere else instead, or simply go hungry. If that is what she does, she is not really choosing outside her feasible set. Rather, our previous description of the feasible set should be expanded to include the option of ordering none of the four available dishes. Thus, the customer's true feasible set is

$$F_5 = \{\text{fish, pasta, omelette, chicken, none of the previous four}\}$$

In the end, she can only avoid choosing something from this by choosing more than one item. If this is not allowed, then F_5 is her true feasible set.

Subsets

Let A and B be any two sets. Then A is a **subset** of B if it is true that every member of A is also a member of B. So A is smaller than B in some sense, even though A and B could actually be equal. This relationship is expressed symbolically by $A \subset B$:

$$A \subset B \iff [x \in A \Rightarrow x \in B]$$

A special case of a subset is when A is a *proper subset* of B, meaning that $A \subset B$ and $A \neq B$.[5]

Set Operations

Sets can be combined in many different ways. Especially important are three operations: union, intersection, and the difference of sets, as shown in Table 1.2.

TABLE 1.2

Notation	*Name*	*The set consists of*
$A \cup B$	A **union** B	The elements that belong to at least one of the sets A and B.
$A \cap B$	A **intersection** B	The elements that belong to both A and B.
$A \setminus B$	A **minus** B	The elements that belong to A, but not to B.

Thus,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}$$

Example 1.11

Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 6\}$. Find $A \cup B$, $A \cap B$, $A \setminus B$, and $B \setminus A$.

Solution $A \cup B = \{1, 2, 3, 4, 5, 6\}$, $A \cap B = \{3\}$, $A \setminus B = \{1, 2, 4, 5\}$, $B \setminus A = \{6\}$.

[5] Sometimes the notation $A \subset B$ is reserved for the case when A is a subset of B satisfying $A \neq B$, just as $a < b$ is reserved for when $a \leq b$ and $a \neq b$. Then $A \subseteq B$ is used to denote that A is a subset of B. However, there is rarely any need to specify that A is a proper subset of B, and when there is, this can easily be done verbally.

An economic example can be obtained by considering particular sets of taxpayers in 1990. Let A be the set of all those taxpayers who had an income of at least \$15,000 and let B be the set of all who had a net worth of at least \$150,000. Then $A \cup B$ would be those taxpayers who earned at least \$15,000 or who had a net worth of at least \$150,000, whereas $A \cap B$ are those taxpayers who earned at least \$15,000 and who also had a net worth of at least \$150,000. Finally, $A \setminus B$ would be those who earned at least \$15,000 but who had less than \$150,000 in net worth.

If two sets A and B have no elements in common, they are said to be **disjoint**. The symbol "$\emptyset$" denotes the set that has no elements. It is called the **empty set**. Thus, sets A and B are disjoint if and only if $A \cap B = \emptyset$.

A collection of sets is often referred to as a family of sets. When considering a certain family of sets, it is usually natural to think of each set in the family as a subset of one particular fixed set Ω, hereafter called the **universal set.** In the previous example, the set of all taxpayers in 1990 would be an obvious choice for a universal set.

If A is a subset of the universal set Ω, then according to the definition of difference, $\Omega \setminus A$ is the set of elements of Ω that are not in A. This set is called the **complement** of A in Ω and is sometimes denoted by $\complement A$, so that $\complement A = \Omega \setminus A$. Other ways of denoting the complement of A include A^c and $\tilde{A}$.

When using the notation $\complement A$, it is important to be clear about which universal set Ω is used to construct the complement.

Example 1.12

Let the universal set Ω be the set of all students at a particular university. Moreover, let F denote the set of female students, M the set of all mathematics students, C the set of students in the university choir, B the set of all biology students, and T the set of all tennis players. Describe the members of the following sets: $\Omega \setminus M$, $M \cup C$, $F \cap T$, $M \setminus (B \cap T)$, and $(M \setminus B) \cup (M \setminus T)$.

Solution $\Omega \setminus M$ consists of those students who are not studying mathematics, $M \cup C$ of those students who study mathematics and/or are in the university choir. The set $F \cap T$ consists of those female students who play tennis. The set $M \setminus (B \cap T)$ has those mathematics students who do not both study biology and play tennis. Finally, the last set $(M \setminus B) \cup (M \setminus T)$ has those students who either are mathematics students not studying biology or mathematics students who do not play tennis. Do you see that the last two sets are equal? (For arbitrary sets M, B, and T, it is true that $(M \setminus B) \cup (M \setminus T) = M \setminus (B \cap T)$. It will be easier to verify this equality after you have read the following discussion of Venn diagrams.)

Venn Diagrams

When considering the relationships between several sets, it is instructive and extremely helpful to represent each set by a region in a plane. The region is drawn so that all the elements belonging to a certain set are contained within some closed region of the plane. Diagrams constructed in this manner are called **Venn diagrams.** The definitions discussed in the previous section can be illustrated as in Fig. 1.6.

By using the definitions directly, or by illustrating sets with Venn diagrams, one can derive formulas that are universally valid regardless of which sets are being considered. For example, the formula $A \cap B = B \cap A$ follows immediately from the definition of the intersection between two sets. It is somewhat more difficult to verify directly from the definitions that the following relationship is valid for all sets A, B, and C:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad [*]$$

With the use of a Venn diagram, however, we easily see that the sets on the right- and left-hand sides of the equality sign both represent the shaded set in Fig. 1.7. The equality in [*] is therefore valid.

It is important that the three sets A, B, and C in a Venn diagram be drawn in such a way that all possible relations between an element and each of the three sets are represented. In other words, the following eight different sets all should be nonempty: (1): $(A \cap B) \setminus C$; (2): $(B \cap C) \setminus A$; (3): $(C \cap A) \setminus B$; (4): $A \setminus (B \cup C)$; (5): $B \setminus (C \cup A)$; (6): $C \setminus (A \cup B)$; (7): $A \cap B \cap C$; and (8): $C(A \cup B \cup C)$. (See Fig. 1.8.) Notice, however, that this way of representing sets in the plane easily becomes unmanageable if four or more sets are involved, because then there would have to be at least $16 (= 2^4)$ regions in any such Venn diagram.

From the definition of intersection and union (or by the use of Venn diagrams), it easily follows that $A \cup (B \cup C) = (A \cup B) \cup C$ and that $A \cap (B \cap C) = (A \cap B) \cap C$.

FIGURE 1.6 Venn diagrams.

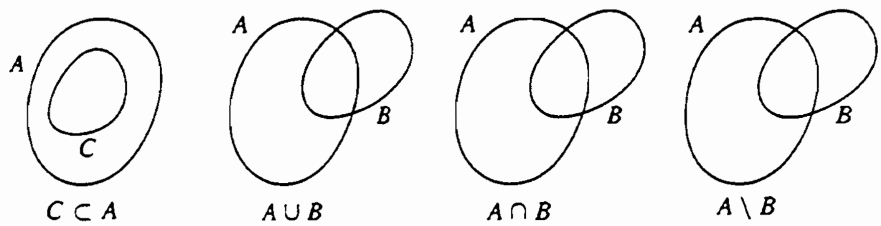


FIGURE 1.7

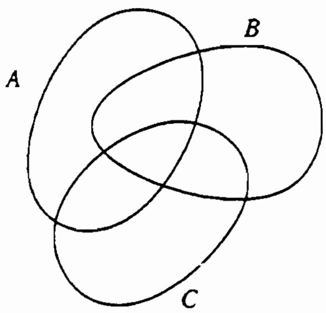
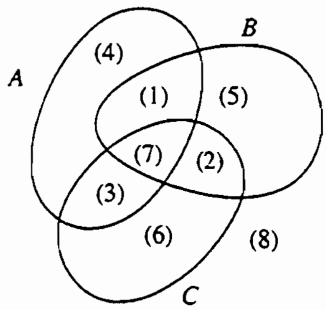


FIGURE 1.8



Consequently, it does not matter where the parentheses are placed. In such cases, the parentheses can be dropped and the expressions written as $A \cup B \cup C$ and $A \cap B \cap C$. Note, however, that the parentheses cannot generally be moved in the expression $A \cap (B \cup C)$, because this set is not always equal to $(A \cap B) \cup C$. Prove this fact by considering the case where $A = \{1, 2, 3\}$, $B = \{2, 3\}$, and $C = \{4, 5\}$, or by using a Venn diagram.

Problems

1. Let $A = \{2, 3, 4\}$, $B = \{2, 5, 6\}$, $C = \{5, 6, 2\}$, and $D = \{6\}$.
   a. Determine if the following statements are true: $4 \in C$; $5 \in C$; $A \subset B$; $D \subset C$; $B = C$; and $A = B$.
   b. Find $A \cap B$; $A \cup B$; $A \setminus B$; $B \setminus A$; $(A \cup B) \setminus (A \cap B)$; $A \cup B \cup C \cup D$; $A \cap B \cap C$; and $A \cap B \cap C \cap D$.
2. **a.** Are the greatest painter among the poets and the greatest poet among the painters one and the same person?
   b. Are the oldest painter among the poets and the oldest poet among the painters one and the same person?
3. With reference to Example 1.12, write the following statements in set terminology:
   a. All biology students are mathematics students.
   b. There are female biology students in the university choir.
   c. Those female students who neither play tennis nor belong to the university choir all study biology.
4. Let F, M, C, B, and T be the sets in Example 1.12. Describe the following sets: $F \cap B \cap C$; $M \cap F$; and $((M \cap B) \setminus C) \setminus T$.
5. Justify the following formulas by either using the definitions or by using Venn diagrams:
   a. $A \cup B = B \cup A$ **b.** $A \cup A = A$
   c. $A \cap A = A$ **d.** $A \cap \emptyset = \emptyset$
   e. $A \cup \emptyset = A$ **f.** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
6. Determine which of the following formulas are true. If any formula is false, find a counterexample to demonstrate this, using a Venn diagram if you find it helpful.
   a. $A \setminus B = B \setminus A$ **b.** $A \subset B \iff A \cup B = B$
   c. $A \subset B \iff A \cap B = A$ **d.** $A \cap B = A \cap C \Longrightarrow B = C$
   e. $A \cup B = A \cup C \Longrightarrow B = C$ **f.** $A \setminus (B \setminus C) = (A \setminus B) \setminus C$
7. Make a complete list of all the different subsets of the set $\{a, b, c\}$. How many are there if the empty set and the set itself are included? Do the same for the set $\{a, b, c, d\}$.

8. A survey revealed that 50 people liked coffee, 40 liked tea, 35 liked both coffee and tea, and 10 did not like either coffee or tea. How many persons in all responded to the survey?

9. If A is a set with a finite number of elements, let $n(A)$ denote the number of elements in A. If A and B are arbitrary finite sets, prove the following:

a. $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

b. $n(A \setminus B) = n(A) - n(A \cap B)$

10. If A and B are two arbitrary sets, define the **symmetric difference** between A and B as

$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

Obviously, $A \triangle B = B \triangle A$, whereas $A \setminus B \neq B \setminus A$ (in general). Prove by using a Venn diagram, or in some other way, the following:

a. $A \triangle B = (A \cup B) \setminus (A \cap B)$

b. $(A \triangle B) \triangle C$ consists of those elements that occur in just one of the sets A, B, and C, or else in all three.

11. One of the following identities is not generally valid. Which one?

a. $(A \triangle B) \triangle C = A \triangle (B \triangle C)$

b. $(A \cap C) \triangle B = (A \triangle B) \cap (C \triangle B)$

c. $A \triangle A = \emptyset$

12. **a.** A thousand people took part in a survey to reveal which newspaper, A, B, or C, they had read on a certain day. The responses showed that 420 had read A, 316 had read B, and 160 had read C. Of these responses, 116 had read both A and B, 100 had read A and C, 30 had read B and C, and 16 had read all three papers.

(i) How many had read A, but not B?

(ii) How many had read C, but neither A nor B?

(iii) How many had read neither A, B, nor C?

b. Denote the complete set of all 1000 persons in the survey by Ω (the universal set). Applying the notation in Problem 9, we have $n(A) = 420$ and $n(A \cap B \cap C) = 16$, for example. Describe the numbers given in part (a) in a similar manner. Why is the following equation valid?

$$n(\Omega \setminus (A \cup B \cup C)) = n(\Omega) - n(A \cup B \cup C)$$

c. Prove that if A, B, and C are arbitrary finite sets, then

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C)$$
$$-n(B \cap C) + n(A \cap B \cap C)$$

2

Functions of One Variable: Introduction

... mathematics is not so much a subject as a way of studying any subject, not so much a science as a way of life.
—*G. Temple* (1981)

Functions are of fundamental importance in practically every area of pure and applied mathematics, including mathematics applied to economics. The language of mathematical economics is full of terms like supply and demand functions, cost functions, production functions, consumption functions, and so on. Here and in the next chapter, we present a general discussion of functions of one real variable, illustrated by some very important examples.

2.1 Introduction

One variable is a function of another if the first variable *depends* upon the second. For instance, the area of a circle is a function of its radius. If the radius r is given, then the area A is determined. In fact $A = \pi r^2$, where π is the numerical constant 3.14159....

The measurement of temperature provides another example of a function. If C denotes the temperature expressed in degrees Centigrade (or Celsius), this is a function of F, the same temperature measured in degrees Fahrenheit, because $C = \frac{5}{9}(F - 32)$.

In ordinary conversation, we sometimes use the word "function" in a similar way. For example, we might say that the infant mortality rate of a country is a function of the quality of its health care, or that a country's national product is

TABLE 2.1 *Personal consumption expenditure in the United States, 1985–1991*

Year	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*
Personal consumption[1]	2,667.4	2,850.6	3,052.2	3,296.2	3,523.1	3,748.4	3,887.7

[1] In billions of dollars.

a function of the level of investment. In both these cases, it would be a major research task to obtain a formula that represents the function precisely.

One does not need a mathematical formula to convey the idea that one variable is a function of another: A table can also show the relationship. For instance, Table 2.1 shows the growth of annual total personal consumption expenditures, measured in current dollars, in the United States for the period 1985–1991. It is taken from figures in the *Economic Report of the President* dated January 1993. This table defines consumption expenditures as a function of the year. No allowance is made for inflation.

The dependence between two variables can also be illustrated by means of a graph or chart. Consider the following two examples.

In Fig. 2.1, we have drawn a curve that allegedly played an important role some years ago in the discussion of "supply side economics." It shows the presumed relationship between a country's income tax rate and its total income tax revenue. Obviously, if the income tax rate is 0%, then tax revenue is 0. However, if the tax rate is 100%, then tax revenue will also be (about) 0, because virtually no one is willing to work if his or her entire income is going to be confiscated. These ideas are obvious to virtually all competent economists (in cases like Problem 1 of Section 3.2). Nevertheless, a controversy was created by the American economist Arthur Laffer, who claimed to have drawn this curve on a restaurant napkin, and then later popularized its message with the public. Economists have hotly disputed what is the percentage rate a at which the government collects the maximum tax revenue.

FIGURE 2.1 The "Laffer curve," which relates tax revenue to tax rates.

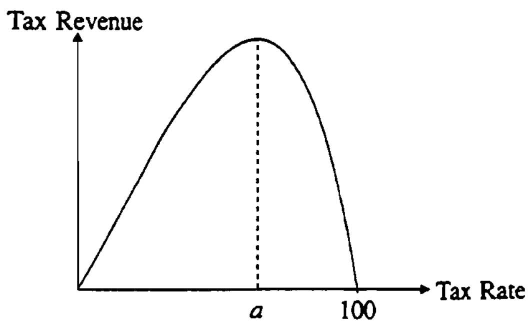


Figure 2.2 reproduces a postage stamp showing how Norway's gross national product grew during the first 100 years of the lifetime of its Central Bureau of Statistics.

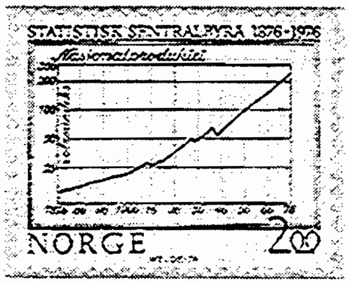


FIGURE 2.2 The national product of Norway (*volume index*) 1876–1976.

All of the relationships just discussed have one characteristic in common: A definite rule relates each value of one variable to a definite value of another variable.

Notice that in all of the examples, it is implicitly assumed that the variables are subject to certain constraints. For instance, in the temperature example, F cannot be less than -459.67, the absolute zero point (which corresponds to -273.15 degrees Centigrade). In Table 2.1, only the years between 1985 and 1991 are relevant.

2.2 Functions of One Real Variable

The examples we studied in the preceding section lead to the following general definition of a real valued function of one real variable:

> A **function** of a real variable x with **domain** D is a rule that assigns a unique real number to each number x in D. [2.1]

The word "rule" is used in a very broad sense. *Every* rule with the properties described in [2.1] is called a function, whether that rule is given by a formula, described in words, defined by a table, illustrated by a curve, or expressed by any other means.

Functions are often given letter names, such as f, g, F, or ϕ. If f is a function and x is a number in its domain D, then $f(x)$ denotes the number that the function f assigns to x. The symbol $f(x)$ is pronounced "f of x." It is important to note the difference between f, which is a symbol for the function (the rule), and $f(x)$, which denotes the value of f at x.

If f is a function, we sometimes let y denote the value of f at x, so

$$y = f(x) \qquad [*]$$

Then we call x the **independent variable**, or the **argument** of f, whereas y is called the **dependent variable**, because the value y (in general) depends on the value of x. In economics, x is often called the *exogenous* variable, whereas y is the *endogenous* variable.

A function is often defined by a particular formula of the type [*], such as $y = 8x^2 + 3x + 2$. The function is then the rule that assigns the number $8x^2 + 3x + 2$ to x.

Functional Notation

To become familiar with the relevant notation, it helps to look at some examples of functions that are defined by formulas.

Example 2.1

A function is defined for all numbers by the following rule:

Assign to any number the third power of that number. [1]

This function will assign $0^3 = 0$ to 0, $3^3 = 27$ to 3, $(-2)^3 = -8$ to -2, and $(1/4)^3 = 1/64$ to $1/4$. In general, it assigns the number x^3 to the number x. If we denote the function by f, then

$$f(x) = x^3 \qquad [2]$$

So $f(0) = 0^3 = 0$, $f(3) = 3^3 = 27$, $f(-2) = (-2)^3 = -8$, $f(1/4) = (1/4)^3 = 1/64$.

Substituting a for x in the formula for f gives $f(a) = a^3$, whereas

$$f(a+1) = (a+1)^3 = (a+1)(a+1)(a+1) = a^3 + 3a^2 + 3a + 1 \qquad [3]$$

Note: A common error is to presume that $f(a) = a^3$ implies $f(a+1) = a^3 + 1$. The error can be illustrated by looking at a simple interpretation of f. If a is the edge of a cube measured in meters, then $f(a) = a^3$ is the volume of the cube measured in cubic meters. Suppose that each edge of the cube has its length increased by 1 m. Then the volume of the new cube is $f(a+1) = (a+1)^3$ cubic meters. The number $a^3 + 1$ can be interpreted as the number obtained when the volume of a cube with edge a is increased by 1 m^3. In fact,

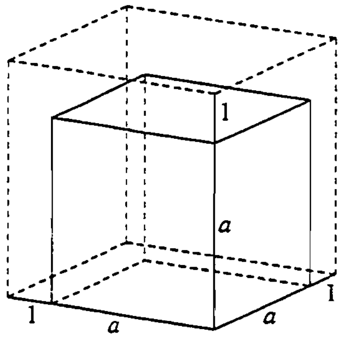


FIGURE 2.3 Volume $f(a+1)=(a+1)^3$.

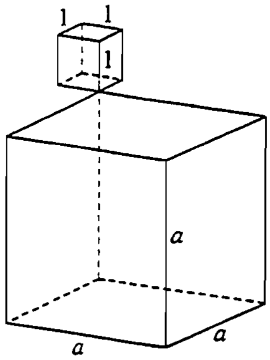


FIGURE 2.4 Volume a^3+1.

$f(a+1) = (a+1)^3$ is quite different from a^3+1, as illustrated in Figs. 2.3 and 2.4.

Example 2.2

The total dollar cost of producing x units of a product is given by

$$C(x) = 100x\sqrt{x} + 500$$

Find the cost of producing 16, 100, and a units. Suppose the firm produces a units; find the *increase* in the cost from producing one additional unit.[1]

Solution The cost of producing 16 units is found by substituting 16 for x in the formula for $C(x)$:

$$C(16) = 100 \cdot 16\sqrt{16} + 500 = 100 \cdot 16 \cdot 4 + 500 = 6900$$

Similarly,

$$C(100) = 100 \cdot 100 \cdot \sqrt{100} + 500 = 100{,}500$$
$$C(a) = 100a\sqrt{a} + 500$$

The cost of producing $a+1$ units is $C(a+1)$, so that the increase in cost is

$$\begin{aligned} C(a+1) - C(a) &= 100(a+1)\sqrt{a+1} + 500 - 100a\sqrt{a} - 500 \\ &= 100\left[(a+1)\sqrt{a+1} - a\sqrt{a}\right] \end{aligned}$$

[1] This is the concept that economists often call **marginal cost**. However, they should really call it **incremental cost**. In Section 4.3, we will explain the difference between the two.

So far we have used x to denote the independent variable, but we could just as well have used almost any other symbol. For example, all of the following formulas define exactly the same function (and hence we can set $f = g = \phi$):

$$f(x) = \frac{x^2 - 3}{x^4 + 1}, \qquad g(t) = \frac{t^2 - 3}{t^4 + 1}, \qquad \phi(\xi) = \frac{\xi^2 - 3}{\xi^4 + 1} \qquad [*]$$

For that matter, we could also express the function in [*] as follows:

$$f(\cdot) = \frac{(\cdot)^2 - 3}{(\cdot)^4 + 1}$$

Here it is understood that the dot between the parentheses can be replaced by an arbitrary number or an arbitrary letter or even another function (like $1/y$). Thus,

$$f(1) = \frac{(1)^2 - 3}{(1)^4 + 1} = -1, \qquad f(k) = \frac{k^2 - 3}{k^4 + 1}, \qquad \text{and} \qquad f(1/y) = \frac{(1/y)^2 - 3}{(1/y)^4 + 1}$$

In economic theory, we often study functions that depend on a number of parameters in addition to the independent variable. A typical example follows.

Example 2.3

Suppose that the cost of producing x units of a commodity is

$$C(x) = Ax\sqrt{x} + B \qquad (A \text{ and } B \text{ are positive constants}) \qquad [1]$$

Find the cost of producing 0, 10, and $x + h$ units.

Solution The cost of producing 0 units is

$$C(0) = A \cdot 0 \cdot \sqrt{0} + B = 0 + B = B$$

(Parameter B simply represents fixed costs. These are the costs that must be paid whether or not anything is actually produced, such as a taxi driver's annual license fee.) Similarly,

$$C(10) = A \cdot 10\sqrt{10} + B$$

Finally, substituting $x + h$ for x in (1) gives

$$C(x + h) = A(x + h)\sqrt{x + h} + B$$

The Domain and the Range

The definition of a function is incomplete unless its domain has been specified. The domain of the function f defined by $f(x) = x^3$ (see Example 2.1) is the set of all real numbers. In Example 2.2, where $C(x) = 100x\sqrt{x} + 500$ denotes the cost of producing x units of a product, the domain was not specified, but the natural domain is the set of numbers $0, 1, 2, \ldots, x_0$, where x_0 is the maximum number of items the firm can produce. If output x is a continuous variable, the natural domain is the closed interval $[0, x_0]$.

If a function is defined using an algebraic formula, we adopt the convention that the domain consists of all values of the independent variable for which the formula gives a meaningful value (unless another domain is explicitly mentioned).

Example 2.4

Find the domains of

(a) $$f(x) = \frac{1}{x+3}$$

(b) $$g(x) = \sqrt{2x+4}$$

Solution

(a) For $x = -3$, the formula reduces to the meaningless expression "1/0." For all other values of x, the formula makes $f(x)$ a well-defined number. Thus, the domain consists of all numbers $x \neq -3$.

(b) The expression $\sqrt{2x+4}$ is defined for all x such that $2x+4$ is nonnegative. Solving the inequality $2x + 4 \geq 0$ for x gives $x \geq -2$. Hence, the domain of g is the interval $[-2, \infty)$.

Let f be a function with domain D. The set of all values $f(x)$ that the function assumes is called the **range** of f. Often, we denote the domain of f by D_f, and the range by R_f. These concepts are illustrated in Fig. 2.5, using the idea of the graph of a function. (Graphs are discussed in the next section, but you probably have been exposed to them before.)

Alternatively, we can think of any function f as an engine operating so that if the number x in the domain is an input, the output is the number $f(x)$. (See Fig. 2.6.) The range of f is then all the numbers we get as output using all numbers x in the domain as inputs. If we try to use as an input a number not in the domain, the engine does not work, and there is no output.

Example 2.5

Show that the number 4 belongs to the range of the function defined by $g(x) = \sqrt{2x+4}$. Find the entire range of g. (Remember that $\sqrt{u}$ denotes the nonnegative square root of u.)

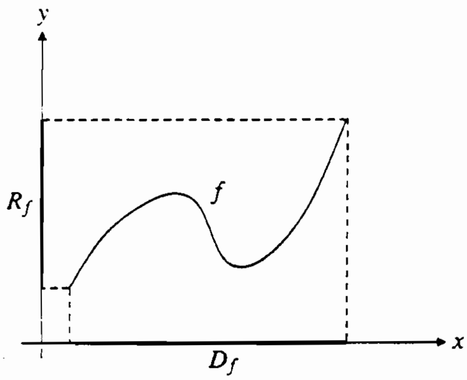


FIGURE 2.5 The domain and the range of f.

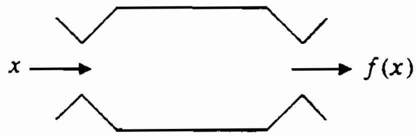
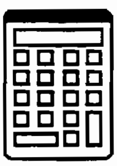


FIGURE 2.6 Function engine.

Solution To show that a number such as 4 is in the range of g, we must find a number x such that $g(x) = 4$. That is, we must solve the equation $\sqrt{2x+4} = 4$ for x. By squaring both sides of the equation, we get $2x + 4 = 4^2 = 16$, that is, $x = 6$. Because $g(6) = 4$, the number 4 does belong to the range R_g.

In order to determine the whole range of g, we must answer the question: As x runs through the whole of the interval $[-2, \infty)$, what are all the possible values of $\sqrt{2x+4}$? For $x = -2$, $\sqrt{2x+4} = 0$, and $\sqrt{2x+4}$ can never be negative. We claim that whatever number $y_0 \geq 0$ is chosen, there exists a number x_0 such that $\sqrt{2x_0+4} = y_0$. Squaring each side of this last equation gives $2x_0 + 4 = y_0^2$. Hence, $2x_0 = y_0^2 - 4$, which implies that $x_0 = \frac{1}{2}(y_0^2 - 4)$. Because $y_0^2 \geq 0$, we have $x_0 = \frac{1}{2}\left(y_0^2 - 4\right) \geq \frac{1}{2}(-4) = -2$. Hence, for every number $y_0 \geq 0$, there is a number $x_0 \geq -2$ such that $g(x_0) = y_0$. The range of g is, therefore, $[0, \infty)$.

Even if we have a function that is completely specified by a formula, including a specific domain, it is not always easy to find the range of the function. For example, without using the methods of differential calculus, it is not at all simple to find R_f when $f(x) = 3x^3 - 2x^2 - 12x - 3$ and $D_f = [-2, 3]$.



Many pocket calculators have some special functions built into them. For example, many have the $\sqrt{\ }$ function that, given a number x, assigns the square root of the number, $\sqrt{x}$. If we enter a nonnegative number such as 25, and press the square-root key, then the number 5 appears. If we enter -3, then the word "Error" is shown, which is the way the calculator tells us that $\sqrt{-3}$ is not defined.

The concept of a function is entirely abstract. In Example 2.2, we studied a function that finds the production cost $C(x)$ in dollars associated with the number of units x of a commodity. Here x and $C(x)$ are concrete, measurable quantities. On the other hand, the letter C, which is the name of the function, does not represent a physical quantity; rather, it represents the dependence of cost upon the number of units produced, a purely abstract concept.

Problems

1. Let $f(x) = x^2 + 1$.
a. Compute $f(0)$, $f(-1)$, $f(1/2)$, and $f(\sqrt{2})$.
b. For what x is it true that (i) $f(x) = f(-x)$? (ii) $f(x+1) = f(x)+f(1)$? (iii) $f(2x) = 2f(x)$?

2. Suppose $F(x) = 10$, for all x. Find $F(0)$, $F(-3)$, and $F(a+h) - F(a)$.

3. Let $f(t) = a^2 - (t-a)^2$ (a is a constant).
a. Compute $f(0)$, $f(a)$, $f(-a)$, and $f(2a)$.
b. Compute $3f(a) + f(-2a)$.

4. Let f be defined for all x by

$$f(x) = \frac{x}{1+x^2}$$

a. Compute $f(-1/10)$, $f(0)$, $f(1/\sqrt{2})$, $f(\sqrt{\pi})$, and $f(2)$.
b. Show that $f(x) = -f(-x)$ for all x, and that $f(1/x) = f(x)$, for $x \neq 0$.

5. The cost of producing x units of a commodity is given by

$$C(x) = 1000 + 300x + x^2$$

a. Compute $C(0)$, $C(100)$, and $C(101) - C(100)$.
b. Compute $C(x+1) - C(x)$, and explain in words the meaning of the difference.

6. Let $F(t) = \sqrt{t^2 - 2t + 4}$. Compute $F(0)$, $F(-3)$, and $F(t+1)$.

7. H. Schultz has estimated the demand for cotton in the United States for the period 1915–1919 to be $D(P) = 6.4 - 0.3P$ [with appropriate units for the price P and the quantity $D(P)$].
a. Find the demand if the price is 8, 10, and 10.22.
b. If the demand is 3.13, what is the price?

8. The cost of removing $p\%$ of the impurities in a lake is given by

$$b(p) = \frac{10p}{105 - p}$$

a. Find $b(0)$, $b(50)$, and $b(100)$.
b. What does $b(50+h) - b(50)$ mean? ($h \geq 0$.)

9. **a.** If $f(x) = 100x^2$, show that for all t, $f(tx) = t^2 f(x)$.
b. If $P(x) = x^{1/2}$, show that for all $t \geq 0$, $P(tx) = t^{1/2} P(x)$.

10. Only for special "additive" functions is it true that $f(a+b) = f(a) + f(b)$ for all a and b. Determine whether $f(2+1) = f(2) + f(1)$ for the following:

a. $f(x) = 2x^2$ b. $f(x) = -3x$ c. $f(x) = \sqrt{x}$

11. a. If $f(x) = Ax$, show that $f(a+b) = f(a) + f(b)$, for all a and b.
b. If $f(x) = 10^x$, show that $f(a+b) = f(a) \cdot f(b)$, for all natural numbers a and b.

12. A student claims that $(x+1)^2 = x^2 + 1$. Can you use a geometric argument to show that this is wrong?

13. Find the domains of the functions defined by the following equations:

a. $y = \sqrt{5-x}$ b. $y = \dfrac{2x-1}{x^2-x}$

c. $y = \sqrt{\dfrac{x-1}{(x-2)(x+3)}}$ d. $y = (x+1)^{1/2} + 1/(x-1)^{1/2}$

14. Consider the function f defined by the formula

$$f(x) = \frac{3x+6}{x-2}$$

a. Find the domain of f.
b. Show that the number 5 is in the range of f by finding a number x such that $(3x+6)/(x-2) = 5$.
c. Show that the number 3 is not in the range of f.

15. Find the domain and the range $g(x) = 1 - \sqrt{x+2}$.

16. Let $f(x) = |x|$. Which of the the following rules are valid for all possible pairs of numbers x and y?

a. $f(x+y) = f(x) + f(y)$ b. $f(x+y) \le f(x) + f(y)$
c. $f(xy) = f(x) \cdot f(y)$ d. $f(2x) = 2f(x)$
e. $f(-2x) = -2f(x)$ f. $f(x) = \sqrt{x^2}$
g. $f(-2x) = 2f(x)$ h. $|f(x) - f(y)| \le |x - y|$

17. Let

$$f(x) = \frac{ax+b}{cx-a}$$

where a, b, and c are constants, and $c \neq 0$. Assuming that $x \neq a/c$, show that

$$f\left(\frac{ax+b}{cx-a}\right) = x$$

2.3 Graphs

Three examples of equations in two variables x and y are

$$y = 2x - 1, \qquad x^2 + y^2 = 16, \qquad x\sqrt{y} = 2 \qquad [*]$$

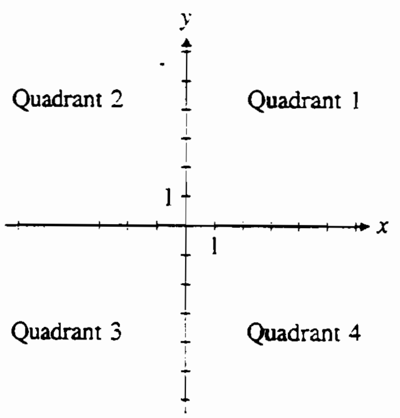
In this section, we shall explain how *any* equation in two variables can be represented by a curve (a graph) in a coordinate system. In particular, any function given by an equation $y = f(x)$ has such a representation, that helps us to visualize the equation or the function. This is because the shape of the graph reflects the properties of the equation or the function.

A Coordinate System in the Plane

In Section 1.4, we claimed that real numbers can be represented by a number line. Analogously, every *pair* of real numbers can be represented by a point in a plane. Draw two perpendicular lines, called respectively the *x-axis* (or the *horizontal axis*) and the *y-axis* (or the *vertical axis*). The intersection point O is called the *origin*. We measure the real numbers along each of these lines, as shown in Fig. 2.7. Often, we measure the numbers on the two axes so that the length on the x-axis that represents the distance between x and $x+1$ is the same length as that along the y-axis that represents the distance between y and $y+1$. But this does not have to be the case.

Figure 2.7 illustrates a **rectangular**, or a **Cartesian, coordinate system**, that we call the **xy-plane**. The coordinate axes separate the plane into four quadrants, which can be numbered as in Fig. 2.7. Any point P in the plane can be represented by a pair (a, b) of real numbers. These can be found by dropping perpendiculars onto the axes. The point represented by (a, b) lies at the intersection of the vertical straight line $x = a$ with the horizontal straight line $y = b$. Conversely, any pair of real numbers represents a unique point in the plane. For example, in Fig. 2.8, the ordered pair $(3, 4)$ corresponds to the point P that lies at the intersection of $x = 3$

FIGURE 2.7 A coordinate system.



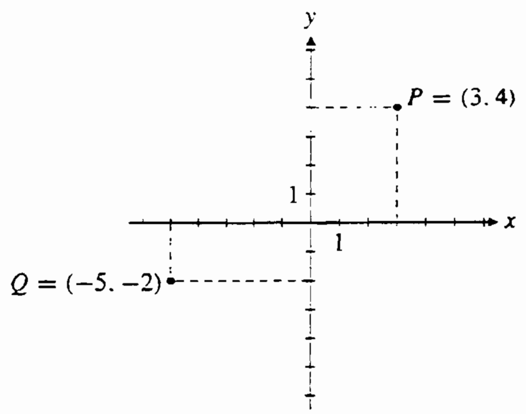


FIGURE 2.8 The points (3, 4) and (−5, −2).

with $y = 4$. Thus, P lies 3 units to the right of the y-axis and 4 units above the x-axis. We call $(3, 4)$ the **coordinates** of P. Similarly, Q lies 5 units to the left of the y-axis and 2 units below the x-axis, so the coordinates of Q are $(-5, -2)$.

Note that we call (a, b) an **ordered pair**, because the order of the two numbers in the pair is important. For instance, $(3, 4)$ and $(4, 3)$ represent two different points.

Example 2.6

Draw coordinate systems and indicate the coordinates (x, y) that satisfy each of the following three conditions:

(a) $x = 3$
(b) $x \geq 0$ and $y \geq 0$
(c) $-2 \leq x \leq 1$ and $-2 \leq y \leq 3$

Solution

(a) See Fig. 2.9, which represents a straight line.
(b) See Fig. 2.10, which represents the first quadrant.
(c) See Fig. 2.11, which represents a rectangle.

FIGURE 2.9 A straight line.

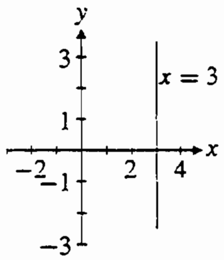


FIGURE 2.10 The first quadrant.

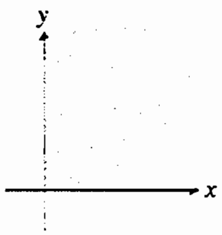
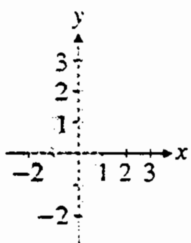


FIGURE 2.11 A rectangle.



Graphs of Equations in Two Variables

A solution of an equation in two variables x and y is a pair (a, b) that satisfies the equation when we substitute a for x and b for y. The **solution set** of the equation is the set of all possible solutions. If we plot all the ordered pairs of the solution set in a coordinate system, we obtain a curve that is called the **graph** of the equation.

Example 2.7

Find some numerical solutions for each of the equations $y = 2x - 1$, $x^2 + y^2 = 16$, and $x\sqrt{y} = 2$, and try to sketch the graphs.

Solution For $y = 2x - 1$, point $(0, -1)$ is a solution, because if $x = 0$, then $y = 2 \cdot 0 - 1 = -1$. Other solutions are $(1, 1)$, $(3, 5)$, and $(-1, -3)$. In Fig. 2.12, we have plotted the four solutions, and they all appear to lie on a straight line. There exist infinitely many other solutions, so we can never write them all down.

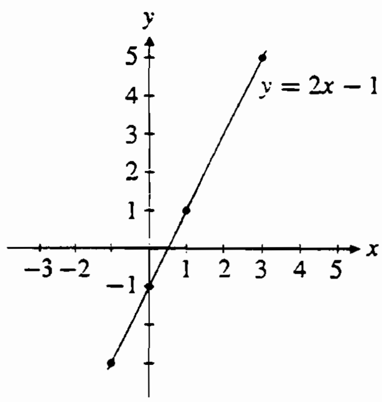


FIGURE 2.12 $y = 2x - 1$.

For $x^2 + y^2 = 16$, point $(4, 0)$ is a solution. Some other solutions are shown in Table 2.2.

TABLE 2.2 *Solutions of* $x^2 + y^2 = 16$

x	-4	-3	-1	0	1	3	4
y	0	$\pm\sqrt{7}$	$\pm\sqrt{15}$	± 4	$\pm\sqrt{15}$	$\pm\sqrt{7}$	0

Figure 2.13 shows the plot of the points in the table, and the graph appears to be a circle.

From $x\sqrt{y} = 2$, we obtain $y = 4/x^2$, and it is easy to fill in Table 2.3. The graph is shown in Fig. 2.14.

Note: When plotting the graph of an equation such as $x^2 + y^2 = 16$, we must try to find a sufficient number of solution pairs (x, y), otherwise we might miss some

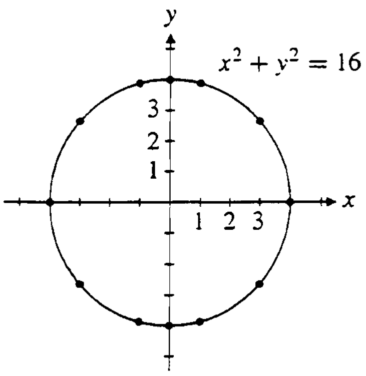


FIGURE 2.13 $x^2 + y^2 = 16$.

TABLE 2.3 *Solutions of* $x\sqrt{y} = 2$

x	1	2	4	6
y	4	1	1/4	1/9

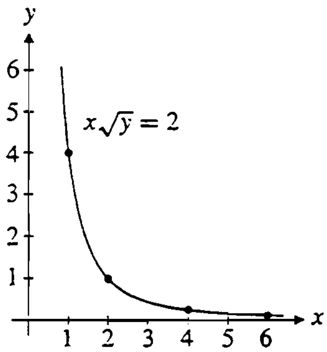


FIGURE 2.14 $x\sqrt{y} = 2$.

important features of the graph. Actually, by merely plotting a finite set of points, we can never be entirely sure that there are no wiggles or bumps we have missed. We shall see in what follows that the graph of the equation $x^2 + y^2 = 16$ really is a circle. For more complicated equations, we have to use differential calculus to decide how many bumps and wiggles there are.

The Distance Between Two Points in the Plane

Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be the two points shown in Fig. 2.15. By Pythagoras' theorem, the distance d between these points satisfies the equation

$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$. Therefore, note that because $(x_1 - x_2)^2 = (x_2 - x_1)^2$ and $(y_1 - y_2)^2 = (y_2 - y_1)^2$, it does not make any difference which point is P_1 and which is P_2.

The Distance Formula

The distance between points (x_1, y_1) and (x_2, y_2) is

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad [2.2]$$

To prove the distance formula, we considered two points in the first quadrant. It turns out that the same formula is valid regardless of where the two points, P_1 and P_2, lie.

Example 2.8

Find the distance d between points $P_1 = (-4, 3)$ and $P_2 = (5, -1)$. (See Fig. 2.16.)

Solution Using [2.2] with $x_1 = -4$, $y_1 = 3$, and $x_2 = 5$, $y_2 = -1$, we have

$$d = \sqrt{(-4-5)^2 + (3-(-1))^2}$$
$$= \sqrt{(-9)^2 + 4^2} = \sqrt{81+16} = \sqrt{97} \approx 9.85$$

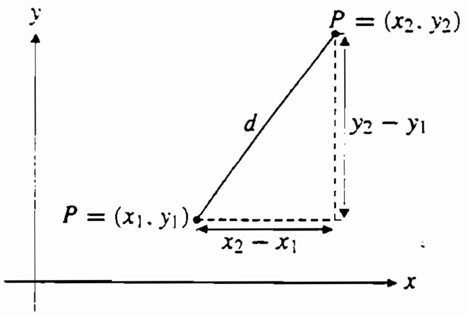


FIGURE 2.15

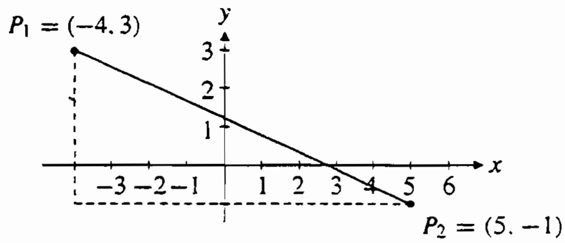


FIGURE 2.16

Circles

Let (a, b) be a point in the plane. *The circle with radius r and center at (a, b) is the set of all points (x, y) whose distance from (a, b) is equal to r.* Considering Fig. 2.17 and using the distance formula gives $\sqrt{(x-a)^2 + (y-b)^2} = r$. Squaring each side yields

The Equation of a Circle

The equation of a circle with center at (a, b) and radius r is

$$(x - a)^2 + (y - b)^2 = r^2 \qquad [2.3]$$

Note that if we let $a = b = 0$ and $r = 4$, then [2.3] reduces to $x^2 + y^2 = 16$. This is the equation of a circle with center at (0, 0) and radius 4, as shown in Fig. 2.13.

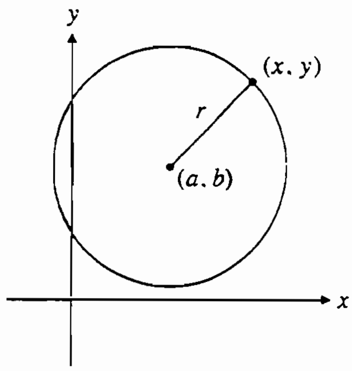


FIGURE 2.17 Circle with center (*a*, *b*) and radius *r*.

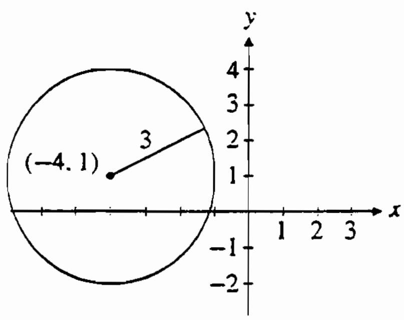
Example 2.9

Find the equation of the circle with center (−4, 1) and radius 3.

Solution Here $a = -4$, $b = 1$, and $r = 3$ (see Fig. 2.18). So the general formula in [2.3] becomes the specific equation

$$(x + 4)^2 + (y - 1)^2 = 9 \qquad [1]$$

FIGURE 2.18 Circle with center (−4, 1) and radius 3.



Expanding the squares in [1] gives

$$x^2 + 8x + 16 + y^2 - 2y + 1 = 9 \qquad [2]$$

This can be written as

$$x^2 + y^2 + 8x - 2y + 8 = 0 \qquad [3]$$

Note: The equation of the circle given in [3] has the disadvantage that we cannot immediately read off its center and radius. If we are given equation [3], we can "argue backwards" in order to deduce [1] via [2]. We then say that we have "completed the squares," which is actually one of the oldest tricks in mathematics. (See Section A.8, Appendix A.) The method is illustrated in Problem 9.

Problems

1. Draw a Cartesian coordinate system and plot the points $(2, 3)$, $(-3, 2)$, $(-3/2, 1/4)$, $(4, 0)$, and $(0, 4)$.

2. Sketch the six sets of points (x, y) satisfying the following conditions:

a. $y = 4$ **b.** $x < 0$ **c.** $x \geq 1$ and $y \geq 2$

d. $|x| = 2$ **e.** $y = x$ **f.** $y \geq x$

3. Sketch the graphs of each of the following equations:

a. $y = 4x - 3$ **b.** $xy = 1$ **c.** $y^2 = x$

4. Try to sketch the graphs of each of the following equations:

a. $x^2 + 2y^2 = 6$ **b.** $y + \sqrt{x - 1} = 0$ **c.** $y^2 - x^2 = 1$

5. Find the distance between each pair of points:

a. $(1, 3)$ and $(2, 4)$ **b.** $(-1, 2)$ and $(-3, -3)$

c. $(3/2, -2)$ and $(-5, 1)$ **d.** (x, y) and $(2x, y + 3)$

e. (a, b) and $(-a, b)$ **f.** $(a, 3)$ and $(2 + a, 5)$

6. The distance between $(2, 4)$ and $(5, y)$ is $\sqrt{13}$. Find y. (Explain geometrically why there must be two values of y. What would happen if the distance were 2?)

7. Find the approximate distance between each pair of points:

a. $(3.998, 2.114)$ and $(1.130, -2.416)$ **b.** $(\pi, 2\pi)$ and $(-\pi, 1)$

8. Find the equations of the following circles:

a. Center at $(2, 3)$ and radius 4.

b. Center at $(2, 5)$ and passes through $(-1, 3)$.

9. We can show that the graph of $x^2 + y^2 + 8x - 2y + 8 = 0$ is a circle by arguing like this: First, rearrange the equation to read $(x^2 + 8x \ldots) +$

$(y^2 - 2y \ldots) = -8$. Completing the two squares gives $(x^2 + 8x + 4^2) + (y^2 - 2y + (-1)^2) = -8 + 4^2 + (-1)^2 = 9$. Thus, the equation becomes $(x+4)^2 + (y-1)^2 = 9$, whose graph is a circle with center $(-4, 1)$ and radius $\sqrt{9} = 3$. Use this method to find the center and the radius of the two circles with equations:

a. $x^2 + y^2 + 10x - 6y + 30 = 0$ **b.** $3x^2 + 3y^2 + 18x - 24y = -39$

10. Point P moves in the plane so that it is always equidistant from each of the points $A = (3, 2)$ and $B = (5, -4)$. Find a simple equation that the coordinates (x, y) of P must satisfy. Illustrate the problem and its solution geometrically. (*Hint:* Compute the square of the distance from P to A and to B, respectively.)

11. Prove that if the distance from a point (x, y) to the point $(-2, 0)$ is twice the distance from (x, y) to $(4, 0)$, then (x, y) must lie on the circle with center $(6, 0)$ and radius 4.

Harder Problems

12. Try to sketch the graph of the equation $\sqrt{x} + \sqrt{y} = 1$.

13. A firm has two plants A and B located 60 kilometers apart at the two points $(0, 0)$ and $(60, 0)$. See Fig. 2.19. It supplies one identical product priced at $\$p$ per unit. Shipping costs per kilometer per unit are \$10 from A and \$5 from B. An arbitrary purchaser is located at point (x, y).

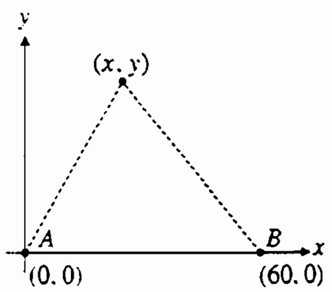
a. Give economic interpretations for the expressions:

$$p + 10\sqrt{x^2 + y^2} \quad \text{and} \quad p + 5\sqrt{(x-60)^2 + y^2}$$

b. Find the equation for the curve that separates the markets of the two firms, assuming that customers buy from the firm for which total costs are lower.

14. Generalize Problem 13 to the case where $A = (0, 0)$ and $B = (a, 0)$, and assume that shipping costs per kilometer are r and s dollars, respectively. Show that the curve separating the markets is a circle, and find its center and radius.

FIGURE 2.19



15. Show that the graph of

$$x^2 + y^2 + Ax + By + C = 0 \qquad (A,\ B,\text{ and } C \text{ are constants})$$

is a circle if $A^2 + B^2 > 4C$. Find its center and radius. (See Problem 9.) What happens if $A^2 + B^2 \leq 4C$?

2.4 Graphs of Functions

The **graph** of a function f is the set of all points $(x, f(x))$, where x belongs to the domain of f. This is simply the graph of the equation $y = f(x)$. Typical examples of graphs of functions are given in Figs. 2.20 and 2.21.

In Fig. 2.20, we show the graph of $f(x) = x^2 - 3x$. It is found by computing points $(x, f(x))$ on the graph and then drawing a smooth curve through the points.

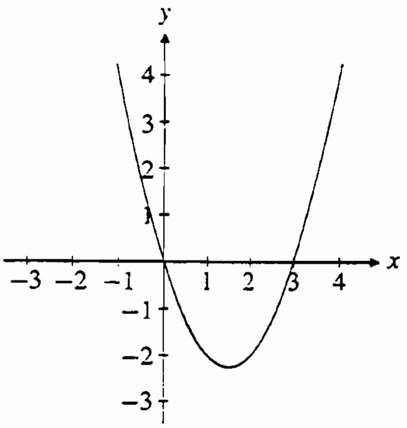
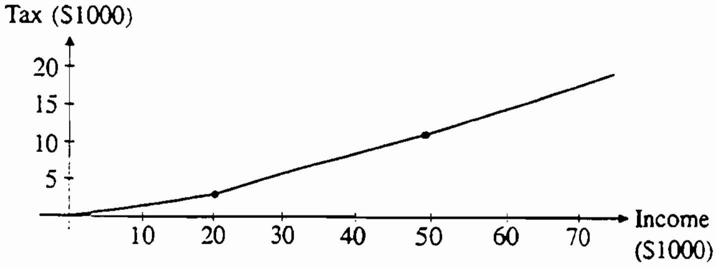


FIGURE 2.20 The graph of $f(x) = x^2 - 3x$.

The function whose graph is shown in Fig. 2.21 is of a type often encountered in economics. It is defined by different formulas on different intervals.

Example 2.10 (U.S. Federal Income Tax (1991) for Single Persons)
In Fig. 2.21, we show the graph of this income tax function. Income up to

FIGURE 2.21 U.S. federal income tax.



\$20,250 was taxed at 15%, income between \$20,251 and \$49,300 was taxed at 28%, and income above \$49,300 was taxed at 31%.

Graphs of different functions can have innumerable different shapes. However, not all curves in the plane are graphs of functions. A function assigns to each point x in the domain only one y-value. *The graph of a function therefore has the property that a vertical line through any point on the x-axis has at most one point of intersection with the graph.* This simple *vertical line test* is illustrated in Figs. 2.22 and 2.23.

The graph of the circle $x^2 + y^2 = 16$, as shown in Fig. 2.13, is a typical example of a graph that does *not* represent a function, because it does not pass the vertical line test. The vertical line $x = a$ for any a with $-4 < a < 4$ intersects the circle at *two* points. When we solve the equation $x^2 + y^2 = 16$ for y, we obtain $y = \pm\sqrt{16 - x^2}$. Note that the upper semicircle alone is the graph of the function $y = \sqrt{16 - x^2}$ and the lower semicircle is the graph of the function $y = -\sqrt{16 - x^2}$. Both these functions are defined on the interval $[-4, 4]$.

Choice of Units

A function of one variable is a rule assigning numbers in its range to numbers in its domain. When we describe an empirical relationship by means of a function, we must first choose the units of measurement. For instance, we might measure

FIGURE 2.22 This graph represents a function.

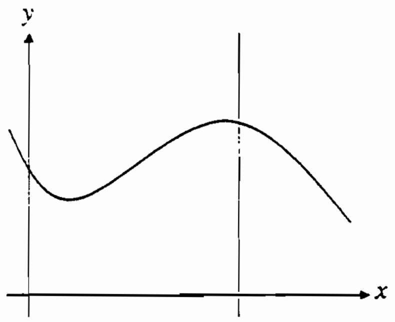
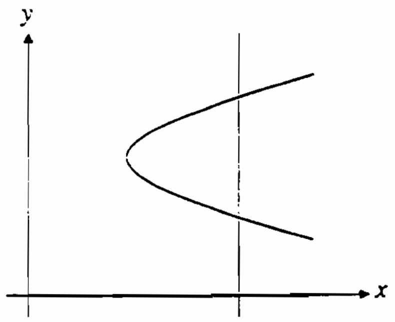


FIGURE 2.23 This graph does not represent a function.



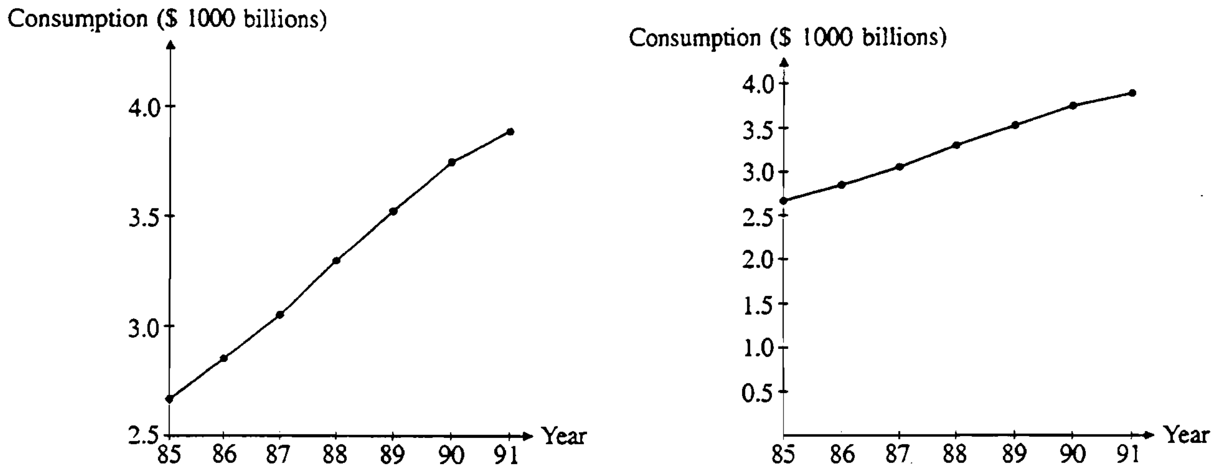


FIGURE 2.24 Graphical representations of the function defined in Table 2.1 with different units of measurement.

time in years, days, or weeks. We might measure money in dollars, yen, or francs. The choice we make may influence the visual impression conveyed by the graph of the function.

Figure 2.24 illustrates a standard trick that is often used to influence people's impressions of empirical relationships. In both diagrams, time is measured in years and consumption in billions of dollars. They both graph the same function. (Which graph would you think the Republicans in the United States might prefer for their advertising, and which is more to the liking of the Democrats?)

Shifting Graphs

Given the graph of a function f, it is sometimes useful to know how to find the graphs of the related functions:

$$f(x)+c, \qquad f(x+c), \qquad -f(x), \qquad \text{and} \qquad f(-x) \qquad [2.4]$$

Problem 3 of this section asks you to study these graphs in general. Here we consider a simple economic example.

Example 2.11

Suppose a person earning y (dollars) in a given year pays $f(y)$ (dollars) in income tax. It is decided to reduce taxes. One proposal is to allow all individuals to deduct d dollars from their taxable income before tax is calculated. An alternative proposal involves calculating income tax on the full amount of taxable income and then allowing each person a "tax credit" that deducts d dollars from the total tax due. Illustrate graphically the two proposals for a "normal" tax function f, and mark off the the income y^*, where the two proposals give the same tax.

Solution Figure 2.25 illustrates the solution. First, draw the graph of the original tax function, $T = f(y)$. If taxable income is y and the deduction is

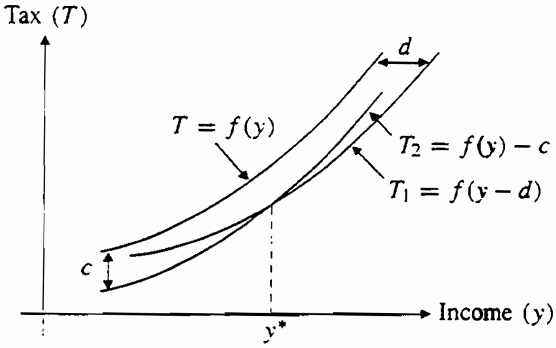


FIGURE 2.25 The graphs of $T_1 = f(y - c)$ and $T_2 = f(y) - d$.

d, then $y-d$ is the reduced taxable income, and so the tax liability is $f(y-d)$. By shifting the graph of the original tax function d units to the right, we obtain the graph of $T_1 = f(y-d)$.[2] The graph of $T_2 = f(y)-c$ is obtained by lowering the graph of $T = f(y)$ by c units. The income y^* that gives the same tax under the two different schemes is the value of y satisfying the equation

$$f(y-d) = f(y) - c$$

This value of y is marked y^* in the figure.

Problems

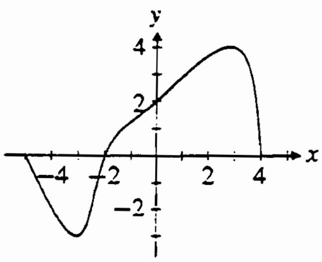
1. Determine the domain on which each of the following equations defines y as a function of x:

   a. $y = x + 2$ **b.** $y = \pm\sqrt{x}$ **c.** $y = x^4$ **d.** $y^4 = x$

   e. $x^2 - y^2 = 1$ **f.** $y = \dfrac{x}{x-3}$ **g.** $y^3 = x$ **h.** $x^3 + y^3 = 1$

2. The graph of the function f is given in Fig. 2.26.

FIGURE 2.26



[2]As an example: $y = x^2$ is a parabola, whereas $y = (x-1)^2$ is a parabola obtained by shifting the first parabola 1 unit to the right.

a. Find $f(-5)$, $f(-3)$, $f(-2)$, $f(0)$, $f(3)$, and $f(4)$ by examining the graph.
b. Find the domain and the range of f.

3. Explain how to get the graphs of the four functions defined by [2.4] based on the graph of $y = f(x)$.

4. Use the rules obtained in Problem 3 to sketch the graphs of the following:

a. $y = x^2 + 1$ b. $y = (x + 3)^2$
c. $y = 3 - (x + 1)^2$ d. $y = 2 - (x + 2)^{-2}$

2.5 Linear Functions

A *linear relationship* between the variables x and y takes the form

$$y = ax + b \qquad (a \text{ and } b \text{ are constants})$$

The graph of the equation is a straight line. If we let f denote the function that assigns y to x, then $f(x) = ax + b$, and f is called a **linear** function.[3] The number a is called the **slope** of the function and of the line. Take an arbitrary value of x. Then $f(x+1)-f(x) = a(x+1)+b-ax-b = a$. This shows that the slope a measures the change in the value of the function when x increases by 1 unit.

If the slope a is positive, the line slants upward to the right, and the larger the value of a, the steeper is the line. On the other hand, if a is negative, then the line slants downward to the right, and the absolute value of a measures the steepness of the line. For example, when $a = -3$, the steepness is 3. In the special case when $a = 0$, then $y = ax + b = b$ for all x, and the line is parallel to the x-axis. The three cases are illustrated in Figs. 2.27 to 2.29. If $x = 0$, then $y = ax+b = b$, and b is called the **y-intercept** (or often just the intercept).

FIGURE 2.27

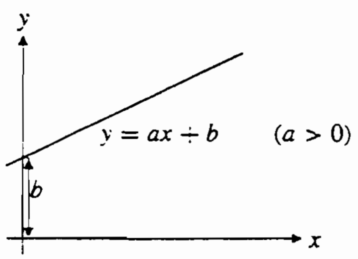
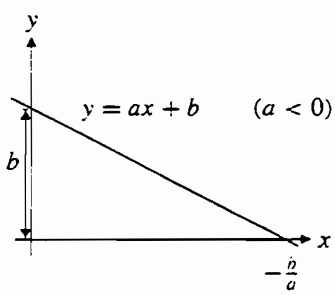


FIGURE 2.28



[3]Actually, mathematicians usually reserve the term "linear" for the functions defined by $y = ax$ (with the y-intercept $b = 0$). They call $y = ax + b$ with $b \neq 0$ an "affine" function. Most economists call $f(x) = ax + b$ a linear function.

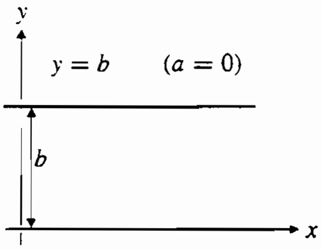


FIGURE 2.29

Example 2.12

Find and interpret the slopes of the following straight lines:

(a) $$C = 55.73x + 182{,}100{,}000$$

which is the estimated cost function for the U.S. Steel Corp. over the period 1917–1938 (C is the total cost in dollars per year, and x is the production of steel in tons per year).

(b) $$q = -0.15p + 0.14$$

which is the estimated annual demand function for rice in India for the period 1949–1964 (p is the price, and q is consumption per person).

Solution

(a) The slope is 55.73, which means that if production increases by 1 ton, then the cost *increases* by \$55.73.

(b) The slope is -0.15, which tells us that if the price increases by 1 unit, then the quantity demanded *decreases* by 0.15 unit.

Finding the Slope

Consider an arbitrary, nonvertical (straight) line in the plane. Pick two distinct points on the line, $P = (x_1, y_1)$ and $Q = (x_2, y_2)$, as shown in Fig. 2.30. Because the line is not vertical and because P and Q are distinct, $x_1 \neq x_2$. The slope of the line is the ratio $(y_2 - y_1)/(x_2 - x_1)$. If we denote the slope by a, then the following holds.

The **slope of a straight line** l is

$$a = \frac{y_2 - y_1}{x_2 - x_1}, \qquad x_1 \neq x_2 \qquad [2.5]$$

where (x_1, y_1) and (x_2, y_2) are any two distinct points on l.

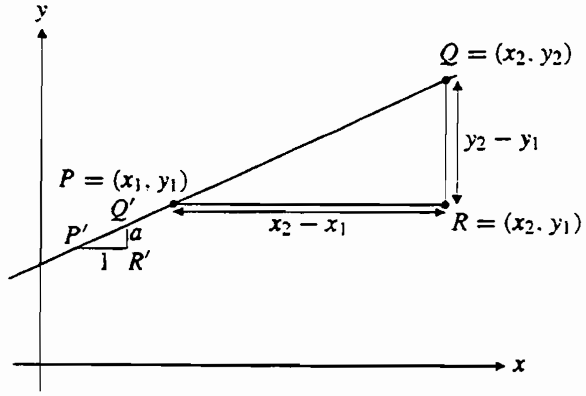


FIGURE 2.30 Slope $a = (y_2 - y_1)/(x_2 - x_1)$.

Multiplying both the numerator and the denominator of the fraction in [2.5] by -1, we obtain the fraction $(y_1 - y_2)/(x_1 - x_2)$. This shows that it does not make any difference which point is P and which is Q. Moreover, using the properties of similar triangles, we see by studying the two triangles PQR and $P'Q'R'$ in Fig. 2.30 that the number a in [2.5] is equal to the change in the value of y when x increases by 1 unit.

Example 2.13

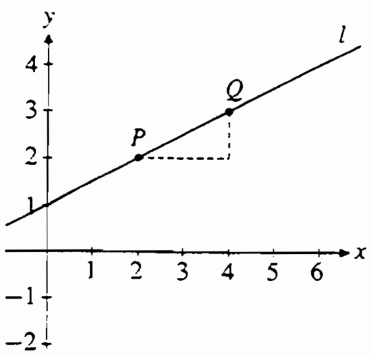
Determine the slopes of the three straight lines l, m, and n in Figs. 2.31–2.33 using [2.5].

Solution The lines l, m, and n all pass through $P = (2, 2)$. In Fig. 2.31, Q is $(4, 3)$. In Fig. 2.32, Q is $(1, -2)$. And in Fig. 2.33, Q is $(5, -1)$. Therefore, the respective slopes of the lines l, m, and n are

$$a_l = \frac{3-2}{4-2} = \frac{1}{2}, \qquad a_m = \frac{-2-2}{1-2} = 4, \qquad a_n = \frac{-1-2}{5-2} = -1$$

The following example illustrates a problem that is important in differential calculus, as will be seen in Chapter 4.

FIGURE 2.31 The line l.



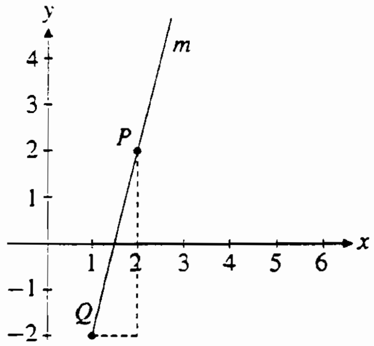


FIGURE 2.32 The line *m*.

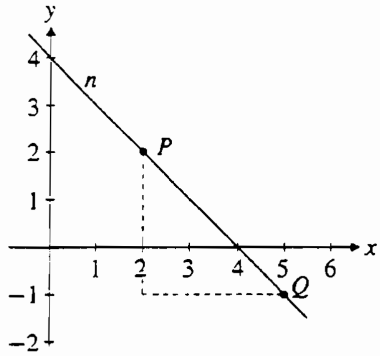


FIGURE 2.33 The line *n*.

Example 2.14

Find an expression for the slope of the line through the two points (x_0, x_0^2) and $(x_0 + h, (x_0 + h)^2)$, where $h \neq 0$.

Solution Apply formula [2.5] with $(x_1, y_1) = (x_0, x_0^2)$ and $(x_2, y_2) = (x_0 + h, (x_0 + h)^2)$ to obtain

$$a = \frac{(x_0 + h)^2 - x_0^2}{x_0 + h - x_0} = \frac{x_0^2 + 2x_0h + h^2 - x_0^2}{h} = \frac{2x_0h + h^2}{h} = 2x_0 + h$$

The Point–Slope and Point–Point Formulas

Let us find the equation of a straight line l passing through point $P = (x_1, y_1)$ with slope a. If (x, y) is any other point on the line, the slope a is given by formula [2.5]:

$$\frac{y - y_1}{x - x_1} = a$$

Multiplying each side by $x - x_1$, we obtain $y - y_1 = a(x - x_1)$. Hence:

Point–Slope Formula of a Straight Line

The equation of the straight line passing through (x_1, y_1) with slope a is

$$y - y_1 = a(x - x_1) \qquad [2.6]$$

Note that when using equation [2.6], x_1 and y_1 are fixed numbers giving the coordinates of the fixed point. On the other hand, x and y are variables denoting an arbitrary point on the line.

Example 2.15

Find the equation of the line through $(-2, 3)$ with slope -4. Then find the point at which this line intersects the x-axis.

Solution The point–slope formula with $(x_1, y_1) = (-2, 3)$ and $a = -4$ gives

$$y - 3 = (-4)[x - (-2)] \quad \text{or} \quad y - 3 = -4(x + 2) \qquad [1]$$

The line intersects the x-axis at the point where $y = 0$, that is, where $0 - 3 = -4(x + 2)$ or $-3 = -4x - 8$. Solving for x, we get $x = -5/4$, so the point of intersection with the x-axis is $(-5/4, 0)$.

Often we need to find the equation of a straight line that passes through two given points. Combining [2.5] with [2.6], we obtain the following formula:

Point–Point Formula of a Straight Line

The equation of the straight line passing through (x_1, y_1) and (x_2, y_2), where $x_1 \neq x_2$, is obtained as follows:

1. Compute the slope of the line:

$$a = \frac{y_2 - y_1}{x_2 - x_1}$$

2. Substitute the expression for a into the point–slope formula $y - y_1 = a(x - x_1)$. The result is

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \qquad [2.7]$$

Example 2.16
Find the equation of the line passing through $(-1, 3)$ and $(5, -2)$.

Solution Let $(x_1, y_1) = (-1, 3)$ and $(x_2, y_2) = (5, -2)$. Then the point–point formula gives

$$y - 3 = \frac{-2 - 3}{5 - (-1)}[x - (-1)] \quad \text{or} \quad y - 3 = -\frac{5}{6}(x + 1)$$

$$\text{or} \quad 5x + 6y = 13$$

Linear Models

Linear relations occur frequently in applied models. The relationship between the Celsius and Fahrenheit temperature scales is an example of an exact linear relation between two variables. Most of the linear models in economics are approximations to more complicated models. Two typical relations are those shown in Example 2.12. Statistical methods have been devised to construct linear functions that approximate the actual data as closely as possible. Let us consider a very naïve attempt to construct a linear model based on some data.

Example 2.17

In a United Nations report, the European population in 1960 was estimated as 641 million, and in 1970 the estimate was 705 million. Use these estimates to construct a linear function of t that approximates the population in Europe (in millions), where t is the number of years from 1960 ($t = 0$ is 1960, $t = 1$ is 1961, and so on). Make use of the equation to estimate the population in 1975 and in 2000. How do you estimate the population in 1930 on the basis of this linear relationship?

Solution If P denotes the population in millions, we construct an equation of the form $P = at + b$. We know that the graph must pass through the points $(t_1, P_1) = (0, 641)$ and $(t_2, P_2) = (10, 705)$. Using the formula in [2.7], replacing x and y with t and P, respectively, we obtain

$$P - 641 = \frac{705 - 641}{10 - 0}(t - 0) = \frac{64}{10}t$$

or

$$P = 6.4t + 641 \qquad [1]$$

In Table 2.4, we have compared our estimates with UN forecasts. Note that because $t = 0$ corresponds to 1960, $t = -30$ will correspond to 1930.

Note that the slope of line [1] is 6.4. This means that if the European population had developed according to [1], then the annual increase in the population would have been constant and equal to 6.4 million.

Actually, Europe's population grew unusually fast during the 1960s. Of course, it grew unusually slowly when millions died during the war years 1939–1945. We see that formula [1] does not give very good results compared to the UN estimates. (For a better way to model population growth, see Example 3.12 in Section 3.5.)

Example 2.18 (The Consumption Function)

In Keynesian macroeconomic theory, total consumption expenditure on goods and services, C, is assumed to be a function of national income

TABLE 2.4 *Population estimates for Europe*

Year	*1930*	*1975*	*2000*
t	−30	15	40
UN estimates	573	728	854
Formula [1] gives	449	737	897

Y, with

$$C = f(Y) \tag{2.8}$$

In many models, the consumption function is assumed to be linear, so that

$$C = a + bY$$

The slope b is called the **marginal propensity to consume**. If C and Y are measured in millions of dollars, the number b tells us by how many millions of dollars consumption increases if the national income increases by 1 million dollars. The number b will usually lie between 0 and 1.

In a study of the U.S. economy for the period 1929–1941, T. Haavelmo found the following consumption function:

$$C = 95.05 + 0.712\,Y$$

Here, the marginal propensity to consume is equal to 0.712.

Example 2.19

Some other economic examples of linear functions are the following demand and supply schedules:

$$D = a - bP$$
$$S = \alpha + \beta P$$

Here a and b (both positive) are parameters of the demand function D, while α and β (both positive) are parameters of the supply function. Such functions play an important role in quantitative economics. It is often the case that the market for a particular commodity, such as a specific brand of $3\frac{1}{2}$-inch computer diskettes, can be represented approximately by linear demand and supply functions. The equilibrium price P^e must equate demand and supply, so that $D = S$ at $P = P^e$. Thus,

$$a - bP^e = \alpha + \beta P^e$$

Adding $bP^e - \alpha$ to each side gives

$$a - bP^e + bP^e - \alpha = \alpha + \beta P^e + bP^e - \alpha$$

Thus, $a - \alpha = (\beta + b)P^e$. The corresponding equilibrium quantity is $Q^e = a - bP^e$. Hence, equilibrium occurs when

$$P^e = \frac{a - \alpha}{\beta + b}, \qquad Q^e = a - b\frac{a - \alpha}{\beta + b} = \frac{a\beta + \alpha b}{\beta + b}$$

If the four parameters, a, b, α, and β, were all known, then the model would be complete and the equilibrium price and quantity could be predicted. Suppose that there is a later shift in the supply or demand function—for instance, suppose supply increases so that S becomes $\tilde{\alpha} + \beta P$, where $\tilde{\alpha} > \alpha$. Then we could predict that the new equilibrium price and quantity would be

$$\tilde{P}^e = \frac{a - \tilde{\alpha}}{\beta + b}, \qquad \tilde{Q}^e = \frac{a\beta + \tilde{\alpha}b}{\beta + b}$$

Here $\tilde{P}^e$ is less than P^e, whereas $\tilde{Q}^e$ is greater than Q^e. In fact,

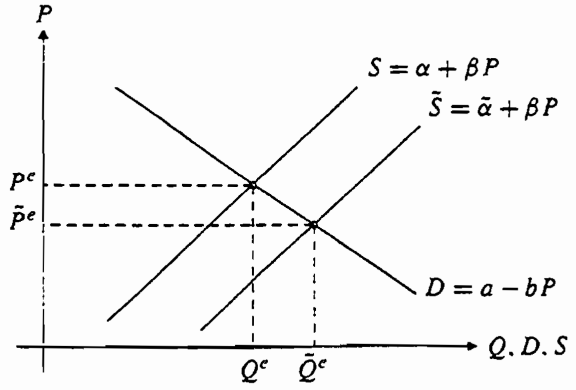
$$\tilde{P}^e - P^e = \frac{\alpha - \tilde{\alpha}}{\beta + b} \quad \text{and} \quad \tilde{Q}^e - Q^e = \frac{(\tilde{\alpha} - \alpha)b}{\beta + b} = -b(\tilde{P}^e - P^e)$$

This is in accord with Fig. 2.34. The rightward shift in the supply curve from S to $\tilde{S}$ moves the equilibrium down and to the right along the unchanged demand curve.

A peculiarity of Fig. 2.34 is that, although quantity is a function of price, here we measure price on the vertical axis and quantity on the horizontal axis. This has been standard practice in elementary price theory since the work of Alfred Marshall late in the nineteenth century.

The trouble with this method of analysis comes when the parameters are not known, so the supply and demand curves cannot be drawn with any certainty. Indeed, if all an economist observes is a decrease in price and an increase in quantity from (P^e, Q^e) to $(\tilde{P}^e, \tilde{Q}^e)$, there is no way of knowing (without more information) whether this results from just a rightward shift in the supply curve, as illustrated in Fig. 2.34, or from some combination of a shift to the right (or left) in demand *and* a shift to

FIGURE 2.34



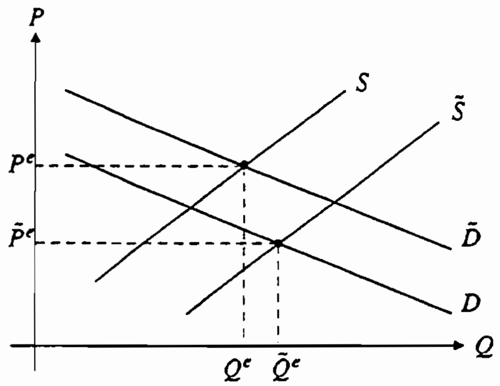


FIGURE 2.35

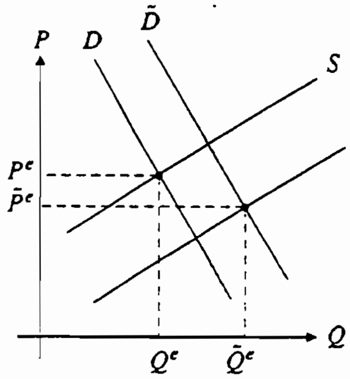


FIGURE 2.36

the right in supply, as illustrated in Figs. 2.35 and 2.36. All that can be said is that, because the equilibrium price falls and the quantity increases, there must have been some rightward shift in supply—but demand could have fallen, risen, or stayed the same. Moreover, there is also the possibility that the demand and supply curves could have changed their slopes—that is, the parameters b and β could also have changed.

The General Equation for a Straight Line

Any nonvertical line in the plane has the equation $y = ax + b$. A vertical line, that is parallel to the y-axis, will intersect the x-axis at some point $(c, 0)$. Every point on the line has the same x-coordinate c, so the line must be

$$x = c$$

This is the equation for a straight line through $(c, 0)$ parallel to the y-axis.

The equations $y = ax + b$ and $x = c$ can both be written as

$$Ax + By + C = 0 \qquad [2.9]$$

for suitable values of the constants A, B, and C. Specifically, $y = ax + b$ corresponds to $A = a$, $B = -1$, and $C = b$, whereas $x = c$ corresponds to $A = 1$, $B = 0$, and $C = -c$. Conversely, every equation of the form [2.9] represents a straight line in the plane, disregarding the uninteresting case when $A = B = 0$. If $B = 0$, it follows from [2.9] that $Ax = -C$, or $x = -C/A$. This is the equation for a straight line parallel to the y-axis. On the other hand, if $B \neq 0$, solving [2.9] for y yields

$$y = -\frac{A}{B}x - \frac{C}{B}$$

This is the equation for a straight line with slope $-A/B$. Equation [2.9] thus deserves to be called the *general equation for a straight line in the plane.*

Graphical Solutions of Linear Equations

Section A.9 of Appendix A deals with algebraic methods for solving a system of linear equations in two unknowns. The equations are linear, so their graphs are straight lines. The coordinates of any point on a line satisfy the equation of that line. Thus, the coordinates of any point of intersection of these lines will satisfy both equations. This means that a point of intersection solves the system.

Example 2.20

Solve each of the following three pairs of equations graphically:

(a) $x + y = 5$ and $x - y = -1$

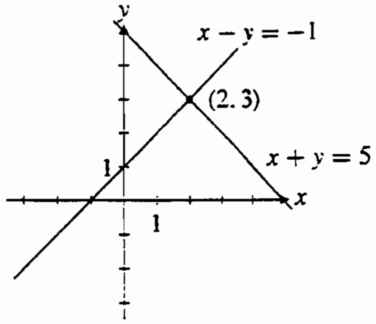
(b) $3x + y = -7$ and $x - 4y = 2$

(c) $3x + 4y = 2$ and $6x + 8y = 24$

Solution

(a) Figure 2.37 shows the graphs of the straight lines $x + y = 5$ and $x - y = -1$. There is only one point of intersection $(2, 3)$. The solution of the system is, therefore, $x = 2$, $y = 3$.

FIGURE 2.37



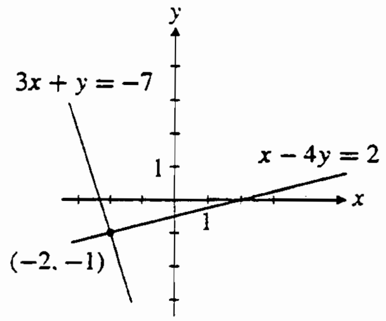


FIGURE 2.38

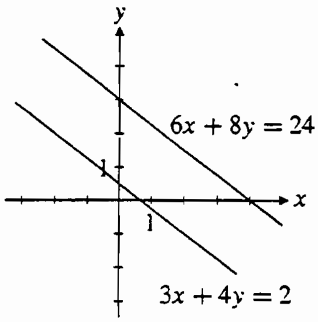


FIGURE 2.39

(b) Figure 2.38 shows the graphs of the straight lines $3x + y = -7$ and $x - 4y = 2$. There is only one point of intersection $(-2, -1)$. The solution of the system is, therefore, $x = -2$, $y = -1$.

(c) Figure 2.39 shows the graphs of the straight lines $3x + 4y = 2$ and $6x + 8y = 24$. These lines are parallel and have no point of intersection. The system has no solutions.

Linear Inequalities

This chapter concludes by discussing how to represent linear inequalities geometrically. Consider two examples.

Example 2.21

Sketch in the xy-plane the set of all pairs of numbers (x, y) that satisfy the inequality $2x + y \leq 4$. (Using set notation, this is $\{(x, y) : 2x + y \leq 4\}$.)

Solution The inequality can be written as $y \leq -2x+4$. The set of points (x, y) that satisfy the equation $y = -2x + 4$ is a straight line. Therefore, the set of points (x, y) that satisfy the inequality $y \leq -2x + 4$ must have y-values below those of points on the line $y = -2x + 4$. So it must consist of all points that lie on or below this straight line. See Fig. 2.40.

Example 2.22

A person has $\$m$ to spend on the purchase of two commodities. The prices of the two commodities are $\$p$ and $\$q$ per unit. Suppose x units of the first

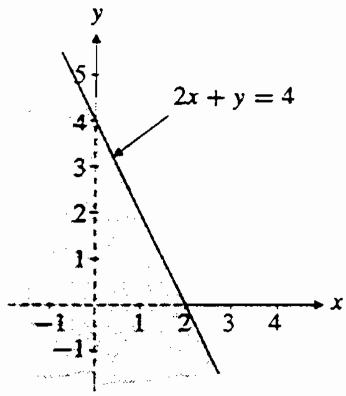


FIGURE 2.40 $\{(x, y) : 2x + y \leq 4\}$.

commodity and y units of the second commodity are bought. Assuming one cannot purchase negative units of x and y, the *budget set* is

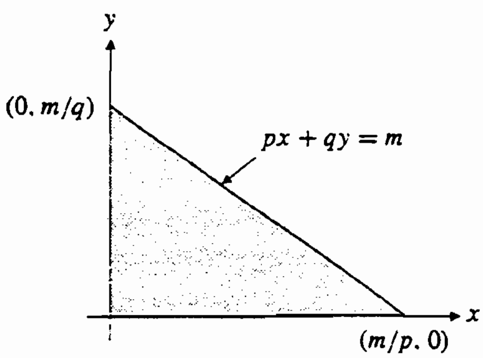
$$B = \{(x, y) : \ px + qy \leq m, \ x \geq 0, \ y \geq 0\}$$

as in (1.7) in Section 1.7. Sketch the budget set B in the xy-plane. Find the slope of the budget line $px + qy = m$, and its points of intersection with the two coordinate axes.

Solution The set of points (x, y) that satisfy $x \geq 0$ and $y \geq 0$ was sketched in Fig. 2.10. It is the first (nonnegative) quadrant. If we impose the additional requirement that $px + qy \leq m$, we obtain the triangular domain B shown in Fig. 2.41.

If we solve $px + qy = m$ for y, we get $y = (-p/q)x + m/q$, so the slope is $-p/q$. The budget line intersects the x-axis when $y = 0$. Then $px = m$, so $x = m/p$. The budget line intersects the y-axis when $x = 0$. Then $qy = m$, so $y = m/q$. So the two points of intersection are $(m/p, 0)$ and $(0, m/q)$, as shown in Fig. 2.41.

FIGURE 2.41 **Budget set:** $px + qy \leq m$, $x \geq 0$, **and** $y \geq 0$.



Problems

1. Find the slopes of the lines passing through the following points by using the formula in [2.5].

   a. (2, 3) and (5, 8) **b.** (−1, −3) and (2, −5) **c.** $(\frac{1}{2}, \frac{3}{2})$ and $(\frac{1}{3}, -\frac{1}{5})$

2. The consumption function $C = 4141 + 0.78\,Y$ for the UK was estimated for the period 1949–1975. What is the marginal propensity to consume?

3. Find the slopes of the five lines L_1 to L_5 shown in Fig. 2.42, and give equations describing them. (L_3 is horizontal.)

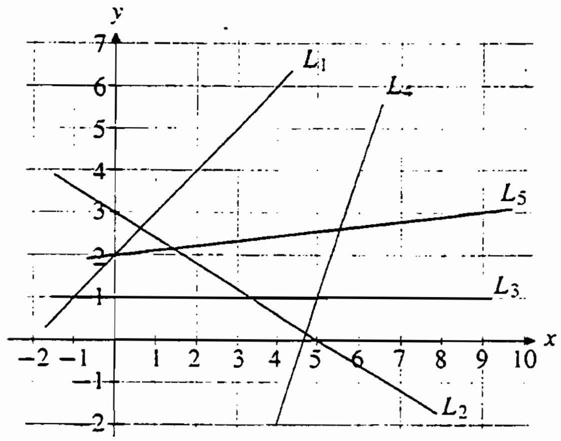


FIGURE 2.42

4. Draw graphs for the following equations:

   a. $3x + 4y = 12$ **b.** $\frac{x}{10} - \frac{y}{5} = 1$ **c.** $x = 3$

5. Decide which of the following relationships are linear:

   a. $5y + 2x = 2$ **b.** $P = 10(1 - 0.3t)$ **c.** $C = (0.5x + 2)(x - 3)$

   d. $p_1x_1 + p_2x_2 = R$ (p_1, p_2, and R constants)

6. **a.** Determine the relationship between Centigrade and Fahrenheit temperature scales when you know that (i) the relation is linear; (ii) water freezes at 0°C and 32°F; and (iii) water boils at 100°C and 212°F.

   b. Which temperature is measured by the same number in both Centigrade and Fahrenheit scales?

7. Determine the equations and draw graphs for the following straight lines:

   a. L_1 passes through (1, 3) and has a slope of 2.

   b. L_2 passes through (−2, 2) and (3, 3).

   c. L_3 passes through the origin and has a slope of −1/2.

   d. L_4 passes through $(a, 0)$ and $(0, b)$ (suppose $a \neq 0$).

8. A line L passes through the point (1, 1) and has a slope of 3. A second line M passes through (−1, 2) and (3, −1). Find the equations for L and M

and their point of intersection, P. Also determine the equation for the line N that passes through $(-1, -1)$ and is parallel to M. Draw the figure.

9. The total cost y of producing x units of some commodity is a linear function. Records show that on one occasion, 100 units were made at a total cost of \$200, and on another occasion, 150 units were made at a total cost of \$275. Express the linear equation for total cost in terms of the number of units x produced.

10. Find the equilibrium price in the model in Example 2.19 for the following.

a. $D = 75 - 3P, \quad S = 20 + 2P$ **b.** $D = 100 - 0.5P, \quad S = 10 + 0.5P$

11. According to 20th report of the International Commission on Whaling, the number N of fin whales in the Antarctic for the period 1958–1963 was given by

$$N = -17{,}400t + 151{,}000, \qquad 0 \leq t \leq 5$$

where $t = 0$ corresponds to January 1958, $t = 1$ corresponds to January 1959, and so on.

a. According to this equation, how many fin whales would there be left in April 1960?

b. If the decrease continued at the same rate, when would there be no fin whales left? (Actually, the 1993 estimate was approximately 21,000.)

12. The expenditure of a household on consumer goods, C, is related to the household's income, y, in the following way: When the household's income is \$1000, the expenditure on consumer goods is \$900, and whenever income is increased by \$100, the expenditure on consumer goods is increased by \$80. Express the expenditure on consumer goods as a function of income, assuming a linear relationship.

13. Solve the following three systems of equations graphically:

a. $x - y = 5$ and $x + y = 1$

b. $x + y = 2$, $x - 2y = 2$ and $x - y = 2$

c. $3x + 4y = 1$ and $6x + 8y = 6$

14. Show that $-1/[x_0(x_0 + h)]$ is the slope of the line passing through P and Q in Fig. 2.43.

15. The following table shows the total consumption and net national income in some country for the period from 1955–1960, measured in millions of dollars. Plot the points from the table in the YC-plane. Draw the straight line through the "extreme points" (21.3, 17.4) and (24.7, 20.4). Find the equation for this line. What is the interpretation of its slope?

Year	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*
Total consumption (C)	17.4	18.0	18.4	18.6	19.3	20.4
Net national product (Y)	21.3	22.4	23.0	22.6	23.4	24.7

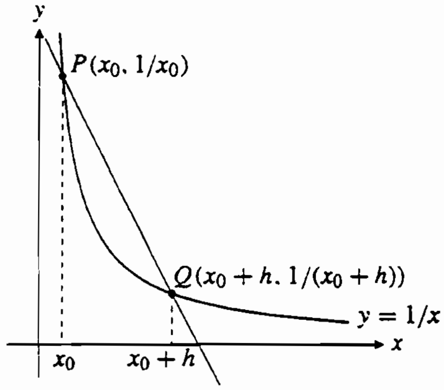


FIGURE 2.43

16. Sketch in the xy-plane the set of all pairs of numbers (x, y) that satisfy the following inequalities:

a. $2x + 4y \geq 5$ **b.** $x - 3y + 2 \leq 0$ **c.** $100x + 200y \leq 300$

17. Sketch in the xy-plane the set of all pairs of numbers (x, y) that satisfy all the following three inequalities: $3x + 4y \leq 12$; $x - y \leq 1$; and $3x + y \geq 3$.

3

Polynomials, Powers, and Exponentials

The paradox is now fully established that the utmost abstractions are the true weapons with which to control our thought of concrete facts.
—*A. N. Whitehead*

The linear functions and associated linear models that were studied in some detail in the previous chapter are particularly simple. Not surprisingly, most economic applications require much more accuracy than is possible with only linear functions, and so economists most often use more complicated functions.

3.1 Quadratic Functions

Many economic models involve functions that either decrease down to some minimum value and then increase, or else increase up to some maximum value and then decrease. Simple functions with this property are the general **quadratic** functions

$$f(x) = ax^2 + bx + c \qquad (a,\ b, \text{ and are } c \text{ constants, } a \neq 0) \qquad [3.1]$$

(If $a = 0$, the function is linear, hence, the restriction $a \neq 0$.) Figure 2.20 of Section 2.4 shows the graph of $f(x) = x^2 - 3x$, which is obtained from [3.1] by choosing $a = 1$, $b = -3$, and $c = 0$. In general, the graph of $f(x) = ax^2 + bx + c$ is called a **parabola**. The shape of this parabola roughly

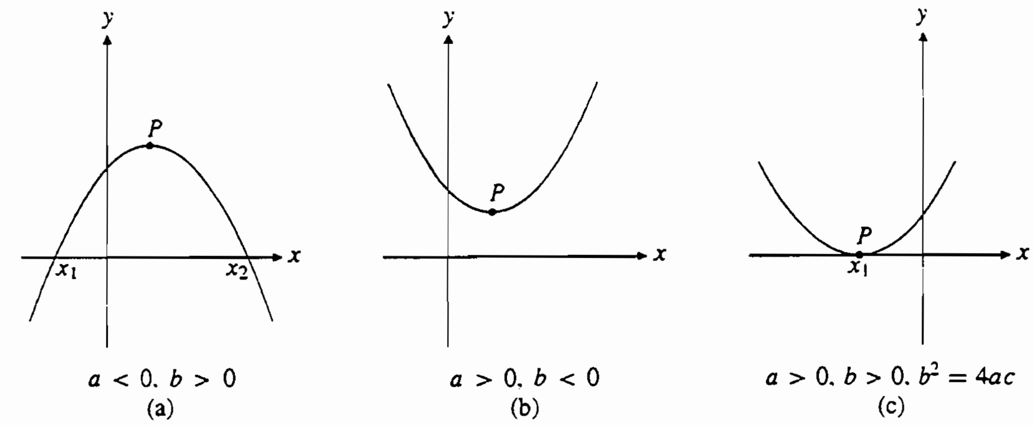


FIGURE 3.1 Graphs of the parabola $y = ax^2 + bx + c$.

resembles ∩ when $a < 0$ and ∪ when $a > 0$. Three typical cases are illustrated in Fig. 3.1.

In order to understand the function $f(x) = ax^2 + bx + c$ in more detail, we are interested in the answers to the following questions:

1. For what values of x (if any) is $ax^2 + bx + c = 0$?
2. What are the coordinates of the maximum/minimum point P?

In the case of question 1, we have to find solutions to the equation $f(x) = 0$. Geometrically, this involves determining points of intersection of the parabola with the x-axis. These points are called the **zeros** of the quadratic function. In Fig. 3.1(a), the zeros are given by x_1 and x_2, in Fig. 3.1(b) there are no zeros, whereas the graph in Fig. 3.1(c) has x_1 as its only point of intersection with the x-axis. In Section A.8 of Appendix A it is proved that, in the case when $b^2 \geq 4ac$ and $a \neq 0$, then

$$ax^2 + bx + c = 0 \iff x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad [3.2]$$

To derive this formula, we used the method known as "completing the square." This technique will also help us answer question 2. In fact, when $a \neq 0$, the function defined by [3.1] can be expressed as

$$f(x) = a\left[x^2 + 2\left(\frac{b}{2a}\right)x + \left(\frac{b}{2a}\right)^2\right] - a\left(\frac{b}{2a}\right)^2 + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} \qquad [3.3]$$

Consider the expression after the second equality sign of [3.3]. When x varies, only the value of $a(x + b/2a)^2$ changes. This term is equal to 0 when $x = -b/2a$, and if $a > 0$, it is never less than 0. This means that when $a > 0$, then the function $f(x)$ attains its minimum when $x = -b/2a$, and the value of $f(x)$ is then equal to $f(-b/2a) = -(b^2 - 4ac)/4a = c - b^2/4a$. If $a < 0$ on the other hand, then $a(x + b/2a)^2 \leq 0$ for all x, and the squared term is equal to 0 when $x = -b/2a$. Hence, $f(x)$ attains its maximum when $x = -b/2a$ in this second case. To summarize, we have shown the following:

If $a > 0$, then $f(x) = ax^2 + bx + c$ has a **minimum** at

$$\left(-\frac{b}{2a}, c - \frac{b^2}{4a}\right)$$

If $a < 0$, then $f(x) = ax^2 + bx + c$ has a **maximum** at

$$\left(-\frac{b}{2a}, c - \frac{b^2}{4a}\right)$$

[3.4]

If you find it difficult to follow the argument leading up to [3.4], you should study the following special examples very carefully.

Example 3.1

Complete the square as in [3.3] for the following functions and then find the maximum/minimum point of each:

(a) $f(x) = x^2 - 4x + 3$

(b) $f(x) = -2x^2 + 40x - 600$

(c) $f(x) = \frac{1}{3}x^2 + \frac{2}{3}x - \frac{8}{3}$

Solution

(a) $x^2 - 4x + 3 = (x^2 - 4x) + 3 = (x^2 - 4x + 4) - 4 + 3 = (x - 2)^2 - 1$

The expression $(x - 2)^2 - 1$ attains its smallest value, which is -1, at $x = 2$.

(b)
$$\begin{aligned} -2x^2 + 40x - 600 &= -2(x^2 - 20x) - 600 \\ &= -2(x^2 - 20x + 100) + 200 - 600 \\ &= -2(x - 10)^2 - 400 \end{aligned}$$

The expression $-2(x - 10)^2 - 400$ attains its largest value, which is -400, at $x = 10$.

(c) $\frac{1}{3}x^2 + \frac{2}{3}x - \frac{8}{3} = \frac{1}{3}(x^2 + 2x) - \frac{8}{3}$

$$= \frac{1}{3}(x^2 + 2x + 1) - \frac{1}{3} - \frac{8}{3}$$

$$= \frac{1}{3}(x + 1)^2 - 3$$

The expression $\frac{1}{3}(x + 1)^2 - 3$ attains its smallest value, which is -3, at $x = -1$.

A useful exercise is to solve the three cases in Example 3.1 by using the expressions set out in [3.4] directly, substituting appropriate values for the three parameters a, b, and c. You should then check that the same results are obtained.

Problems

1. a. Let $f(x) = x^2 - 4x$. Complete the following table:

x	−1	0	1	2	3	4	5
$f(x)$							

b. Using the table in part (a), sketch the graph of f.
c. Using [3.3], determine the minimum point.
d. Solve the equation $f(x) = 0$.

2. a. Let $f(x) = -\frac{1}{2}x^2 - x + \frac{3}{2}$. Complete the following table:

x	−4	−3	−2	−1	0	1	2
$f(x)$							

b. Use the information in part (a) to sketch the graph of f.
c. Using [3.3], determine the maximum point.
d. Solve the equation $-\frac{1}{2}x^2 - x + \frac{3}{2} = 0$ for x.
e. Show that $f(x) = -\frac{1}{2}(x - 1)(x + 3)$, and use this to study how the sign of f varies with x. Compare the result with the graph.

3. Complete the squares as in [3.3] for the following quadratic functions, and then determine the maximum/minimum points:

a. $x^2 + 4x$ b. $x^2 + 6x + 18$ c. $-3x^2 + 30x - 30$
d. $9x^2 - 6x - 44$ e. $-x^2 - 200x + 30{,}000$ f. $x^2 + 100x - 20{,}000$

4. Find the zeros of each quadratic function in Problem 3, and write each function in the form $a(x - x_1)(x - x_2)$ (if possible).

5. Use the formula in [3.2] to find solutions to the following equations, where p and q are positive parameters.

   a. $x^2 - 3px + 2p^2 = 0$ b. $x^2 - (p+q)x + pq = 0$

   c. $x^2 + px + q = 0$

6. A person is given a rope of length L with which to enclose a rectangular area.

   a. If one of the sides is x, show that the area of the enclosure is $A(x) = Lx/2 - x^2$, where $0 \le x \le L/2$. Find x such that the area of the rectangle is maximized.

   b. Will a circle of circumference L enclose an area that is larger than the one we found in part (a)? (It is reported that certain surveyors in antiquity wrote contracts with farmers to sell them rectangular pieces of land in which only the circumference was specified. As a result, the lots were long narrow rectangles.)

7. Consider the function given by the formula $A = 500x - x^2$ in Example 1.1 of Section 1.3. What choice of x gives the largest value for the area A?

8. a. Solve $x^4 - 5x^2 + 4 = 0$. (*Hint:* Put $x^2 = u$ and form a quadratic equation in u.)

   b. Solve the equations (i) $x^4 - 8x^2 - 9 = 0$ and (ii) $x^6 - 9x^3 + 8 = 0$.

9. A model occurring in the theory of efficient loan markets involves the function

$$U(x) = 72 - (4+x)^2 - (4-rx)^2$$

where r is a constant. Find the value of x for which $U(x)$ attains its largest value.

10. Find the equation for the parabola $y = ax^2 + bx + c$ that passes through the three points $(1, -3)$, $(0, -6)$, and $(3, 15)$. (*Hint:* Determine a, b, and c.)

Harder Problems

11. The graph of a function f is said to be *symmetric* about the line $x = p$ if

$$f(p-t) = f(p+t) \qquad \text{(for all } t\text{)}$$

Show that the parabola $f(x) = ax^2 + bx + c$ is symmetric about the line $x = -b/2a$. (*Hint:* Use [3.3].)

12. Let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be arbitrary real numbers. We claim that the following inequality (called the **Cauchy–Schwarz inequality**) is always valid:

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \le (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2) \qquad [3.5]$$

a. Check the inequality for (i) $a_1 = 1$, $a_2 = 3$, $b_1 = 2$, and $b_2 = 5$; and for (ii) $a_1 = -3$, $a_2 = 2$, $b_1 = 5$, and $b_2 = -2$. (In both cases, $n = 2$.)

b. Prove [3.5] by means of the following trick: first, define f for all x by

$$f(x) = (a_1x + b_1)^2 + \cdots + (a_nx + b_n)^2$$

We see that $f(x) \geq 0$ for all x. Write $f(x)$ as $Ax^2 + Bx + C$, where the expressions for A, B, and C are related to the terms in [3.5]. Because $Ax^2 + Bx + C \geq 0$ for all x, we must have $B^2 - 4AC \leq 0$. Why? The conclusion follows.

3.2 Examples of Quadratic Optimization Problems

Much of mathematical economics is concerned with optimization problems. Economics, after all, is the science of choice, and optimization problems are the form in which choice is usually expressed mathematically.

A general discussion of such problems must be postponed until we have developed the necessary tools from calculus. Here we show how the simple results from the previous section on maximizing quadratic functions can be used to illustrate some basic economic ideas.

Example 3.2 (A Monopoly Problem)

Consider a firm that is the only seller of the commodity it produces, possibly a patented medicine, and so enjoys a monopoly. The total costs of the monopolist are assumed to be given by the quadratic function

$$C = \alpha Q + \beta Q^2, \qquad Q \geq 0 \qquad [1]$$

of its output level Q, where α and β are positive constants. For each Q, the price P at which it can sell its output is assumed to be determined from the linear "inverse" demand function

$$P = a - bQ, \qquad Q \geq 0 \qquad [2]$$

where a and b are constants with $a > 0$ and $b \geq 0$. So for any nonnegative Q, the total revenue R is given by the quadratic function

$$R = PQ = (a - bQ)Q$$

and profit by the quadratic function[1]

$$\pi(Q) = R - C = (a - bQ)Q - \alpha Q - \beta Q^2$$
$$= (a - \alpha)Q - (b + \beta)Q^2 \qquad [3]$$

The monopolist's objective is to maximize $\pi = \pi(Q)$. By using [3.4], we see that there is a maximum of π (for the monopolist M) at

$$Q^M = \frac{a - \alpha}{2(b + \beta)} \quad \text{with} \quad \pi^M = \frac{(a - \alpha)^2}{4(b + \beta)} \qquad [4]$$

This is valid if $a > \alpha$; if $a \leq \alpha$, the firm will not produce, but will have $Q^M = 0$ and $\pi^M = 0$. The two cases are illustrated in Figs. 3.2 and 3.3. The associated price and cost can be found by routine algebra.

If we put $b = 0$ in [2], then $P = a$ for all Q. In this case, the firm's choice of quantity does not influence the price at all and so the firm is said to be *perfectly competitive*. By replacing a by P in [3] and putting $b = 0$, we see that profit is maximized for a perfectly competitive firm at

$$Q^* = \frac{P - \alpha}{2\beta} \quad \text{with} \quad \pi^* = \frac{(P - \alpha)^2}{4\beta} \qquad [5]$$

provided that $P > \alpha$. If $P \leq \alpha$, then $Q^* = 0$ and $\pi^* = 0$.

FIGURE 3.2 The profit function, $a > \alpha$.

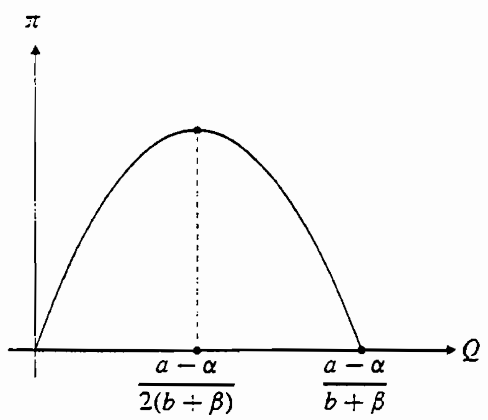
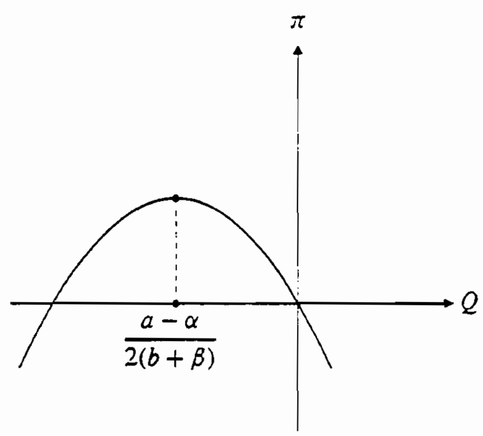


FIGURE 3.3 The profit function, $a \leq \alpha$.



[1] Previously, π has been used to denote the constant ratio 3.14159... between the circumference of a circle and its diameter. In economics, this constant is not used very often, so π has come to denote profit, or probability.

Solving the first equation in [5] for P yields $P = \alpha + 2\beta Q^*$. Thus,

$$P = \alpha + 2\beta Q \quad [6]$$

represents the supply curve of this perfectly competitive firm for $P > \alpha$ when $Q^* > 0$, whereas for $P \leq \alpha$, the profit maximizing output Q^* is 0. The supply curve relating the price on the market to the firm's choice of output quantity is shown in Fig. 3.4; it includes points between the origin and $(0, x)$.

Let us return to the monopoly firm (which has no supply curve). If it could somehow be made to act like a competitive firm, taking price as given, it would be on the supply curve [6]. Given the demand curve $P = a - bQ$, equilibrium between supply and demand occurs when [6] is also satisfied, and so $P = a - bQ = \alpha + 2\beta Q$. Solving the second equation for Q, and then substituting for P and π in turn, we see that the equilibrium level of output, the corresponding price, and the profit would be

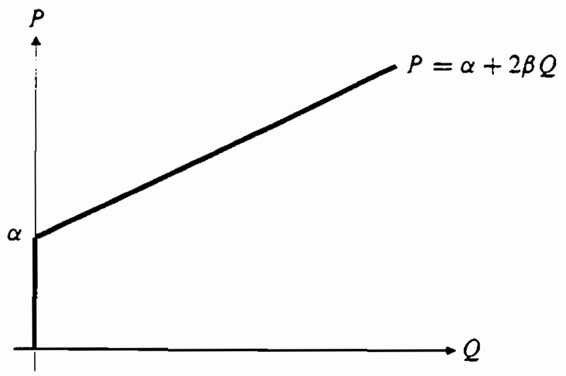
$$Q^e = \frac{a - \alpha}{b + 2\beta}, \qquad P^e = \frac{2a\beta + \alpha b}{b + 2\beta}, \qquad \pi^e = \frac{\beta(a - \alpha)^2}{(b + 2\beta)^2} \quad [7]$$

In order to have the monopolist mimic a competitive firm by choosing to be at (Q^e, P^e), it may be desirable to tax (or subsidize) the output of the monopolist. Suppose that the monopolist is required to pay a specific tax of t per unit of output. Because the tax payment tQ is added to the firm's costs, the new total cost function is

$$\begin{aligned} C &= \alpha Q + \beta Q^2 + tQ \\ &= (\alpha + t)Q + \beta Q^2 \end{aligned} \quad [8]$$

Carrying out the same calculations as before, but with α replaced by $\alpha + t$,

FIGURE 3.4 The supply curve of a perfectly competitive firm.



gives the monopolist's choice of output as

$$Q_t^M = \begin{cases} \dfrac{a-\alpha-t}{2(b+\beta)}, & \text{if } a \geq \alpha + t \\ 0, & \text{otherwise} \end{cases} \qquad [9]$$

So $Q_t^M = Q^e$ when $(a-\alpha-t)/2(b+\beta) = (a-\alpha)/(b+2\beta)$. Solving this equation for t yields $t = -(a-\alpha)b/(b+2\beta)$. Note that t is actually negative, indicating the desirability of *subsidizing* the output of the monopolist in order to encourage additional production. (Of course, subsidizing monopolists is usually felt to be unjust, and many additional complications need to be considered carefully before formulating a desirable policy for dealing with monopolists. Still the previous analysis suggests that if justice requires lowering a monopolist's price or profit, this is much better done directly than by taxing output.)

Problem

1. If a cocoa shipping firm sells Q tons of cocoa in England, the price received is given by $P = \alpha_1 - \frac{1}{3}Q$. On the other hand, if it buys Q tons from its only source in Ghana, the price it has to pay is given by $P = \alpha_2 + \frac{1}{6}Q$. In addition, it costs γ per ton to ship cocoa from its supplier in Ghana to its customers in England (its only market). The numbers α_1, α_2, and γ are all positive.
   a. Express the cocoa shipper's profit as a function of Q, the number of tons shipped.
   b. Assuming that $\alpha_1 - \alpha_2 - \gamma > 0$, find the profit maximizing shipment of cocoa. What happens if $\alpha_1 - \alpha_2 - \gamma \leq 0$?
   c. Suppose the government of Ghana imposes an export tax on cocoa of t per ton. Find the new expression for the shipper's profits and the new quantity shipped.
   d. Calculate the government's export tax revenue as a function of t, and advise it on how to obtain as much tax revenue as possible.

3.3 Polynomials

After considering linear and quadratic functions, the logical next step is to examine **cubic functions** of the form

$$f(x) = ax^3 + bx^2 + cx + d \qquad (a, b, c, \text{and } d \text{ are constants}; a \neq 0) \qquad [3.6]$$

It is relatively easy to understand the behavior of linear and quadratic functions from their graphs. Cubic functions are considerably more complicated, because

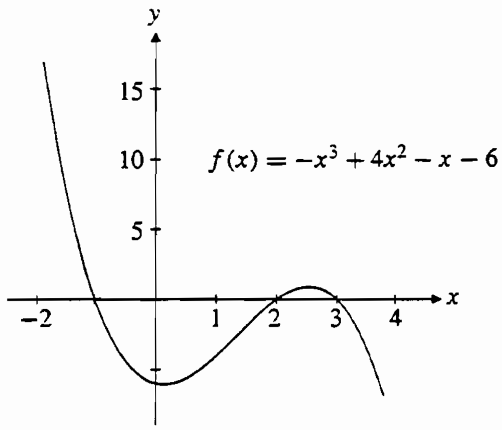


FIGURE 3.5 A cubic function.

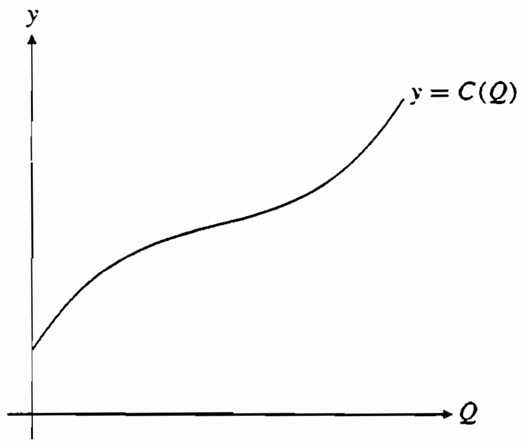


FIGURE 3.6 A cubic cost function.

the shape of their graphs changes drastically as the coefficients a, b, c, and d vary. Two examples are given in Figs. 3.5 and 3.6.

Cubic functions do occasionally appear in economic models. Let us look at a typical example.

Example 3.3

Consider a firm producing a single commodity. The total cost of producing Q units of the commodity is $C(Q)$. Cost functions often have the following properties: First, $C(0)$ is positive, because an initial fixed expenditure is involved. When production increases, costs also increase. In the beginning, costs increase rapidly, but the rate of increase slows down as production equipment is used for a higher proportion of each working week. However, at high levels of production, costs again increase at a fast rate, because of technical bottlenecks and overtime payments to workers, for example. The cubic cost function $C(Q) = aQ^3 + bQ^2 + cQ + d$ exhibits this type of behavior provided that $a > 0$, $b < 0$, $c > 0$, and $d > 0$ with $3ac > b^2$. Such a function is sketched in Fig. 3.6.

General Polynomials

Linear, quadratic, and cubic functions are all examples of **polynomials**. The function P defined for all x by

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \qquad (a\text{'s are constants; } a_n \neq 0) \quad [3.7]$$

is called the **general polynomial of degree n**. When $n = 4$, we obtain $P(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$, which is the general polynomial of degree 4.

Numerous problems in mathematics and its applications involve polynomials. Often, one is particularly interested in finding the number and location of the zeros of $P(x)$—that is, the values of x such that $P(x) = 0$. The equation

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \qquad [3.8]$$

is called the **general nth-order equation**. It will soon be shown that this equation has *at most n* (real) solutions, also called **roots**, but it need not have any.

According to the **fundamental theorem of algebra**, every polynomial of the form [3.7] can be written as a product of polynomials of first or second degree. Here is a somewhat complicated case:

$$x^5 - x^4 + x - 1 = (x - 1)(x^4 + 1) = (x - 1)\left(x^2 - \sqrt{2}x + 1\right)\left(x^2 + \sqrt{2}x + 1\right)$$

Integer Roots

Suppose that x_0 is an integer that satifies the cubic equation $-x^3 + 4x^2 - x - 6 = 0$, or, equivalently, $-x^3 + 4x^2 - x = 6$. Then x_0 must also satisfy the equation

$$x_0(-x_0^2 + 4x_0 - 1) = 6 \qquad [*]$$

Because x_0 is an integer, it follows that x_0^2, $4x_0$, and $-x_0^2 + 4x_0 - 1$ must also be integers. But because x_0 multiplied by the integer $-x_0^2 + 4x_0 - 1$ is equal to 6, the number x_0 must be a factor of 6—that is, 6 must be divisible by x_0. Now, the only integers by which 6 is divisible are ±1, ±2, ±3, and ±6. Direct substitution into the left-hand side (LHS) of equation [*] reveals that of these eight possibilities, -1, 2, and 3 are roots of the equation. A third degree equation has at most three roots, so we have found all of them. In general, we can state the following result:

> Suppose that $a_n, a_{n-1}, \ldots, a_1, a_0$ are all integers. Then all possible integer roots of the equation
>
> $$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \qquad [3.9]$$
>
> must be factors of the constant term a_0.

Proof If x_0 is an integer root, then x_0 must satisfy the equation

$$x_0(a_n x_0^{n-1} + a_{n-1}x_0^{n-2} + \cdots + a_1) = -a_0$$

Both factors on the left are integers, so $-a_0$ must be divisible by each of them, and in particular by x_0. So must a_0.

Example 3.4
Find all possible integer roots to the equation $\frac{1}{2}x^3 - x^2 + \frac{1}{2}x - 1 = 0$.

Solution We multiply both sides of the equation by 2 to obtain an equation whose coefficients are all integers:

$$x^3 - 2x^2 + x - 2 = 0$$

According to [3.9], all integer solutions of the equation must be factors of -2. So only ± 1 and ± 2 can be integer solutions. A check shows that $x = 2$ is the only integer solution. In fact, because $x^3-2x^2+x-2 = (x-2)(x^2+1)$, there is only one real root.

The Remainder Theorem

Let $P(x)$ and $Q(x)$ be two polynomials for which the degree of $P(x)$ is greater than or equal to the degree of $Q(x)$. Then there always exist unique polynomials $q(x)$ and $r(x)$ such that

$$P(x) = q(x)Q(x) + r(x) \qquad [3.10]$$

where the degree of $r(x)$ is less than the degree of $Q(x)$. This fact is called the **remainder theorem**. When x is such that $Q(x) \neq 0$, then [3.10] can be written in the form

$$\frac{P(x)}{Q(x)} = q(x) + \frac{r(x)}{Q(x)} \qquad [3.11]$$

If $r(x) = 0$ in [3.10] and [3.11], we say that $Q(x)$ *is a factor of* $P(x)$, or that $P(x)$ *is divisible by* $Q(x)$. Then $P(x) = q(x)Q(x)$ or $P(x)/Q(x) = q(x)$, which is the *quotient*. When $r(x) \neq 0$, it is the *remainder*.

An important special case is when $Q(x) = x - a$. Then $Q(x)$ is of degree 1, so the remainder $r(x)$ must have degree 0, and is therefore a constant. In this special case, for all x,

$$P(x) = q(x)(x-a) + r$$

For $x = a$ in particular, we get $P(a) = r$. *Hence, $x - a$ divides $P(x)$ if and only if $P(a) = 0$.* This is an important observation that can be formulated as follows:

Polynomial $P(x)$ has the factor $x - a \iff P(a) = 0$	[3.12]

It follows from [3.12] that an nth-degree polynomial $P(x)$ can have *at most* n different zeros. To see this, note that each zero $x = a_1$, $x = a_2$, ..., $x = a_k$ gives rise to a different factor of the form $x - a$. From this it follows that $P(x)$ can be expressed as $P(x) = A(x)(x - a_1) \dots (x - a_k)$ for some polynomial $A(x)$. Thus, $P(x)$ has degree $\geq k$, and so k cannot exceed n.

Example 3.5

Prove that the polynomial $f(x) = -2x^3 + 2x^2 + 10x + 6$ has a zero at $x = 3$, and factorize the polynomial.

Solution Inserting $x = 3$ into the polynomial yields

$$f(3) = -2 \cdot 3^3 + 2 \cdot 3^2 + 10 \cdot 3 + 6 = -54 + 18 + 30 + 6 = 0$$

So $x - 3$ is a factor. It follows that the cubic function $f(x)$ can be expressed as the product of $(x - 3)$ with a second degree polynomial. In fact,

$$f(x) = -2x^3 + 2x^2 + 10x + 6 = -2(x - 3)(x^2 + ax + b)$$

We must determine a and b. Expanding the last expression yields

$$f(x) = -2x^3 + (6 - 2a)x^2 + (6a - 2b)x + 6b$$

If this polynomial $f(x)$ is to equal $-2x^3 + 2x^2 + 10x + 6$ for all x, then the coefficients of like powers of x must be equal; thus, $6 - 2a = 2$, $6a - 2b = 10$, and $6b = 6$. Hence, $b = 1$ and $a = 2$. Because $x^2 + 2x + 1 = (x + 1)^2$, we conclude that

$$\begin{aligned} f(x) &= -2x^3 + 2x^2 + 10x + 6 = -2(x - 3)(x^2 + 2x + 1) \\ &= -2(x - 3)(x + 1)^2 \end{aligned}$$

The factorization procedure used in this example is called the *method of undetermined coefficients*. (Here a and b were the undetermined coefficients.) The alternative "long-division" method for factorizing polynomials will be considered next.

Polynomial Division

One can divide polynomials in much the same way as one divides numbers. Consider first a simple numerical example:

$$
\begin{array}{rl}
2735 \div 5 & = 500 + 40 + 7 \\
\underline{2500} & \\
235 & \\
\underline{200} & \\
35 & \\
\underline{35} & \\
0 & \quad \text{remainder}
\end{array}
$$

Hence, $2735 \div 5 = 547$. Note that the horizontal lines instruct you to subtract the numbers above the lines. (You might be more accustomed to a different way of arranging the numbers, but the idea is the same.)

Consider next

$$(-x^3 + 4x^2 - x - 6) \div (x - 2)$$

We write the following:

$$
\begin{array}{llll}
(-x^3 + 4x^2 - x - 6) \div (x - 2) & = -x^2 & + 2x & + 3 \\
\underline{-x^3 + 2x^2} & \longleftarrow \boxed{-x^2(x-2)} & & \\
\quad 2x^2 - x - 6 & & & \\
\quad \underline{2x^2 - 4x} & \longleftarrow \boxed{2x(x-2)} \longleftarrow & & \\
\qquad 3x - 6 & & & \\
\qquad \underline{3x - 6} & \longleftarrow \boxed{3(x-2)} \longleftarrow & & \\
\qquad\quad 0 & \text{remainder} & &
\end{array}
$$

(You can omit the boxes, but they should help you to see what is going on.) We conclude that $(-x^3 + 4x^2 - x - 6) \div (x - 2) = -x^2 + 2x + 3$. Because it is easy to see that $-x^2 + 2x + 3 = -(x + 1)(x - 3)$, we have

$$-x^3 + 4x^2 - x - 6 = -(x + 1)(x - 3)(x - 2)$$

Polynomial Division with a Remainder

The division $2734 \div 5$ gives 546 and leaves the remainder 4. So $2734/5 = 546 + 4/5$. We consider a similar form of division for polynomials.

Example 3.6

$(x^4 + 3x^2 - 4) \div (x^2 + 2x)$

Solution

$$\begin{array}{l} (x^4 \qquad + 3x^2 \qquad - 4) \div (x^2 + 2x) = x^2 - 2x + 7 \\ \underline{x^4 + 2x^3} \\ \quad - 2x^3 + 3x^2 \qquad - 4 \\ \quad \underline{- 2x^3 - 4x^2} \\ \qquad\qquad 7x^2 \qquad - 4 \\ \qquad\qquad \underline{7x^2 + 14x} \\ \qquad\qquad\quad - 14x - 4 \quad \text{remainder} \end{array}$$

(The polynomial $x^4 + 3x^2 - 4$ has no terms in x^3 and x, so we inserted some extra space between the powers of x to make room for the terms in x^3 and x that arise in the course of the calculations.) We conclude that

$$x^4 + 3x^2 - 4 = (x^2 - 2x + 7)(x^2 + 2x) + (-14x - 4)$$

Hence,

$$\frac{x^4 + 3x^2 - 4}{x^2 + 2x} = x^2 - 2x + 7 - \frac{14x + 4}{x^2 + 2x} \qquad [*]$$

Rational Functions

A **rational function** is a function $R(x) = P(x)/Q(x)$ that can be expressed as the ratio of two polynomials $P(x)$ and $Q(x)$. This function is defined for all x where $Q(x) \neq 0$. The rational function $R(x)$ is called **proper** if the degree of $P(x)$ is less than the degree of $Q(x)$. When the degree of $P(x)$ is greater than or equal to that of $Q(x)$, then $R(x)$ is called an **improper** rational function. By using polynomial division, any improper rational function can be written as a polynomial plus a proper rational function, as in [3.11] and Example 3.6.

Problems

1. By making use of [3.9], find all integer roots of the following equations:
   a. $x^2 + x - 2 = 0$ **b.** $x^3 - x^2 - 25x + 25 = 0$ **c.** $x^5 - 4x^3 - 3 = 0$
2. Find all integer roots of the following equations:
   a. $x^4 - x^3 - 7x^2 + x + 6 = 0$ **b.** $2x^3 + 11x^2 - 7x - 6 = 0$
   c. $x^4 + x^3 + 2x^2 + x + 1 = 0$ **d.** $\frac{1}{4}x^3 - \frac{1}{4}x^2 - x + 1 = 0$
3. Perform the following divisions:
   a. $(x^2 - x - 20) \div (x - 5)$ **b.** $(x^3 - 1) \div (x - 1)$
   c. $(-3x^3 + 48x) \div (x - 4)$

4. Perform the following divisions:
   a. $(2x^3 + 2x - 1) \div (x - 1)$ b. $(x^4 + x^3 + x^2 + x) \div (x^2 + x)$
   c. $(3x^8 + x^2 + 1) \div (x^3 - 2x + 1)$ d. $(x^5 - 3x^4 + 1) \div (x^2 + x + 1)$
5. Which of the following divisions leave no remainder? (a and b are constants; n is a natural number.)
   a. $(x^3 - x - 1)/(x - 1)$ b. $(2x^3 - x - 1)/(x - 1)$
   c. $(x^3 - ax^2 + bx - ab)/(x - a)$ d. $(x^{2n} - 1)/(x + 1)$
6. Write the following polynomials as products of linear factors:
   a. $p(x) = x^3 + x^2 - 12x$ b. $q(x) = 2x^3 + 3x^2 - 18x + 8$
7. Find possible formulas for each of the three polynomials with graphs in Fig. 3.7.

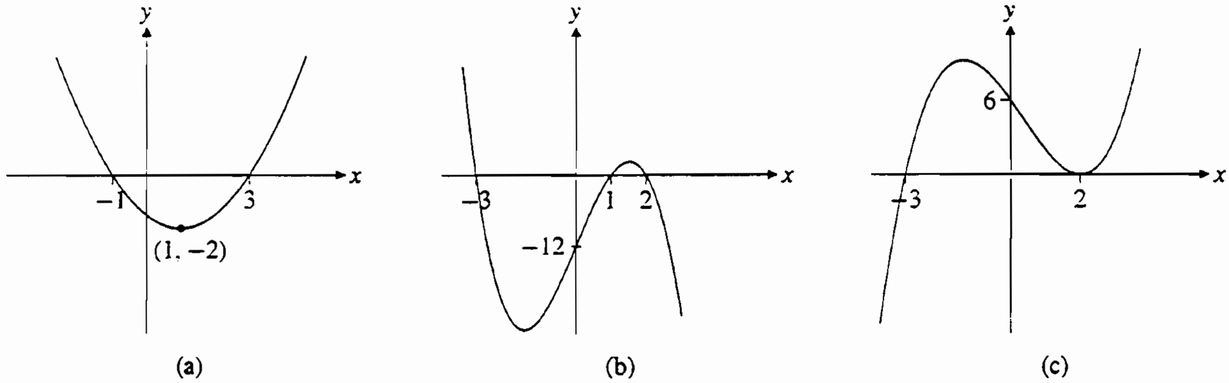


FIGURE 3.7

3.4 Power Functions

Consider the power function f defined by the formula

$$f(x) = x^r \qquad [3.13]$$

We know the meaning of x^r if r is any integer—that is, $r = 0, \pm 1, \pm 2, \ldots$. In fact, if r is a natural number, x^r is the product of r x's. Also if $r = 0$, then $x^r = x^0 = 1$ for all $x \neq 0$, and if $r = -n$, then $x^r = 1/x^n$ for $x \neq 0$. In addition, for $r = 1/2$, $x^r = x^{1/2} = \sqrt{x}$, defined for all $x \geq 0$. (See Section A.2 of Appendix A.) This section extends the definition of x^r so that it has meaning for any rational number r.

Here are some examples of why powers with rational exponents are needed:

1. The flow of blood (in liters per second) through the heart of an individual is approximately proportional to $x^{0.7}$, where x is the body weight.
2. The formula $S \approx 4.84V^{2/3}$ gives the approximate surface S of a ball as a function of its volume V. (See Example 3.10, which follows.)

3. The formula $Y = 2.262K^{0.203}L^{0.763}(1.02)^t$ appears in a study of the growth of national output, and shows how powers with fractional exponents can arise in economics. (Here Y is the net national product, K is capital stock, L is labor, and t is time.)

These examples illustrate the need to define x^r for $r = 0.7 = 7/10$, $r = 2/3$, $r = 0.203$, and $r = 0.763 = 763/1000$. In general, we want to define x^r for $x > 0$ when r is an arbitrary rational number.

The following basic power rules (discussed in Section A.1, Appendix A) are valid for all integers r and s:

$$\text{(i) } a^r a^s = a^{r+s} \qquad \text{(ii) } (a^r)^s = a^{rs} \tag{3.14}$$

When extending the definition of x^r so that it also applies to rational exponents r, it is natural to require that these rules retain their validity.

Let us first examine the meaning of $a^{1/n}$, where n is a natural number, and a is positive. For example, what does $5^{1/3}$ mean? If rule [3.14](ii) is still to apply in this case, we must have $(5^{1/3})^3 = 5$. This implies that $5^{1/3}$ must be a solution of the equation $x^3 = 5$. This equation can be shown to have a unique positive solution, denoted by $\sqrt[3]{5}$, the *cube root of* 5. (See Example 7.2 in Section 7.1.) Therefore, we must define $5^{1/3}$ as $\sqrt[3]{5}$. In general, $(a^{1/n})^n = a$. Thus, $a^{1/n}$ is a solution of the equation $x^n = a$. This equation can be shown to have a unique positive solution denoted by $\sqrt[n]{a}$, the *nth root of* a:

$$a^{1/n} = \sqrt[n]{a} \tag{3.15}$$

In words: *if a is positive and n is a natural number, then $a^{1/n}$ is the unique positive number that, raised to the nth power, gives a—that is, $(a^{1/n})^n = a^1 = a$.* For example,

$$27^{1/3} = \sqrt[3]{27} = 3 \qquad \text{because} \qquad 3^3 = (27^{1/3})^3 = 27$$

$$\left(\tfrac{1}{625}\right)^{1/4} = \sqrt[4]{\tfrac{1}{625}} = \tfrac{1}{5} \qquad \text{because} \qquad \left(\tfrac{1}{5}\right)^4 = \left[\left(\tfrac{1}{625}\right)^{1/4}\right]^4 = \tfrac{1}{625}$$

Usually, we write $a^{1/2}$ as $\sqrt{a}$ rather than $\sqrt[2]{a}$ (see Section A.2 of Appendix A).

We proceed to define $a^{p/q}$ whenever p is an integer, q is a natural number, and $a > 0$. Consider $5^{2/3}$, for example. We have already defined $5^{1/3}$. For rule [3.14](ii) to apply, we must have $5^{2/3} = (5^{1/3})^2$. So we must define $5^{2/3}$ as $\left(\sqrt[3]{5}\right)^2$. In general, for $a > 0$, we define

$$a^{p/q} = \left(a^{1/q}\right)^p = \left(\sqrt[q]{a}\right)^p, \qquad p \text{ an integer, } q \text{ a natural number} \tag{3.16}$$

Note: If q is an odd number and p is an integer, $a^{p/q}$ can be defined even when $a < 0$. For example, $(-8)^{1/3} = \sqrt[3]{-8} = -2$, because $(-2)^3 = -8$. However, in defining $a^{p/q}$ when $a < 0$, the fraction p/q must be reduced to lowest terms. If not, we would get contradictions such as "$-2 = (-8)^{1/3} = (-8)^{2/6} = \sqrt[6]{(-8)^2} = \sqrt[6]{64} = 2$."

Example 3.7

Compute $625^{0.75}$ and $32^{-3/5}$.

Solution $625^{0.75} = 625^{3/4} = (625^{1/4})^3 = 5^3 = 125$

$$32^{-3/5} = \left(32^{1/5}\right)^{-3} = 2^{-3} = 1/8$$

Many scientific calculators have a power key, often denoted by $\boxed{y^x}$. For instance, suppose we let $y = 625$ and $x = 0.75$, then instruct the calculator to compute y^x (the way this is done varies from calculator to calculator). The display may show the number 125.000—or possibly, if 7 decimals are shown, 125.0000001. This shows that the key $\boxed{y^x}$ does not always give an exact answer, even in simple cases. Try it with 2^3, and check the value for $32^{-3/5}$. Simple pocket calculators are usually exact enough for practical purposes, however.

With this definition of $a^{p/q}$, we can show that rules [3.14] are still valid when r and s are rational numbers. In particular,

$$a^{p/q} = \left(a^{1/q}\right)^p = \left(a^p\right)^{1/q} = \sqrt[q]{a^p}$$

Thus, to compute $a^{p/q}$, we could either first take the qth root of a and raise the result to p, or first raise a to the power p and then take the qth root of the result. We obtain the same answer either way. For example,

$$625^{0.75} = 625^{3/4} = \left(625^3\right)^{1/4} = (244140625)^{1/4} = \sqrt[4]{244140625} = 125$$

Note that this procedure involves more difficult computations than the one used in Example 3.7.

Example 3.8

If z denotes demand for coffee in tons per year and p denotes its price per ton, the approximate relationship between them over a specific time period is

$$z = 694{,}500p^{-0.3}$$

(a) Write the formula using roots.

(b) Use a calculator to compute demand when $p = 35{,}000$ and when $p = 55{,}000$.

Solution

(a) $p^{-0.3} = \frac{1}{p^{0.3}} = \frac{1}{p^{3/10}} = \frac{1}{\sqrt[10]{p^3}}$.

so we obtain

$$z = \frac{694,500}{\sqrt[10]{p^3}}$$

(b) $p = 35,000$ gives $z = 694,500 \cdot (35,000)^{-0.3} \approx 30,092$ (tons)

$p = 55,000$ gives $z = 694,500 \cdot (55,000)^{-0.3} \approx 26,276$ (tons)

Note that when price increases, demand decreases.

Using the Power Rules

Powers with rational exponents often occur in economic applications, so you must learn to use them correctly. Before we consider some more examples, note that the power rules can easily be extended to more factors. For instance, we have

$$(abcd)^p = (ab)^p(cd)^p = a^p b^p c^p d^p$$

Example 3.9

Simplify the following expression so that the answer contains only a single exponent for each variable x and y:

$$\left(\frac{5x^{-2}y^{2/3}}{625x^4y^{-4/3}}\right)^{-1/3}$$

Solution One method begins by simplifying the expression inside the parentheses:

$$\left(\frac{5x^{-2}y^{2/3}}{625x^4y^{-4/3}}\right)^{-1/3} = \left(\frac{1}{125} \cdot \frac{x^{-2}}{x^4} \cdot \frac{y^{2/3}}{y^{-4/3}}\right)^{-1/3} = \left(\frac{1}{125} \cdot x^{-6} \cdot y^2\right)^{-1/3}$$

$$= \left(\frac{1}{125}\right)^{-1/3} \left(x^{-6}\right)^{-1/3} \left(y^2\right)^{-1/3} = (125)^{1/3} x^2 y^{-2/3} = \frac{5x^2}{y^{2/3}}$$

Alternatively, we can also raise all the factors to the power $-1/3$ and use the relation $625 = 5^4$ to obtain

$$\left(\frac{5x^{-2}y^{2/3}}{625x^4y^{-4/3}}\right)^{-1/3} = \frac{5^{-1/3}x^{2/3}y^{-2/9}}{(5^4)^{-1/3}x^{-4/3}y^{4/9}} = 5^{-1/3-(-4/3)} \cdot x^{2/3-(-4/3)} \cdot y^{-2/9-4/9}$$

$$= 5^1 x^2 y^{-2/3} = \frac{5x^2}{y^{2/3}}$$

Example 3.10

The formulas for the surface S and the volume V of a ball with radius r are $S = 4\pi r^2$ and $V = (4/3)\pi r^3$. Express S in terms of V.

Solution We must eliminate r. From $V = (4/3)\pi r^3$ we obtain $r^3 = 3V/4\pi$. By raising each side of this equation to the power 1/3 and using $(r^3)^{1/3} = r$, we obtain $r = (3V/4\pi)^{1/3}$. Hence,

$$S = 4\pi r^2 = 4\pi\left[\left(\frac{3V}{4\pi}\right)^{1/3}\right]^2 = 4\pi\frac{(3V)^{2/3}}{(4\pi)^{2/3}}$$
$$= (4\pi)^{1-(2/3)}3^{2/3}V^{2/3} = (4\pi)^{1/3}(3^2)^{1/3}V^{2/3} = \sqrt[3]{36\pi}\,V^{2/3}$$

We have thus shown that

$$S = \sqrt[3]{36\pi}\,V^{2/3} \approx 4.84\,V^{2/3} \qquad [1]$$

Note: Perhaps the most commonly committed error in elementary algebra is to replace $(x+y)^2$ by x^2+y^2 and hence lose the term $2xy$. If we replace $(x+y)^3$ by x^3+y^3, then we lose the terms $3x^2y+3xy^2$. What error do we commit if we replace $(x-y)^3$ by x^3-y^3? Tests also reveal that students who are able to handle these simple power expressions often make mistakes when dealing with more complicated powers. A surprisingly common error is replacing $\left(25-\frac{1}{2}x\right)^{1/2}$ by $25^{1/2}-\left(\frac{1}{2}x\right)^{1/2}$, for example. In general:

> $(x+y)^\alpha$ is usually NOT equal to $x^\alpha + y^\alpha$
> $(x-y-z)^{1/\alpha}$ is usually NOT equal to $x^{1/\alpha} - y^{1/\alpha} - z^{1/\alpha}$

The *only* exception, for general values of x, y, and z, occurs when $\alpha = 1$.

Graphs of Power Functions

We return to the power function $f(x) = x^r$ in [3.13], which is now defined for all rational numbers r provided that $x > 0$. We always have $f(1) = 1^r = 1$, so the graph of the function passes through the point $(1, 1)$ in the xy-plane. The behavior of the graph depends crucially on whether r is positive or negative.

Example 3.11

Sketch the graphs $y = x^{0.3}$ and $y = x^{-1.3}$.

Solution Using a pocket calculator allows us to complete the following table:

x	0	1/3	2/3	1	2	3	4
$y = x^{0.3}$	0	0.72	0.89	1	1.23	1.39	1.52
$y = x^{-1.3}$	*	4.17	1.69	1	0.41	0.24	0.16

*Not defined.

The graphs are shown in Figs. 3.8 and 3.9.

Figure 3.10 illustrates how the graph of $y = x^r$ changes with changing values of the exponent. Try to draw the graphs of $y = x^{-3}$, $y = x^{-1}$, $y = x^{-1/2}$, and $y = x^{-1/3}$.

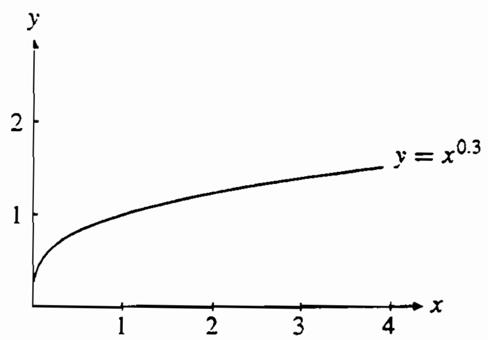


FIGURE 3.8

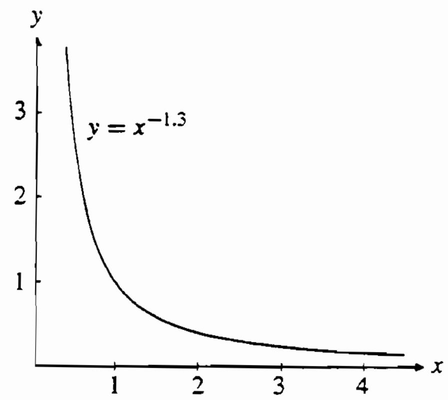


FIGURE 3.9

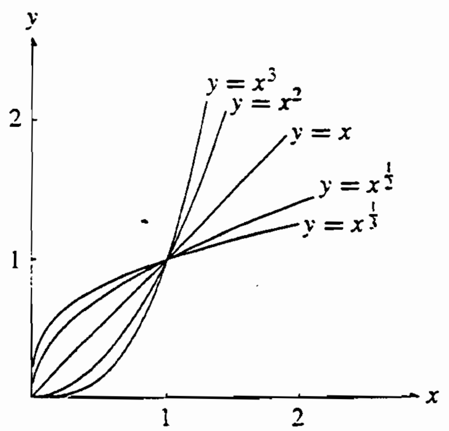


FIGURE 3.10

Problems

1. Compute the following:

   a. $16^{1/4}$ **b.** $243^{-1/5}$ **c.** $5^{1/7} \cdot 5^{6/7}$ **d.** $(4^8)^{-3/16}$

2. Using a pocket calculator or computer, find approximate values for the following:

   a. $100^{1/5}$ b. $16^{-3.33}$ c. $5.23^{1.02} \cdot 2.11^{-3.11}$

3. Compute the following:

   a. $\dfrac{4 \cdot 3^{-1/3}}{\sqrt[6]{81}}$ b. $(0.064)^{-1/3}$ c. $\left(3^2 + 4^2\right)^{-1/2}$

4. How can the number $50^{0.16}$ be expressed as a root?

5. Simplify the following expressions so that each contains only a single exponent of a.

   a. $\{[(a^{1/2})^{2/3}]^{3/4}\}^{4/5}$ b. $a^{1/2}a^{2/3}a^{3/4}a^{4/5}$

   c. $\{[(3a)^{-1}]^{-2}(2a^{-2})^{-1}\}/a^{-3}$ d. $\dfrac{\sqrt[3]{a}\, a^{1/12}\sqrt[4]{a^3}}{a^{5/12}\sqrt{a}}$

6. Solve the following equations for x:

   a. $2^{2x} = 8$ b. $3^{3x+1} = 1/81$ c. $10^{x^2-2x+2} = 100$

7. Which of the following equations are valid for all x and y?

   a. $(2^x)^2 = 2^{x^2}$ b. $3^{x-3y} = \dfrac{3^x}{3^{3y}}$

   c. $3^{-1/x} = \dfrac{1}{3^{1/x}}$ $(x \neq 0)$ d. $5^{1/x} = \dfrac{1}{5^x}$ $(x \neq 0)$

   e. $a^{x+y} = a^x + a^y$ f. $2^{\sqrt{x}} \cdot 2^{\sqrt{y}} = 2^{\sqrt{xy}}$ (x and y positive)

8. Solve the following equations for the variables indicated:

   a. $3K^{-1/2}L^{1/3} = 1/5$ for K
   b. $p - abx_0^{b-1} = 0$ for x_0
   c. $ax(ax+b)^{-2/3} + (ax+b)^{1/3} = 0$ for x
   d. $\left[(1-\lambda)a^{-\rho} + \lambda b^{-\rho}\right]^{-1/\rho} = c$ for b

9. A sphere of capacity 100 m^3 is to have its outside surface painted. One liter of paint covers 5 m^2. How many liters of paint are needed? (*Hint:* Use formula [1] in Example 3.10.)

10. Show by using a pocket calculator (or a computer) that the equation

$$Y = 2.262K^{0.203}L^{0.763}(1.02)^t$$

has an approximate solution for K given by $K \approx 0.018Y^{4.926}L^{-3.759}(0.907)^t$. Then determine K numerically when $Y = 100$, $L = 6$, and $t = 10$.

11. Simplify the following expressions:

   a. $\left(a^{1/3} - b^{1/3}\right)\left(a^{2/3} + a^{1/3}b^{1/3} + b^{2/3}\right)$

   b. $\dfrac{bx^{1/2} - (x-a)b\frac{1}{2}x^{-1/2}}{\left(bx^{1/2}\right)^2}$ $(x > 0)$

3.5 Exponential Functions

A quantity that increases (or decreases) by a fixed factor per unit of time is said to *increase* (or decrease) *exponentially*. If this fixed factor is a, this leads to the study of the exponential function f defined by

$$f(t) = Aa^t \tag{3.17}$$

where a and A are positive constants. Note that if $f(t) = Aa^t$, then $f(t+1) = Aa^{t+1} = Aa^t \cdot a^1 = af(t)$, so the value of f at time $t+1$ is a times the value of f at time t. If $a > 1$, then f is increasing; if $0 < a < 1$, then f is decreasing. Because $f(0) = Aa^0 = A$, we can write $f(t) = f(0)a^t$.

Exponential functions appear in many important economic, social, and physical models. For instance, economic growth, population growth, continuously accumulated interest, radioactive decay, and decreasing illiteracy have all been described by exponential functions. In addition, the exponential function is one of the most important in statistics.

Example 3.12 (Population Growth)

Consider a growing population like that of Europe. In Example 2.13, we constructed a linear function

$$P = 6.4t + 641$$

where P denotes the population in millions, $t = 0$ corresponds to the year 1960 when the population was 641 million, and $t = 10$ corresponds to the year 1970 when the population estimate was 705 million. According to this formula, the annual increase in population would be constant and equal to 6.4 million. This is a very unreasonable assumption. After all, populations tend to grow faster as they get bigger because there are more people to have babies, and the death rate usually decreases or stays the same. In fact, according to UN estimates, the European population was expected to grow by approximately 0.72% annually during the period 1960 to 2000. With a population of 641 million in 1960, the population in 1961 would then be

$$641 + \frac{641 \cdot 0.72}{100} = 641 \cdot \left(1 + \frac{0.72}{100}\right) = 641 \cdot 1.0072$$

which is approximately 645 million. Next year, in 1962, it would have grown to

$$\begin{aligned} 641 \cdot 1.0072 + \frac{641 \cdot 1.0072 \cdot 0.72}{100} &= 641 \cdot 1.0072 \cdot (1 + 0.0072) \\ &= 641 \cdot 1.0072^2 \end{aligned}$$

which is approximately 650 million. Note how the population figure grows by the factor 1.0072 each year. If the growth rate were to continue at 0.72% annually, then t years after 1960 the population would be given by

$$P(t) = 641 \cdot 1.0072^t \qquad [1]$$

Thus, $P(t)$ is an exponential function of the form [3.17]. For the year 2000, corresponding to $t = 40$, the formula yields the estimate $P(40) \approx 854$ million.

Many countries, particularly in Africa and Latin America, have recently had far faster population growth than Europe. For instance, during the 1970s and 1980s, the growth rate of Zimbabwe's population was close to 3.5% annually. If we let $t = 0$ correspond to the census year 1969 when the population was 5.1 million, the population t years after 1969 is given by

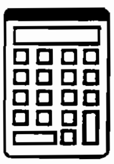
$$P(t) = 5.1 \cdot 1.035^t$$

If we calculate $P(20)$, $P(40)$, and $P(60)$ using this formula, we get roughly 10, 20, and 40. Thus, the population of Zimbabwe roughly doubles after 20 years; during the next 20 years, it doubles again, and so on. We say that the *doubling time* of the population is approximately 20 years. Of course, extrapolating so far into the future is quite dubious, because exponential growth of population cannot go on forever. (If the growth rate continued at 3.5% annually, and the Zimbabwean territory did not expand, in the year 2697, each Zimbabwean would on average have only 1 square meter of land. See Problem 7.)

If $a > 1$ and $A > 0$, the exponential function $f(t) = Aa^t$ is increasing. Its **doubling time** is the time it takes for it to double. Its value at $t = 0$ is A, so the doubling time t^* is given by the equation $f(t^*) = Aa^{t^*} = 2A$, or after cancelling A, by $a^{t^*} = 2$. Thus the doubling time of the exponential function $f(t) = Aa^t$ is the power to which a must be raised in order to get 2.[2] (In Problem 8 you will be asked to show that the doubling time is independent of which year you take as the base.)

Example 3.13

Use your calculator to find the doubling time of



(a) a population (like that of Zimbabwe) increasing at 3.5% annually (thus confirming the earlier calculations)

(b) the population of Kenya in the 1980s (which had the world's highest annual growth rate of 4.2%).

Solution

(a) The doubling time t^* is given by the equation $1.035^{t^*} = 2$. Using a calculator shows that $1.035^{15} \approx 1.68$, whereas $1.035^{25} \approx 2.36$. Thus,

[2]By using natural logarithms as explained in Section 8.2, we find that $t^* = \ln 2 / \ln a$.

t^* must lie between 15 and 25. Because $1.035^{20} \approx 1.99$, t^* is close to 20. In fact, $t^* \approx 20.15$.

(b) The doubling time t^* is given by the equation $1.042^{t^*} = 2$. Using a calculator, we find that $t^* \approx 16.85$. Thus, with a growth rate of 4.2%, Kenya's population would double in less than 17 years.

Example 3.14 (Compound Interest)
A savings account of $\$K$ that increases by $p\%$ interest each year will have increased after t years to

$$K(1 + p/100)^t \qquad [1]$$

(see Section A.1 of Appendix A). According to this formula, \$1 ($K = 1$) earning interest at 8% per annum ($p = 8$) will have increased after t years to

$$(1 + 8/100)^t = 1.08^t \qquad [2]$$

Table 3.1 indicates how this dollar grows over time:

TABLE 3.1 *How \$1 of savings increases with time*

t	1	2	5	10	20	30	50	100	200
$(1.08)^t$	1.08	1.17	1.47	2.16	4.66	10.06	46.90	2,199.76	4,838,949.60

After 30 years, \$1 of savings has increased to more than \$10, and after 200 years, it has grown to more than \$4.8 million! This growth is illustrated in Fig. 3.11. Observe that the expression 1.08^t defines an exponential function of the type [3.17] with $a = 1.08$. Even if a is only slightly larger than 1, $f(t)$ will increase very quickly when t is large.

Example 3.15 (Radioactive Decay)
Measurements indicate that radioactive materials decay by a fixed percentage per unit of time. Plutonium 239, which is a waste product of certain nuclear power plants and is used in the production of nuclear weapons, decays by 50% every 24,400 years. We say, therefore, that the *half-life* of plutonium 239 is 24,400 years. If there are I_0 units of plutonium 239 at time $t = 0$, then after t years, there will be

$$I(t) = I_0 \cdot \left(\tfrac{1}{2}\right)^{t/24,400} = I_0 \cdot 0.9999716^t$$

units remaining. (Observe that this is consistent with $I(24,400) = \frac{1}{2}I_0$.)

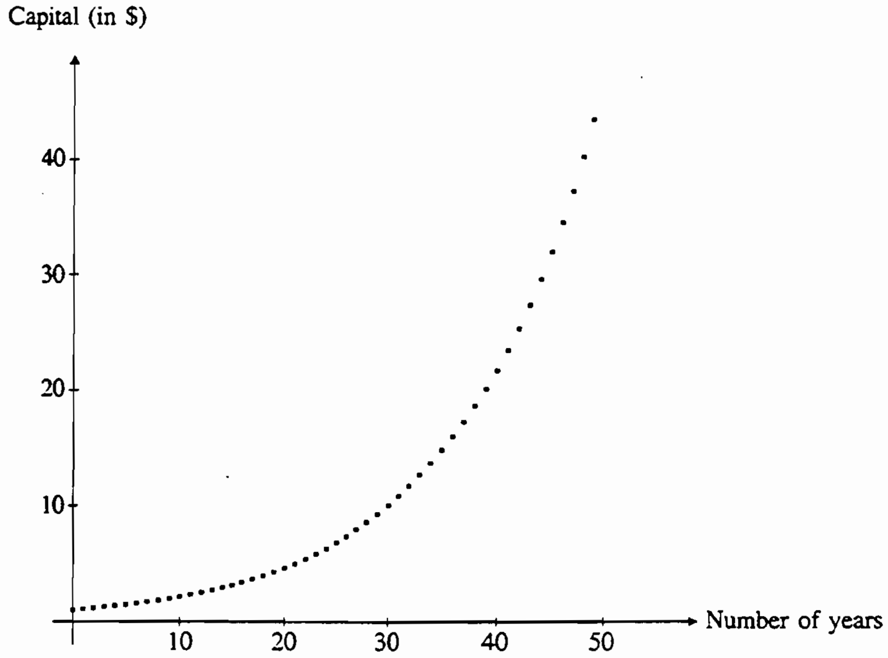


FIGURE 3.11 The growth of $1 of savings after t years when the interest rate is 8% per year.

Chapter 8 discusses the exponential function in much greater detail. Observe the fundamental difference between the two functions

$$f(x) = a^x \qquad \text{and} \qquad g(x) = x^a$$

The second of these two is the **power function** discussed in Section 3.4. For the exponential function a^x, it is the exponent that varies, while the base is constant. For the power function x^a, on the other hand, the exponent is constant, while the base varies.

The most important properties of the exponential function are summed up by the following:

The **general exponential function** with base $a > 0$ is

$$f(x) = Aa^x$$

where $f(0) = A$, and a is the factor by which $f(x)$ changes when x increases by 1.

If $a = 1 + p/100$, where $p > 0$ and $A > 0$, then $f(x)$ will increase by $p\%$ for each unit increase in x.

If $a = 1 - p/100$, where $p > 0$ and $A > 0$, then $f(x)$ will decrease by $p\%$ for each unit increase in x.

Problems

1. If the population of Europe grew at the rate of 0.72% annually, what would be the doubling time?
2. The population of Botswana was estimated to be 1.22 million in 1989, and to be growing at the rate of 3.4% annually.
   a. If $t = 0$ denotes 1989, find a formula for the population at date t.
   b. What is the doubling time?
3. A savings account with an initial deposit of \$100 earns 12% interest per year.
   a. What is the amount of savings after t years?
   b. Make a table similar to Table 3.1. (Stop at 50 years.)
4. Suppose that you are promised \$2 on the first day, \$4 on the second day, \$8 on the third day, \$16 on the fourth day, and so on (so that every day you get twice as much as the day before).
   a. How much will you receive on the tenth day?
   b. Find a function $f(t)$ that indicates how much you will obtain on the tth day.
   c. Explain why $f(20)$ is more than \$1 million. (*Hint:* 2^{10} is a little larger than 10^3.)
5. Fill in the following table and then make a rough sketch of the graphs of $y = 2^x$ and $y = 2^{-x}$.

x	-3	-2	-1	0	1	2	3
2^x							
2^{-x}							

6. Fill in the following table and then sketch the graph of $y = 2^{x^2}$.

x	-2	-1	0	1	2
2^{x^2}					

7. The area of Zimbabwe is approximately $3.91 \cdot 10^{11}$ square meters. Referring to the text following Example 3.12 and using a calculator, solve the equation $5.1 \cdot 1.035^t = 3.91 \cdot 10^{11}$ for t, and interpret your answer. (Recall that $t = 0$ corresponds to 1969.)
8. With $f(t) = Aa^t$, if $f(t + t^*) = 2f(t)$, prove that $a^{t^*} = 2$. Explain why this shows that the doubling time of the general exponential function is independent of the initial time.

9. In 1964 a five-year plan was introduced in Tanzania. One objective was to double the real per capita income over the next 15 years. What is the average annual rate of growth of real income per capita required to achieve this objective?
10. Consider the function f defined for all x by $f(x) = 1 - 2^{-x}$.
    a. Make a table of function values for $x = 0, \pm 1, \pm 2$, and ± 3. Then sketch the graph of f.
    b. What happens to $f(x)$ as x becomes very large and very small?
11. Which of the following equations do *not* define exponential functions of x?

    a. $y = 3^x$ b. $y = x^{\sqrt{2}}$ c. $y = (\sqrt{2})^x$
    d. $y = x^x$ e. $y = (2.7)^x$ f. $y = 1/2^x$
12. Fill in the following table and then sketch the graph of $y = x^2 2^x$.

x	−10	−5	−4	−3	−2	−1	0	1	2
$x^2 2^x$									

13. Find possible exponential functions for the graphs of Fig. 3.12.
14. The radioactive isotope iodine 131, which has a half-life of 8 days, is often used to diagnose disease in the thyroid gland. If there are I_0 units of the material at time $t = 0$, how much remains after t days?

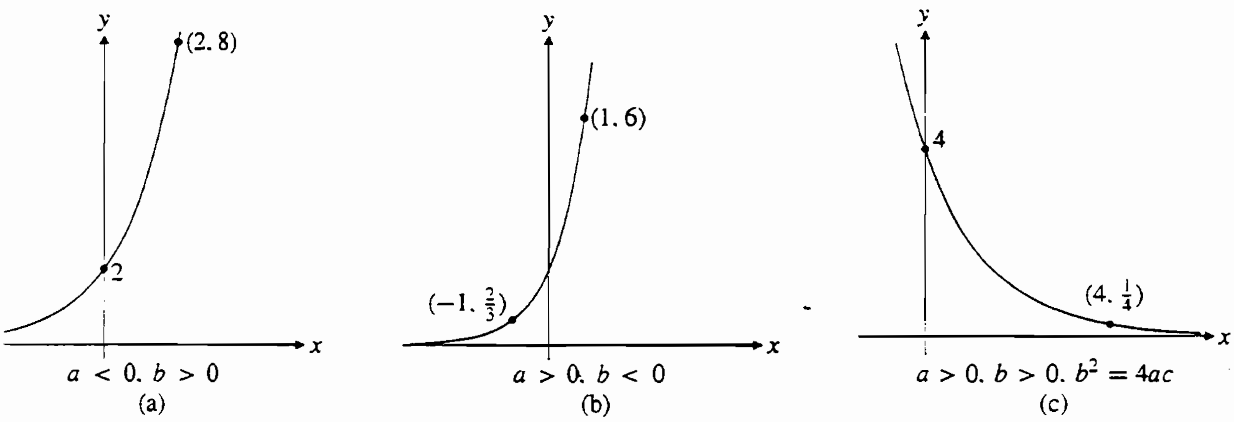


FIGURE 3.12

3.6 The General Concept of a Function

So far we have studied functions of one variable. These are functions whose domain is a set of real numbers, and whose range is also a set of real numbers. Yet a realistic description of many economic phenomena requires considering a large number of variables simultaneously. For example, the demand for a good like butter is a function of several variables such as the price of the good, the prices of complements and substitutes, consumers' incomes, and so on.

Actually, we have already seen many special functions of several variables. For instance, the formula $V = \pi r^2 h$ for the volume V of a cylinder with base radius r and height h involves a function of two variables. (Of course, in this case $\pi \approx 3.14159$ is a mathematical constant.) A change in one of these variables will not affect the value of the other variable. For each pair of positive numbers (r, h), there is a definite value for the volume V. To emphasize that V depends on the values of both r and h, we write

$$V(r, h) = \pi r^2 h$$

For $r = 2$ and $h = 3$, we obtain $V(2, 3) = 12\pi$, whereas $r = 3$ and $h = 2$ give $V(3, 2) = 18\pi$. Also, $r = 1$ and $h = 1/\pi$ give $V(1, 1/\pi) = 1$. Note in particular that $V(2, 3) \neq V(3, 2)$.

In some abstract economic models, it may be enough to know that there is some functional relationship between variables, without specifying the dependence more closely. For instance, suppose a market sells three commodities whose prices per unit are respectively p, q, and r. Then economists generally assume that the demand for one of the commodities by an individual with income m is given by a function $f(p, q, r, m)$ of four variables, without specifying the precise form of that function.

An extensive discussion of functions of several variables begins in Chapter 15. This section introduces an even more general type of function. In fact, general functions of the kind presented here are of fundamental importance in practically every area of pure and applied mathematics, including mathematics applied to economics.

Example 3.16

The following examples indicate how very wide is the concept of a function.

(a) The function that assigns to each triangle in a plane the area of that triangle (measured, say, in cm^2).

(b) The function that determines the social security number, or other identification number, of each taxpayer.

(c) The function that for each point P in a plane determines the point lying 3 units above P.

(d) Let A be the set of possible actions that a person can choose in a certain situation. Suppose that every action $a \in A$ produces a certain result (say, a certain profit $\varphi(a)$). In this way, we have defined a function φ with domain A.

Here is a general definition:

> A **function** from A to B is a rule that assigns to each element of the set A one and only one element of the set B. [3.18]

If we denote the function by f, the set A is called the **domain** of f, and B is called the **target**. The two sets A and B need not consist of numbers, but can be sets of quite arbitrary elements.

The definition of a function requires three objects to be specified:

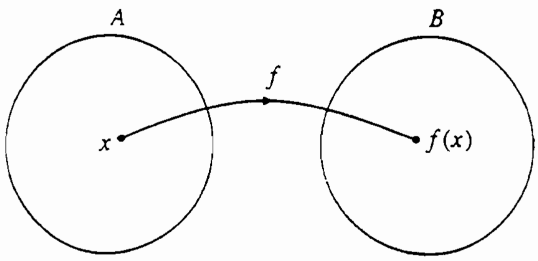
1. A domain A
2. A target B
3. A rule that assigns a *unique* element in B to *each* element in A.

Nevertheless, in many cases, we refrain from specifying the sets A and/or B explicitly when it obvious from the context what these sets are.

An important requirement in the definition of a function is that to each element in domain A, there corresponds a *unique* element in target B. While it is meaningful to talk about the function that assigns the natural mother to every child, the rule that assigns the aunt to any child does not, in general, define a function, because many children have several aunts. Explain why the following rule, as opposed to the one in Example 3.16(c), does not define a function: "to a point P in a horizontal plane, assign a point that lies 3 units from P."

If f is a function with domain A and target B, we often say that f is a **function from A to B**, and write $f : A \to B$. The functional relationship is often represented as in Fig. 3.13. Other words that are sometimes used instead of "function" include **transformation** and **map** or **mapping**. The particular value $f(x)$ is often called the **image** of the element x by the function f. The set of elements in B that are images of at least one element in A is called

FIGURE 3.13 A function from A to B.



the **range** of the function. Thus, the range is a subset of the target. If we denote the range of f by R_f, then $R_f = \{f(x) : x \in A\}$. This is also written as $f(A)$. The range of the function in Example 3.16(a) is the set of all positive numbers. Explain why the range of the function in (c) must be the entire plane.

The definition of a function requires that only *one* element in B be assigned to each element in A. However, different elements in A might be mapped to the same element in B. In Example 3.16(a), for instance, many different triangles have the same area. If each element of B is the image of at most one element in A, function f is called **one-to-one**. Otherwise, if one or more elements of B are the images of more than one element in A, the function is many-to-one.

The social security function in Example 3.16(b) is intended to be one-to-one, because two different taxpayers should always have different numbers. (In very rare instances, errors cause this function to be many-to-one. These always create a great deal of confusion when they are noticed!) Can you explain why the function defined in Example 3.16(c) is also one-to-one, whereas the function that assigns to each child his or her mother is not?

Suppose f is a one-to-one function from a set A to a set B, and assume that the range of f is all of B. Thus:

1. f maps each element of A into an element of B (so f is a function).
2. Two different elements of A are always mapped into different elements of B (so f is one-to-one).
3. For each element v in B, there is an element u in A such that $f(u) = v$ (so the range of f is the whole of B).

We can then define a function g from B to A by the following obvious rule: Assign to each element v of B the element $u = g(v)$ of A that f maps to v—that is, the u satisfying $v = f(u)$. Because of rule 2, there can be only one u in A such that $v = f(u)$, so g is a function. Its domain is B and its target and range are both equal to A. The function g is called the **inverse function** of f. For instance, the inverse of the social security function mentioned earlier is the function that, to each social security number, assigns the person carrying that number. Section 7.6 provides more detail about inverse functions and their properties.

Problems

1. Decide which of the following rules defines a function:
   a. The rule that assigns to each person in a classroom his or her height.
   b. The rule that assigns to a mother her youngest child.
   c. The rule that assigns the circumference of a rectangle to its area.
   d. The rule that assigns the surface area of a spherical ball to its volume.

e. The rule that assigns the pair of numbers $(x+3, y)$ to the pair of numbers (x, y).

2. Decide which of the functions defined in Problem 1 is one-to-one, and which then have an inverse. Determine the inverse when it exists.

3. Each person has red blood cells that belong to one and only one of four blood groups denoted A, B, AB, and O. Consider the function that assigns each person in a team to his or her blood group. Can this function be one-to-one if the team consists of at least five persons?