Chapter 1

A Context for Calculus

Calculus gives us a language to describe how quantities are related to one another, and it gives us a set of computational and visual tools for exploring those relationships. Usually, we want to understand how quantities are related in the context of a particular problem—it might be in chemistry, or public policy, or mathematics itself. In this chapter we take a single context—an infectious disease spreading through a population—to see how calculus emerges and how it is used.

1.1 The Spread of Disease

Making a Model

Many human diseases are contagious: you "catch" them from someone who is already infected. Contagious diseases are of many kinds. Smallpox, polio, and plague are severe and even fatal, while the common cold and the childhood illnesses of measles, mumps, and rubella are usually relatively mild. Moreover, you can catch a cold over and over again, but you get measles only once. A disease like measles is said to "confer immunity" on someone who recovers from it. Some diseases have the potential to affect large segments of a population; they are called *epidemics* (from the Greek words *epi*, upon + *demos*, the people.) *Epidemiology* is the scientific study of these diseases.

Some properties of contagious diseases

An epidemic is a complicated matter, but the dangers posed by contagion—and especially by the appearance of new and uncontrollable diseases—compel us to learn as much as we can about the nature of epidemics. Mathematics offers a very special kind of help. First, we can try to draw out of the situation its essential features and describe them mathematically. This is calculus as *language*. We substitute an "ideal" mathematical world for the real one. This mathematical world is called a **model**. Second, we can use mathematical insights and methods to analyze the model. This is calculus as *tool*. Any conclusion we reach about the model can then be interpreted to tell us something about the reality.

The idea of a mathematical model

To give you an idea how this process works, we'll build a model of an epidemic. Its basic purpose is to help us understand the way a contagious disease spreads through a population—to the point where we can even predict what fraction falls ill, and when. Let's suppose the disease we want to model is like measles. In particular,

- it is mild, so anyone who falls ill eventually recovers;
- it confers permanent immunity on every recovered victim.

In addition, we will assume that the affected population is large but fixed in size and confined to a geographically well-defined region. To have a concrete image, you can imagine the elementary school population of a big city.

At any time, that population can be divided into three distinct classes:

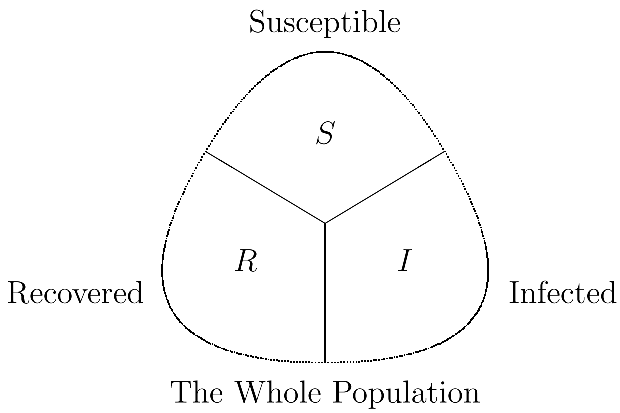
Susceptible: those who have never had the illness and can catch it;

Infected: those who currently have the illness and are contagious;

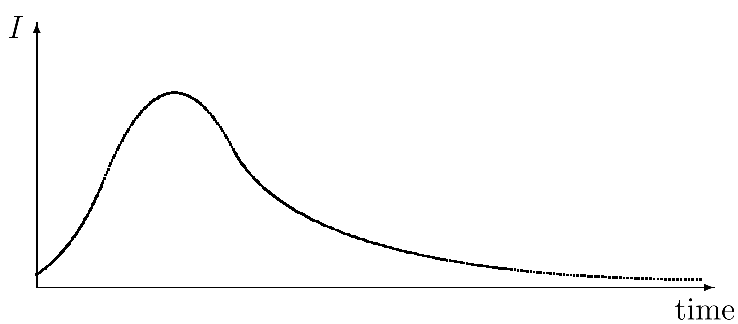
Recovered: those who have already had the illness and are immune.

The quantities that our model analyzes

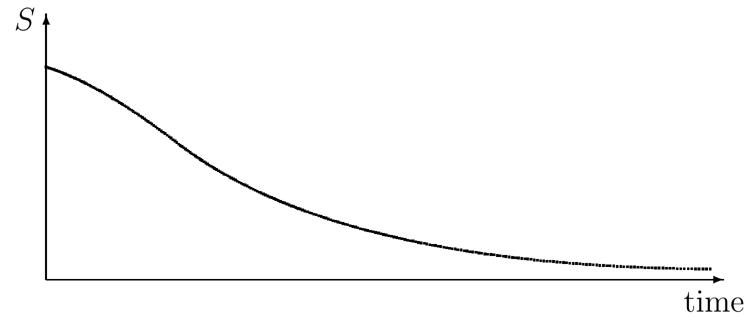
Suppose we let S, I, and R denote the number of people in each of these three classes, respectively. Of course, the classes are all mixed together throughout the population: on a given day, we may find persons who are susceptible, infected, and recovered in the same family. For the purpose of organizing our thinking, though, we'll represent the whole population as separated into three "compartments" as in the following diagram:

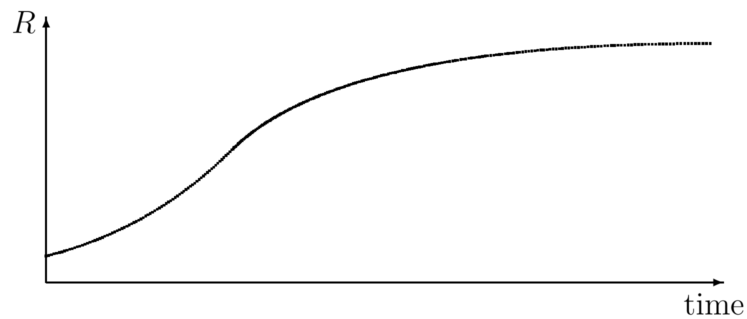


The goal of our model is to determine what happens to the numbers S, I, and R over the course of time. Let's first see what our knowledge and experience of childhood diseases might lead us to expect. When we say there is a "measles outbreak," we mean that there is a relatively sudden increase in the number of cases, and then a gradual decline. After some weeks or months, the illness virtually disappears. In other words, the number I is a **variable**; its value changes over time. One possible way that I might vary is shown in the following graph.



During the course of the epidemic, susceptibles are constantly falling ill. Thus we would expect the number S to show a steady decline. Unless we know more about the illness, we cannot decide whether everyone eventually catches it. In graphical terms, this means we don't know whether the graph of S levels off at zero or at a value above zero. Finally, we would expect more and more people in the recovered group as time passes. The graph of R should therefore climb from left to right. The graphs of S and R might take the following forms:





Some quantitative questions that the graphs raise

While these graphs give us an idea about what might happen, they raise some new questions, too. For example, because there are no scales marked along the axes, the first graph does not tell us how large I becomes when the infection reaches its peak, nor when that peak occurs. Likewise, the second and third graphs do not say how rapidly the population either falls ill or recovers. A good model of the epidemic should give us graphs like these and it should also answer the quantitative questions we have already raised—for example: When does the infection hit its peak? How many susceptibles eventually fall ill?

A Simple Model for Predicting Change

Suppose we know the values of S, I, and R today; can we figure out what they will be tomorrow, or the next day, or a week or a month from now? Basically, this is a problem of predicting the future. One way to deal with it is to get an idea how S, I, and R are *changing*. To start with a very simple example, suppose the city's Board of Health reports that the measles infection has been spreading at the rate of 470 new cases per day for the last several days. If that rate continues to hold steady and we start with 20,000 susceptible children, then we can expect 470 fewer susceptibles with each passing day. The immediate future would then look like this:

days after today	accumulated number of new infections	remaining number of susceptibles
0	0	20 000
1	470	19 530
2	940	19 060
3	1410	18 590
⋮	⋮	⋮

Of course, these numbers will be correct only if the infection continues to spread at its present rate of 470 persons per day. If we want to follow S, I, and R into the future, our example suggests that we should pay attention to the **rates** at which these quantities change. To make it easier to refer to them, let's denote the three rates by S', I', and R'. For example, in the illustration above, S is changing at the rate $S' = -470$ persons per day. We use a minus sign here because S is *decreasing* over time. If S' stays fixed we can express the value of S after t days by the following formula:

Knowing rates, we can predict future values

$$S = 20000 + S' \cdot t = 20000 - 470\,t \text{ persons.}$$

Check that this gives the values of S found in the table when $t = 0$, 1, 2, or 3. How many susceptibles does it say are left after 10 days?

Our assumption that $S' = -470$ persons per day amounts to a mathematical characterization of the susceptible population—in other words, a model! Of course it is quite simple, but it led to a formula that told us what value we could expect S to have at any time t.

The equation $S' = -470$ is a model

The model will even take us backwards in time. For example, two days ago the value of t was -2; according to the model, there were

$$S = 20000 - 470 \times -2 = 20940$$

susceptible children then. There is an obvious difference between going backwards in time and going forwards: we already know the past. Therefore, by letting t be negative we can generate values for S that can be checked against health records. If the model gives good agreement with *known* values of S we become more confident in using it to predict *future* values.

Predictions depend on the initial value, too

To predict the value of S using the rate S' we clearly need to have a starting point—a known value of S from which we can measure changes. In our case that starting point is $S = 20000$. This is called the **initial value** of S, because it is given to us at the "initial time" $t = 0$. To construct the formula $S = 20000 - 470\,t$, we needed to have an initial value as well as a rate of change for S.

In the following pages we will develop a more complex model for all three population groups that has the same general design as this simple one. Specifically, the model will give us information about the rates S', I', and R', and with that information we will be able to predict the values of S, I, and R at any time t.

The Rate of Recovery

Our first task will be to model the recovery rate R'. We look at the process of recovering first, because it's simpler to analyze. An individual caught in the epidemic first falls ill and then recovers—recovery is just a matter of time. In particular, someone who catches measles has the infection for about fourteen days. So if we look at the entire infected population today, we can expect to find some who have been infected less than one day, some who have been infected between one and two days, and so on, up to fourteen days. Those in the last group will recover today. In the absence of any definite information about the fourteen groups, let's assume they are the same size. Then 1/14-th of the infected population will recover today:

$$\text{today the change in the recovered population} = \frac{I \text{ persons}}{14 \text{ days}}.$$

There is nothing special about today, though; I has a value at any time. Thus we can make the same argument about any other day:

$$\text{every day the change in the recovered population} = \frac{I \text{ persons}}{14 \text{ days}}.$$

This equation is telling us about R', the rate at which R is changing. We can write it more simply in the form

The first piece of the S-I-R model

$$R' = \frac{1}{14} I \text{ persons per day.}$$

We call this a **rate equation**. Like any equation, it links different quantities together. In this case, it links R' to I. The rate equation for R is the first part of our model of the measles epidemic.

Are you uneasy about our claim that 1/14-th of the infected population recovers every day? You have good reason to be. After all, during the first few days of the epidemic almost no one has had measles the full fourteen days, so the recovery rate will be much less than $I/14$ persons per day. About a week before the infection disappears altogether there will be no one in the early stages of the illness. The recovery rate will then be much greater than $I/14$ persons per day. Evidently our model is not a perfect mirror of reality!

Don't be particularly surprised or dismayed by this. Our goal is to gain insight into the workings of an epidemic and to suggest how we might intervene to reduce its effects. So we start off with a model which, while imperfect, still captures some of the workings. The simplifications in the model will be justified if we are led to inferences which help us understand how an epidemic works and how we can deal with it. If we wish, we can then refine the model, replacing the simple expressions with others that mirror the reality more fully.

Notice that the rate equation for R' does indeed give us a tool to predict future values of R. For suppose today 2100 people are infected and 2500 have already recovered. Can we say how large the recovered population will be tomorrow or the next day? Since $I = 2100$,

$$R' = \frac{1}{14} \times 2100 = 150 \text{ persons per day}.$$

Thus 150 people will recover in a single day, and twice as many, or 300, will recover in two. At this rate the recovered population will number 2650 tomorrow and 2800 the next day.

These calculations assume that the rate R' holds steady at 150 persons per day for the entire two days. Since $R' = I/14$, this is the same as assuming that I holds steady at 2100 persons. If instead I varies during the two days we would have to adjust the value of R' and, ultimately, the future values of R as well. In fact, I *does* vary over time. We shall see this when we analyze how the infection is transmitted. Then, in chapter 2, we'll see how to make the adjustments in the values of R' that will permit us to predict the value of R in the model with as much accuracy as we wish.

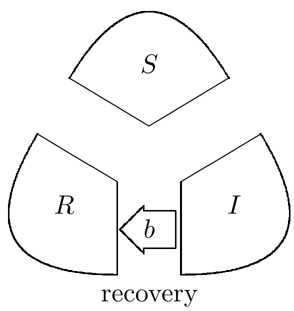
Other diseases. What can we say about the recovery rate for a contagious disease other than measles? If the period of infection of the new illness is k days, instead of 14, and if we assume that $1/k$ of the infected people recover each day, then the new recovery rate is

$$R' = \frac{I \text{ persons}}{k \text{ days}} = \frac{1}{k} I \text{ persons per day.}$$

If we set $b = 1/k$ we can express the recovery rate equation in the form

$$R' = bI \text{ persons per day.}$$

The constant b is called the **recovery coefficient** in this context.



Let's incorporate our understanding of recovery into the compartment diagram. For the sake of illustration, we'll separate the three compartments. As time passes, people "flow" from the infected compartment to the recovered. We represent this flow by an arrow from I to R. We label the arrow with the recovery coefficient b to indicate that the flow is governed by the rate equation $R' = bI$.

The Rate of Transmission

Since susceptibles become infected, the compartment diagram above should also have an arrow that goes from S to I and a rate equation for S' to show how S changes as the infection spreads. While R' depends only on I, because recovery involves only waiting for people to leave the infected population, S' will depend on both S and I, because transmission involves contact between susceptible and infected persons.

Here's a way to model the transmission rate. First, consider a single susceptible person on a single day. On average, this person will contact only a small fraction, p, of the infected population. For example, suppose there are 5000 infected children, so $I = 5000$. We might expect only a couple of them—let's say 2—will be in the same classroom with our "average" susceptible. So the fraction of contacts is $p = 2/I = 2/5000 = .0004$. The 2 contacts themselves can be expressed as $2 = (2/I) \cdot I = pI$ contacts per day per susceptible.

Contacts are proportional to both S and I

To find out how many daily contacts the *whole* susceptible population will have, we can just multiply the average number of contacts per susceptible person by the number of susceptibles: this is $pI \cdot S = pSI$.

Not all contacts lead to new infections; only a certain fraction q do. The more contagious the disease, the larger q is. Since the number of daily contacts is pSI, we can expect $q \cdot pSI$ new infections per day (i.e., to convert contacts to infections, multiply by q). This becomes aSI if we define a to be the product qp.

Recall, the value of the recovery coefficient b depends only on the illness involved. It is the same for all populations. By contrast, the value of a depends on the general health of a population and the level of social interaction between its members. Thus, when two different populations experience the same illness, the values of a could be different. One strategy for dealing with an epidemic is to alter the value of a. Quarantine does this, for instance; see the exercises.

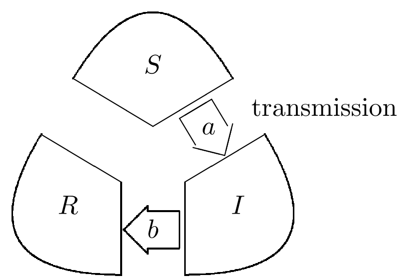
Since each new infection decreases the number of susceptibles, we have the rate equation for S:

$$S' = -aSI \text{ persons per day}.$$

Here is the second piece of the S-I-R model

The minus sign here tells us that S is decreasing (since S and I are positive). We call a the **transmission coefficient**.

Just as people flow from the infected to the recovered compartment when they recover, they flow from the susceptible to the infected when they fall ill. To indicate the second flow let's add another arrow to the compartment diagram. Because this flow is due to the transmission of the illness, we will label the arrow with the transmission coefficient a. The compartment diagram now reflects all aspects of our model.



We haven't talked about the units in which to measure a and b. They must be chosen so that any equation in which a or b appears will balance. Thus, in $R' = bI$ the units on the left are persons/day; since the units for I are persons, the units for b must be 1/(days). The units in $S' = -aSI$ will balance only if a is measured in 1/(person-day).

The reciprocals have more natural interpretations. First of all, $1/b$ is the number of days a person needs to recover. Next, note that $1/a$ is measured in person-days (i.e., persons $\times$ days), which are the natural units in which to measure exposure. Here is why. Suppose you contact 3 infected persons for each of 4 days. That gives you the same exposure to the illness that you get from 6 infected persons in 2 days—both give 12 "person-days" of exposure. Thus, we can interpret $1/a$ as the level of exposure of a typical susceptible person.

Completing the Model

The final rate equation we need—the one for I'—reflects what is already clear from the compartment diagram: every loss in I is due to a gain in R, while every gain in I is due to a loss in S.

Here is the complete S-I-R model

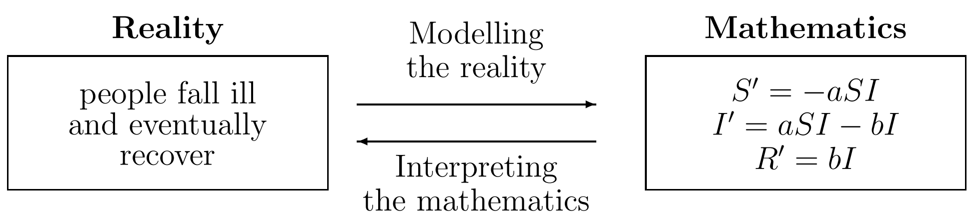
$$\begin{aligned} S' &= -aSI, \\ I' &= aSI - bI, \\ R' &= \phantom{-aSI - {}} bI. \end{aligned}$$

If you add up these three rates you should get the overall rate of change of the whole population. The sum is zero. Do you see why?

You should not draw the conclusion that the only use of rate equations is to model an epidemic. Rate equations have a long history, and they have been put to many uses. Isaac Newton (1642–1727) introduced them to model the motion of a planet around the sun. He showed that the same rate equations could model the motion of the moon around the earth and the motion of an object falling to the ground. Newton created calculus as a tool to analyze these equations. He did the work while he was still an undergraduate—on an extended vacation, curiously enough, because a plague epidemic was raging at Cambridge!

Today we use Newton's rate equations to control the motion of earth satellites and the spacecraft that have visited the moon and the planets. We use other rate equations to model radioactive decay, chemical reactions, the growth and decline of populations, the flow of electricity in a circuit, the change in air pressure with altitude—just to give a few examples. You will have an opportunity in the following chapters to see how they arise in many different contexts, and how they can be analyzed using the tools of calculus.

The following diagram summarizes, in a schematic way, the relation between our model and the reality it seeks to portray.



The model is part of mathematics; it only approximates reality

The diagram calls attention to several facts. First, the model is a part of mathematics. It is distinct from the reality being modelled. Second, the model is based on a simplified interpretation of the epidemic. As such, it will not match the reality exactly; it will be only an **approximation**. Thus, we cannot expect the values of S, I, and R that we calculate from the rate equations to give us the exact sizes of the susceptible, infected, and recovered populations. Third, the connection between reality and mathematics is a two-way street. We have already travelled one way by constructing a mathematical object that reflects some aspects of the epidemic. This is model-building. Presently we will travel the other way. First we need to get mathematical answers to mathematical questions; then we will see what those answers tell us about the epidemic. This is interpretation of the model. Before we begin the interpretation, we must do some mathematics.

Analyzing the Model

Now that we have a model we shall analyze it as a mathematical object. We will set aside, at least for the moment, the connection between the mathematics and the reality. Thus, for example, you should not be concerned when our calculations produce a value for S of 44,446.6 persons at a certain time—a value that will never be attained in reality. In the following analysis S is just a numerical quantity that varies with t, another numerical quantity. Using only mathematical tools we must now extract from the rate equations the information that will tell us just how S and I and R vary with t.

We already took the first steps in that direction when we used the rate equation $R' = I/14$ to predict the value of R two days into the future (see page 7). We assumed that I remained fixed at 2100 during those two days, so the rate $R' = 2100/14 = 150$ was also fixed. We concluded that if $R = 2500$ today, it will be 2650 tomorrow and 2800 the next day.

A glance at the full S-I-R model tells us those first steps have to be modified. The assumption we made—that I remains fixed—is not justified, because I (like S and R) is continually changing. As we shall see, I actually increases over those two days. Hence, over the same two days, R' is not fixed at 150, but is continually increasing also. That means that R' becomes larger than 150 during the first day, so R will be larger than 2650 tomorrow.

Rates are continually changing; this affects the calculations

The fact that the rates are continually changing complicates the mathematical work we need to do to find S, I, and R. In chapter 2 we will develop tools and concepts that will overcome this problem. For the present we'll assume that the rates S', I', and R' stay fixed for the course of an entire day. This will still allow us to produce reasonable estimates for the values of S, I, and R. With these estimates we will get our first glimpse of the predictive power of the S-I-R model. We will also use the estimates as the starting point for the work in chapter 2 that will give us precise values. Let's look at the details of a specific problem.

The Problem. Consider a measles epidemic in a school population of 50,000 children. The recovery coefficient is $b = 1/14$. For the transmission coefficient we choose $a = .00001$, a number within the range used in epidemic studies. We suppose that 2100 people are currently infected and 2500 have already recovered. Since the total population is 50,000, there must be 45,400 susceptibles. Here is a summary of the problem in mathematical terms:

Rate equations:

$$S' = -.00001SI,$$
$$I' = .00001SI - I/14,$$
$$R' = I/14.$$

Initial values: when $t = 0$,

$$S = 45400, \qquad I = 2100, \qquad R = 2500.$$

Tomorrow. From our earlier discussion, $R' = 2100/14 = 150$ persons per day, giving us an estimated value of $R = 2650$ persons for tomorrow. To estimate S we use

$$S' = -.00001\,SI = -.00001 \times 45400 \times 2100 = -953.4 \text{ persons/day.}$$

Hence we estimate that tomorrow

$$S = 45400 - 953.4 = 44446.6 \text{ persons.}$$

Since $S + I + R = 50000$ and we have $S + R = 47096.6$ tomorrow, a final subtraction gives us $I = 2903.4$ persons. (Alternatively, we could have used the rate equation for I' to estimate I.)

The fractional values in the estimates for S and I remind us that the S-I-R model describes the behavior of the epidemic only approximately.

Several days hence. According to the model, we estimate that tomorrow $S = 44446.6$, $I = 2903.4$, and $R = 2650$. Therefore, from the new I we get a new approximation for the value of R' tomorrow; it is

$$R' = \frac{1}{14}I = \frac{1}{14} \times 2903.4 = 207.4 \text{ persons/day.}$$

Hence, two days from now we estimate that R will have the new value $2650 + 207.4 = 2857.4$. Now follow this pattern to get new approximations for S' and I', and then use those to estimate the values of S and I two days from now.

The pattern of steps that just carried you from the first day to the second will work just as well to carry you from the second to the third. Pause now and do all these calculations yourself. See exercises 15 and 16 on page 22. If you round your calculated values of S, I, and R to the nearest tenth, they should agree with those in the following table.

Stop and do the calculations

Estimates for the first three days

t	S	I	R	S'	I'	R'
0	45 400.0	2100.0	2500.0	−953.4	803.4	150.0
1	44 446.6	2903.4	2650.0	−1290.5	1083.1	207.4
2	43 156.1	3986.5	2857.4	−1720.4	1435.7	284.7
3	41 435.7	5422.1	3142.1			

Yesterday. We already pointed out, on page 5, that we can use our models to go *backwards* in time, too. This is a valuable way to see how well the model fits reality, because we can compare estimates that the model generates with health records for the days in the recent past.

To find how S, I, and R change when we go one day into the future we multiplied the rates S', I', and R' by a time step of +1. To find how they change when we go one day into the past we do the same thing, except that we must now use a time step of –1. According to the table above, the rates at time $t = 0$ (i.e., today) are

$$S' = -953.4, \qquad I' = 803.4, \qquad R' = 150.0.$$

Therefore we estimate that, one day ago,

$$\begin{aligned} S = 45400 + (-953.4 \times -1) &= 45400 + 953.4 = 46353.4, \\ I = 2100 + (803.4 \times -1) &= 2100 - 803.4 = 1296.6, \\ R = 2500 + (150.0 \times -1) &= 2500 - 150.0 = 2350.0. \end{aligned}$$

Just as we would expect with a spreading infection, there are more susceptibles yesterday than today, but fewer infected or recovered. In the exercises for section 3 you will have an opportunity to continue this process, tracing the epidemic many days into the past. For example, you will be able to go back and see when the infection started—that is, find the day when the value of I was only about 1.

Go forward a day and then back again

There and back again. What happens when we start with tomorrow's values and use tomorrow's rates to go back one day—back to today? We should get $S = 45400$, $I = 2100$, and $R = 2500$ once again, shouldn't we? Tomorrow's values are

$$S = 44446.6, \qquad I = 2903.4, \qquad R = 2650.0,$$
$$S' = -1290.5, \qquad I' = 1083.1, \qquad R' = 207.4.$$

To go backwards one day we must use a time step of -1. The predicted values are thus

$$\begin{aligned} S &= 44446.6 + (-1290.6 \times -1) &&= 45737.2, \\ I &= 2903.4 + (1083.1 \times -1) &&= 1820.3, \\ R &= 2650.0 + (207.4 \times -1) &&= 2442.6. \end{aligned}$$

These are *not* the values that we had at the start, when $t = 0$. In fact, it's worth noting the difference between the original values and those produced by "going there and back again."

	original value	there and back again	difference
S	45400	45737.1	337.1
I	2100	1820.3	–279.7
R	2500	2442.6	–57.4

Do you see why there are differences? We went forward in time using the rates that were current at $t = 0$, but when we returned we used the rates that were current at $t = 1$. Because these rates were different, we didn't get back where we started. These differences do not point to a flaw in the model; the problem lies with the way we are trying to extract information from the model. As we have been making estimates, we have assumed that the rates don't change over the course of a whole day. We already know that's not true, so the values that we have been getting are not exact. What this test adds to our knowledge is a way to measure just *how* inexact those values are—as we do in the table above.

The differences measure how rough an estimate is

In chapter 2 we will solve the problem of rough estimates by recalculating all the quantities ten times a day, a hundred times a day, or even more. When we do the computations with shorter and shorter time steps we will be able to see how the estimates improve. We will even be able to see how to get values that are mathematically exact!

Delta notation. This work has given us some insights about the way our model predicts future values of S, I, and R. The basic idea is very simple: *determine how S, I, and R change*. Because these changes play such an important role in what we do, it is worth having a simple way to refer to them. Here is the notation that we will use:

Δx stands for a **change** in the quantity x

The symbol "Δ" is the Greek capital letter *delta*; it corresponds to the Roman letter "D" and stands for **difference**.

Delta notation gives us a way to refer to changes of all sorts. For example, in the table on page 13, between day 1 and day 3 the quantities t and S change by

$$\Delta t = 2 \text{ days},$$
$$\Delta S = -3010.9 \text{ persons}.$$

Δ stands for a change, a difference, or a step

We sometimes refer to a change as a **step**. For instance, in this example we can say there is a "t step" of 2 days, and an "S step" of -3010.9 persons. In the calculations that produced the table on page 13 we "stepped into" the future, a day at a time. Finally, delta notation gives us a concise and vivid way to describe the relation between rates and changes. For example, if S changes at the constant rate S', then under a t step of Δt, the value of S changes by

$$\Delta S = S' \cdot \Delta t.$$

Using the computer as a tool. Suppose we wanted to find out what happens to S, I, and R after a month, or even a year. We need only repeat—30 times, or 365 times—the three rounds of calculations we used to go three days into the future. The computations would take time, though. The same is true if we wanted to do ten or one hundred rounds of calculations per day—which is the approach we'll take in chapter 2 to get more accurate values. To save our time and effort we will soon begin to use a computer to do the repetitive calculations.

A computer does calculations by following a set of instructions called a **program**. Of course, if we had to give a million instructions to make the computer carry out a million steps, there would be no savings in labor. The trick is to write a program with just a few instructions that can be repeated over and over again to do all the calculations we want. The usual way to do this is to arrange the instructions in a **loop**. To give you an idea what a loop is, we'll look at the S-I-R calculations. They form a loop. We can see the loop by making a flow chart.

The flow chart. We'll start by writing down the three steps that take us from one day to the next:

Step I Given the current values of S, I, and R, we get current S', I', and R' by using the rate equations

$$\begin{aligned} S' &= -aSI, \\ I' &= aSI - bI, \\ R' &= \phantom{-aSI - {}} bI. \end{aligned}$$

Step II Given the current values of S', I', and R', we find the changes ΔS, ΔI, and ΔR over the course of a day by using the equations

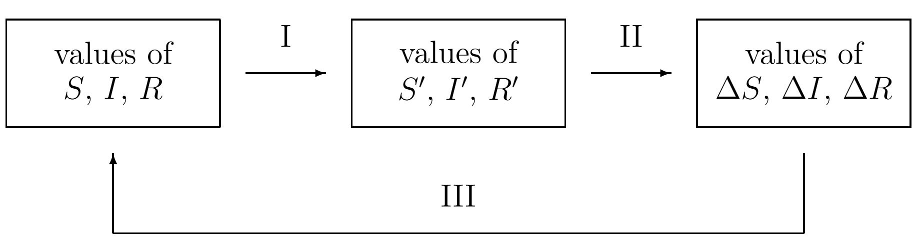
$$\begin{aligned} \Delta S \text{ persons} &= S' \ \frac{\text{persons}}{\text{day}} \times 1 \text{ day}, \\ \Delta I \text{ persons} &= I' \ \frac{\text{persons}}{\text{day}} \times 1 \text{ day}, \\ \Delta R \text{ persons} &= R' \ \frac{\text{persons}}{\text{day}} \times 1 \text{ day}. \end{aligned}$$

Step III Given the current values of ΔS, ΔI, and ΔR, we find the new values of S, I, and R a day later by using the equations

$$\begin{aligned} \text{new } S &= \text{current } S + \Delta S, \\ \text{new } I &= \text{current } I + \Delta I, \\ \text{new } R &= \text{current } R + \Delta R. \end{aligned}$$

Each step takes three numbers as **input**, and produces three new numbers as **output**. Note that the output of each step is the input of the next, so the steps flow together. The diagram below, called a **flow chart**, shows us how the steps are connected.

The flow chart forms a loop



The calculations form a **loop**, because the output of step III is the input for step I. If we go once around the loop, the output of step III gives us the values of S, I, and R *on the following day.* The steps do indeed carry us into the future.

Each step involves calculating three numbers. If we count each calculation as a single instruction, then it takes nine instructions to carry the values of S, I, and R one day into the future. To go a million days into the future, we need add only one more instruction: "Go around the loop a million times." In this way, a computer program with only ten instructions can carry out a million rounds of calculations!

Later in this chapter (section 3) you will find a real computer program that lists these instructions (for three days instead of a million, though). Study the program to see which instructions accomplish which steps. In particular, see how it makes a loop. Then run the program to check that the computer reproduces the values you already computed by hand. Once you see how the program works, you can modify it to get further information—for example, you can find out what happens to S, I, and R thirty days into the future. You will even be able to plot the graphs of S, I, and R.

A computer program will carry out the three steps

Rate equations have always been at the heart of calculus, and they have been analyzed using mechanical and electronic computers for as long as those tools have been available. Now that small powerful computers have begun to appear in the classroom, it is possible for beginning calculus students to explore interesting and complex problems that are modelled by rate equations. Computers are changing how mathematics is done and how it is learned.

Analysis without a computer. A computer is a powerful tool for exploring the S-I-R model, but there are many things we can learn about the model without using a computer. Here is an example.

According to the model, the rate at which the infected population grows is given by the equation

$$I' = .00001\,SI - I/14 \quad \text{persons/day}.$$

In our example, $I' = 803.4$ at the outset. This is a positive number, so I increases initially. In fact, I will continue to increase as long as I' is positive. If I' ever becomes negative, then I decreases. So let's ask the question: when is I' positive, when is it negative, and when is it zero? By factoring out I in the last equation we obtain

$$I' = I\left(.00001\,S - \frac{1}{14}\right) \text{ persons/day}.$$

Consequently $I' = 0$ if either

$$I = 0 \qquad \text{or} \qquad .00001\,S - \frac{1}{14} = 0.$$

The first possibility $I = 0$ has a simple interpretation: there is no infection within the population. The second possibility is more interesting; it says that I' will be zero when

$$.00001\,S - \frac{1}{14} = 0 \qquad \text{or} \qquad S = \frac{100000}{14} \approx 7142.9.$$

If S is *greater* than 100000/14 and I is positive, then you can check that the formula

$$I' = I\left(.00001\,S - \frac{1}{14}\right) \text{ persons/day}$$

tells us I' is positive—so I is increasing. If, on the other hand, S is *less than* 100000/14, then I' is negative and I is decreasing. So $S = 100000/14$ represents a **threshold**. If S falls below the threshold, I decreases. If S exceeds the threshold, I increases. Finally, I reaches its peak when S equals the threshold.

The threshold determines whether there will be an epidemic

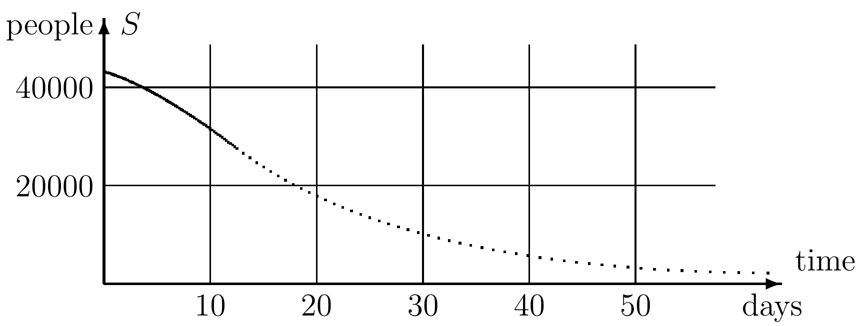
The presence of a threshold value for S is purely a mathematical result. However, it has an interesting interpretation for the epidemic. As long as there are at least 7143 susceptibles, the infection will spread, in the sense that there will be more people falling ill than recovering each day. As new people fall ill, the number of susceptibles declines. Finally, when there are fewer than 7143 susceptibles, the pattern reverses: each day more people will recover than will fall ill.

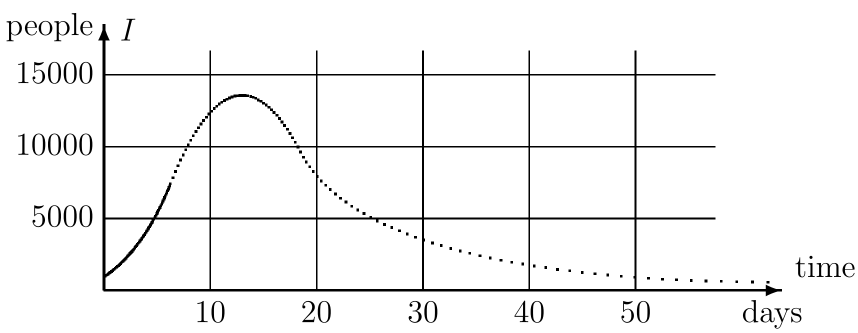
If there were fewer than 7143 susceptibles in the population *at the outset*, then the number of infected would only decline with each passing day. The infection would simply never catch hold. The clear implication is that the noticeable surge in the number of cases that we associate with an "epidemic" disease is due to the presence of a large susceptible population. If the susceptible population lies below a certain threshold value, a surge just isn't possible. This is a valuable insight—and we got it with little effort. We didn't need to make lengthy calculations or call on the resources of a computer; a bit of algebra was enough.

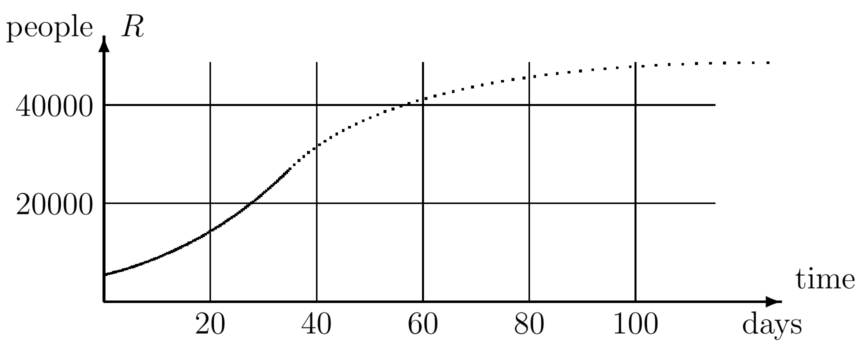
Exercises

Reading a Graph

The graphs on pages 3 and 4 have no scales marked along their axes, so they provide mainly *qualitative* information. The graphs below do have scales, so you can now answer *quantitative* questions about them. For example, on day 20 there are about 18,000 susceptible people. Read the graphs to answer the following questions. (Note: $S + I + R$ is *not* constant in this example, so these graphs cannot be solutions to our model..)







1. When does the infection hit its peak? How many people are infected at that time?

2. Initially, how many people are susceptible? How many days does it take for the susceptible population to be cut in half?

3. How many days does it take for the recovered population to reach 25,000? How many people *eventually* recover? Where did you look to get this information?

4. On what day is the size of the infected population increasing most rapidly? When is it decreasing most rapidly? How do you know?

5. How many people caught the illness at some time during the first 20 days? (Note that this is not the same as the number of people who are infected on day 20.) Explain where you found this information.

6. Copy the graph of R as accurately as you can, and then superimpose a sketch of S on it. Notice the time scales on the *original* graphs of S and R are different. Describe what happened to the graph of S when you superimposed it on the graph of R. Did it get compressed or stretched? Was this change in the horizontal direction or the vertical?

A Simple Model

These questions concern the rate equation $S' = -470$ persons per day that we used to model a susceptible population on pages 4–6.

7. Suppose the initial susceptible population was 20,000 on Wednesday. Use the model to answer the following questions.

a) How many susceptibles will be left ten days later?

b) How many days will it take for the susceptible population to vanish entirely?

c) How many susceptibles were there on the previous Sunday?

d) How many days before Wednesday were there 30,000 susceptibles?

Mark Twain's Mississippi

The Lower Mississippi River meanders over its flat valley, forming broad loops called ox-bows. In a flood, the river can jump its banks and cut off one of these loops, getting shorter in the process. In his book *Life on the Mississippi* (1884), Mark Twain suggests, with tongue in cheek, that some day the river might even vanish! Here is a passage that shows us some of the pitfalls in using rates to predict the future and the past.

> In the space of one hundred and seventy six years the Lower Mississippi has shortened itself two hundred and forty-two miles. That is an average of a trifle over a mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oölitic Silurian Period, just a million years ago next November, the Lower Mississippi was upwards of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-pole. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long, and Cairo [Illinois] and New Orleans will have joined their streets together and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesome returns of conjecture out of such a trifling investment of fact.

Let L be the length of the Lower Mississippi River. Then L is a variable quantity we shall analyze.

8. According to Twain's data, what is the exact **rate** at which L is changing, in miles per year? What approximation does he use for this rate? Is this a reasonable approximation? Is this rate *positive* or *negative*? Explain. In what follows, use Twain's approximation.

9. Twain wrote his book in 1884. Suppose the Mississippi that Twain wrote about had been 1100 miles long; how long would it have become in 1990?

10. Twain does not tell us how long the Lower Mississippi was in 1884 when he wrote the book, but he does say that 742 years later it will be only $1\frac{3}{4}$ miles long. How long must the river have been when he wrote the book?

11. Suppose t is the number of years since 1884. Write a formula that describes how much L has changed in t years. Your formula should complete the equation

$$\text{the change in } L \text{ in } t \text{ years} = \dots\ .$$

12. From your answer to question 10, you know how long the river was in 1884. From question 11, you know how much the length has changed t years after 1884. Now write a formula that describes how long the river is t years later.

13. Use your formula to find what L was a million years ago. Does your answer confirm Twain's assertion that the river was "upwards of 1,300,000 miles long" then?

14. Was the river ever 1,300,000 miles long; will it ever be $1\frac{3}{4}$ miles long? (This is called a **reality check**.) What, if anything, is wrong with the "trifling investment of fact" which led to such "wholesale returns of conjecture" that Twain has given us?

The Measles Epidemic

We consider once again the specific rate equations

$$\begin{aligned} S' &= -.00001\,SI, \\ I' &= .00001\,SI - I/14, \\ R' &= I/14, \end{aligned}$$

discussed in the text on pages 11–14. We saw that at time $t = 1$,

$$S = 44446.6, \qquad I = 2903.4, \qquad R = 2650.0.$$

15. Calculate the current rates of change S', I', and R' when $t = 1$, and then use these values to determine S, I, and R one day later.

16. In the previous question you found S, I, and R when $t = 2$. Using these values, calculate the rates S', I', and R' and then determine the new values of S, I, and R when $t = 3$. See the table on page 13.

17. **Double the time step**. Go back to the starting time $t = 0$ and to the initial values

$$S = 45400, \qquad I = 2100, \qquad R = 2500.$$

Recalculate the values of S, I, and R at time $t = 2$ by using a time step of $\Delta t = 2$. You should perform only a single round of calculations, and use the rates S', I', and R' that are current at time $t = 0$.

18. **There and back again**. In the text we went one day into the future and then back again to the present. Here you'll go forward two days from $t = 0$ and then back again. There are two ways to do this: with a time step of $\Delta t = \pm 2$ (as in the previous question), and with a pair of time steps of $\Delta t = \pm 1$.

a) ($\Delta t = \pm 2$). Using the values of S, I, and R at time $t = 2$ that you just got in the previous question, calculate the rates S', I', and R'. Then using a time step of $\Delta t = -2$, estimate new values of S, I, and R at time $t = 0$. How much do these new values differ from the original values 45,400, 2100, 2500?

b) ($\Delta t = \pm 1$). Now make a new start, using the values

$$S = 43156.1, \qquad I = 3986.5, \qquad R = 2857.4,$$
$$S' = -1720.4, \qquad I' = 1435.7, \qquad R' = 284.7.$$

that occur when $t = 2$ if we make estimates with a time step $\Delta t = 1$. (These values come from the table on page 13) Using two rounds of calculations with a time step of $\Delta t = -1$, estimate another set of new values for S, I, and R at time $t = 0$. How much do these new values differ from the original values 45,400, 2100, 2500?

c) Which process leads to a *smaller* set of differences: a single round of calculations with $\Delta t = \pm 2$, or two rounds of calculations with $\Delta t = \pm 1$? Consequently, which process produces better estimates—in the sense in which we used to measure estimates on page 14?

19. **Quarantine**. One of the ways to treat an epidemic is to keep the infected away from the susceptible; this is called quarantine. The intention is to reduce the chance that the illness will be transmitted to a susceptible person. Thus, quarantine alters the *transmission coefficient*.

a) Suppose a quarantine is put into effect that cuts in half the chance that a susceptible will fall ill. What is the new transmission coefficient?

b) On page 18 it was determined that whenever there were fewer than 7143 susceptibles, the number of infected would decline instead of grow. We called 7143 a *threshold* level for S. Changing the transmission coefficient, as in part (a), changes the threshold level for S. What is the new threshold?

c) Suppose we start with $S = 45{,}400$. Does quarantine eliminate the epidemic, in the sense that the number of infected immediately goes down from 2100, without ever showing an increase in the number of cases?

d) Since the new transmission coefficient is not small enough to guarantee that I never goes up, can you find a smaller value that *does* guarantee I never goes up? Continue to assume we start with $S = 45400$.

e) Suppose the initial susceptible population is 45,400. What is the *largest* value that the transmission coefficient can have and still guarantee that I never goes up? What level of quarantine does this represent? That is, do you have to reduce the chance that a susceptible will fall ill to one-third of what it was with no quarantine at all, to one-fourth, or what?

Other Diseases

20. Suppose the spread of an illness similar to measles is modelled by the following rate equations:

$$\begin{aligned} S' &= -.00002\,SI, \\ I' &= .00002\,SI - .08\,I, \\ R' &= .08\,I. \end{aligned}$$

Note: the initial values $S = 45400$, etc. that we used in the text do not apply here.

a) Roughly how long does someone who catches this illness remain infected? Explain your reasoning.

b) How large does the susceptible population have to be in order for the illness to take hold—that is, for the number of cases to increase? Explain your reasoning.

c) Suppose 100 people in the population are currently ill. According to the model, how many (of the 100 infected) will recover during the next 24 hours?

d) Suppose 30 *new* cases appear during the same 24 hours. What does that tell us about S'?

e) Using the information in parts (c) and (d), can you determine how large the current susceptible population is?

21. a) Construct the appropriate S-I-R model for a measles-like illness that lasts for 4 days. It is also known that a typical susceptible person meets only about 0.3% of infected population each day, and the infection is transmitted in only one contact out of six.

b) How small does the susceptible population have to be for this illness to fade away without becoming an epidemic?

22. Consider the general S-I-R model for a measles-like illness:

$$\begin{aligned} S' &= -aSI, \\ I' &= \quad aSI - bI, \\ R' &= \qquad\qquad bI. \end{aligned}$$

a) The threshold level for S—below which the number of infected will only decline—can be expressed in terms of the transmission coefficient a and the recovery coefficient b. What is that expression?

b) Consider two illnesses with the same transmission coefficient a; assume they differ only in the length of time someone stays ill. Which one has the lower threshold level for S? Explain your reasoning.

What Goes Around Comes Around

Some relatively mild illnesses, like the common cold, return to infect you again and again. For a while, right after you recover from a cold, you are immune. But that doesn't last; after some weeks or months, depending on the illness, you become susceptible again. This means there is now a flow from the recovered population to the susceptible. These exercises ask you to modify the basic S-I-R model to describe an illness where immunity is temporary.

23. Draw a compartment diagram for such an illness. Besides having all the ingredients of the diagram on page 9, it should depict a flow from R to S. Call this **immunity loss**, and use c to denote the coefficient of immunity loss.

24. Suppose immunity is lost after about six weeks. Show that you can set $c = 1/42$ per day, and explain your reasoning carefully. A suggestion: adapt the discussion of recovery in the text.

25. Suppose this illness lasts 5 days and it has a transmission coefficient of .00004 in the population we are considering. Suppose furthermore that the total population is fixed in size (as was the case in the text). Write down rate equations for S, I, and R.

26. We saw in the text that the model for an illness that confers permanent immunity has a threshold value for S in the sense that when S is above the threshold, I increases, but when it is below, I decreases. Does *this* model have the same feature? If so, what is the threshold value?

27. For a mild illness that confers permanent immunity, the size of the recovered population can only grow. This question explores what happens when immunity is only temporary.

a) Will R increase or decrease if

$$S = 45400, \qquad I = 2100, \qquad R = 2500?$$

b) Suppose we shift 20000 susceptibles to the recovered population (so that $S = 25400$ and $R = 22500$), leaving I unchanged. Now, will R increase or will it decrease?

c) Using a total population of 50,000, give two other sets of values for S, I, and R that lead to a decreasing R.

d) In fact, the relative sizes of I and R determine whether R will increase or decrease. Show that

$$\begin{array}{ll} \text{if } I > \frac{5}{42}R, & \text{then } R \text{ will increase;} \\ \text{if } I < \frac{5}{42}R, & \text{then } R \text{ will decrease.} \end{array}$$

Explain your argument clearly. A suggestion: consider the rate equation for R'.

28. **The steady state**. Any illness that confers only temporary immunity can appear to linger in a population forever. You may not always have a cold, but someone does, and eventually you catch another one. ("What goes around comes around.") Individuals gradually move from one compartment to the next. When they return to where they started, they begin another cycle.

Each compartment (in the diagram you drew in exercise 23) has an *inflow* and an *outflow*. It is conceivable that the two exactly balance, so that the *size* of the compartment doesn't change (even though its individual occupants do). When this happens for all three compartments simultaneously, the illness is said to be in a **steady state**. In this question you explore the steady state of the model we are considering. Recall that the total population is 50,000.

a) What must be true if the inflow and outflow to the I compartment are to balance?

b) What must be true if the inflow and outflow to the R compartment are to balance?

c) If neither I nor R is changing, then the model must be at the steady state. Why?

d) What is the value of S at the steady state?

e) What is the value of R at the steady state? A suggestion: you know $R + I = 50000 -$ (the steady state value of S). You also have a connection between I and R at the steady state.

1.2 The Mathematical Ideas

A number of important mathematical ideas have already emerged in our study of an epidemic. In this section we pause to consider them, because they have a "universal" character. Our aim is to get a fuller understanding of what we have done so we can use the ideas in other contexts.

> We often draw out of a few particular experiences a lesson that can be put to good use in new settings. This process is the essence of mathematics, and it has been given a name—abstraction—which means literally "drawing from." Of course abstraction is not unique to mathematics; it is a basic part of the human psyche.

Functions

Functions and their notation

A **function** describes how one quantity depends on another. In our study of a measles epidemic, the relation between the number of susceptibles S and the time t is a function. We write $S(t)$ to denote that S is a function of t. We can also write $I(t)$ and $R(t)$ because I and R are functions of t, too. We can even write $S'(t)$ to indicate that the rate S' at which S changes over time is a function of t. In speaking, we express $S(t)$ as "S of t" and $S'(t)$ as "S prime of t."

You can find functions everywhere. The amount of postage you pay for a letter is a function of the weight of the letter. The time of sunrise is a function of what day of the year it is. The crop yield from an acre of land is a function of the amount of fertilizer used. The position of a car's gasoline gauge (measured in centimeters from the left edge of the gauge) is a function of the amount of gasoline in the fuel tank. On a polygraph ("lie detector") there is a pen that records breathing; its position is a function of the amount of expansion of the lungs. The volume of a cubical box is a function of the length of a side. The last is a rather special kind of function because it can be described by an algebraic formula: if V is the volume of the box and s is the length of a side, then $V(s) = s^3$.

Most functions are *not* described by algebraic formulas, however. For instance, the postage function is given by a set of verbal instructions and the time of sunrise is given by a table in an almanac. The relation between a gas gauge and the amount of fuel in the tank is determined simply by making measurements. There is no algebraic formula that tells us how the number of susceptibles, S, depends upon t, either. Instead, we find $S(t)$ by carrying out the steps in the flow chart on page 16 until we reach t days into the future.

A function has input and output

In the function $S(t)$ the variable t is called the **input** and the variable S is called the **output**. In the sunrise function, the day of the year is the input and the time of sunrise is the output. In the function $S(t)$ we think of S as *depending on* t, so t is also called the independent variable and S the dependent variable. The set of values that the input takes is called the **domain** of the function. The set of values that the output takes is called the **range**.

The idea of a function is one of the central notions of mathematics. It is worth highlighting:

A function is a rule that specifies how the value of one variable, the input, determines the value of a second variable, the output.

Notice that we say *rule* here, and not *formula*. This is deliberate. We want the study of functions to be as broad as possible, to include all the ways one quantity is likely to be related to another in a scientific question.

Some technical details. It is important not to confuse an expression like $S(t)$ with a product; $S(t)$ does *not* mean $S \times t$. On the contrary, the expression $S(1.4)$, for example, stands for the output of the function S when 1.4 is the input. In the epidemic model we then interpret it as the number of susceptibles that remain 1.4 days after today.

We have followed the standard practice in science by letting the single letter S designate both the *function*—that is, the *rule*—and the *output* of that function. Sometimes, though, we will want to make the distinction. In that case we will use two different symbols. For instance, we might write $S = f(t)$. Then we are still using S to denote the output, but the new symbol f stands for the function rule.

The symbols we use to denote the input and the output of a function are just names; if we change them, we don't change the function. For example, here are three ways to describe the same function g:

$$\begin{aligned} g &: \text{multiply the input by 5, then subtract 3} \\ g(x) &= 5x - 3 \\ g(u) &= 5u - 3\,. \end{aligned}$$

It is important to realize that the *formulas* we just wrote in the last two lines are merely shorthand for the instructions stated in the first line. If you keep this in mind, then absurd-looking combinations like $g(g(2))$ can be decoded easily by remembering g of *anything* is just 5 times that thing, minus 3. We could thus evaluate $g(g(2))$ from the inside out (which is usually easier) as

$$g(g(2)) = g(5 \cdot 2 - 3) = g(7) = 5 \cdot 7 - 3 = 32,$$

or we could evaluate it from the outside in as

$$g(g(2)) = 5g(2) - 3 = 5(5 \cdot 2 - 3) - 3 = 5 \cdot 7 - 3 = 32,$$

as before.

Suppose f is some other rule, say $f(t) = t^2 + 4t - 1$. Remember that this is just shorthand for "Take the input (whatever it is), square it, add four times the input, and subtract 1." We could then evaluate

$$f(g(3)) = f(5 \cdot 3 - 3) = f(12) = 12^2 + 4 \cdot 12 - 1 = 144 + 48 - 1 = 191,$$

while

$$g(f(3)) = g(20) = 97.$$

This process of **chaining** functions together by using the output of one function as the input for another turns out to be very important later in this course, and will be taken up again in chapter 3. For now, though, you should treat it simply as part of the formal language of mathematics, requiring a knowledge of the rules but no cleverness. It is analogous to learning how to conjugate verbs in French class—it's not very exciting for its own sake, but it allows us to read the interesting stuff later on!

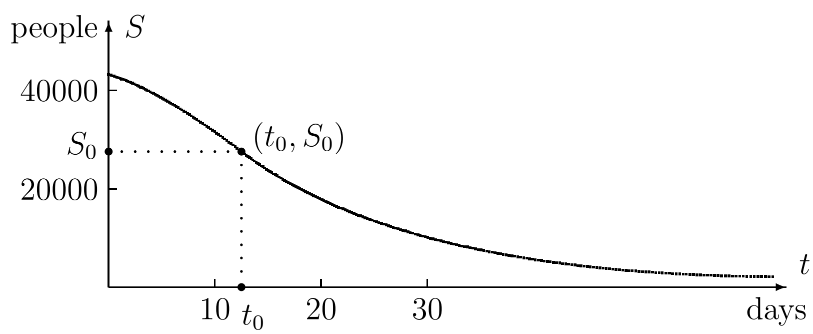
Part of math is simply learning the language

A particularly important class of functions is composed of the **constant functions** which give the same output for every input. If h is the constant function that always gives back 17, then in formula form we would express this as $h(x) = 17$. Constant functions are so simple you might feel you are missing the point, but that's all there is to it!

Graphs

A graph is a function rule given visually

A graph describes a function in a visual form. Sometimes—as with a seismograph or a lie detector, for instance—this is the *only* description we have of a particular function. The usual arrangement is to put the input variable on the horizontal axis and the output on the vertical—but it is a good idea when you are looking at a particular graph to take a moment to check; sometimes, the opposite convention is used! This is often the case in geology and economics, for instance.



Sketched above is the graph of a function $S(t)$ that tells how many susceptibles there are after t days. Given any t_0, we "read" the graph to find $S(t_0)$, as follows: from the point t_0 on the t-axis, go vertically until you reach the graph; then go horizontally until you reach the S-axis. The value S_0 at that point is the output $S(t_0)$. Here t_0 is about 13 and S_0 is about 27,000; thus, the graph says that $S(13) \approx 27000$, or about 27,000 susceptibles are left after 13 days.

Linear Functions

If y depends on x, then Δy depends on Δx

Changes in input and output. Suppose y is a function of x. Then there is some rule that answers the question: What is the value of y for any given x? Often, however, we start by knowing the value of y for a particular x, and the question we really want to ask is: How does y respond to *changes* in x? We are still dealing with the same function—just looking at it from a different point of view. This point of view is important; we use it to analyze functions (like $S(t)$, $I(t)$, and $R(t)$) that are defined by rate equations.

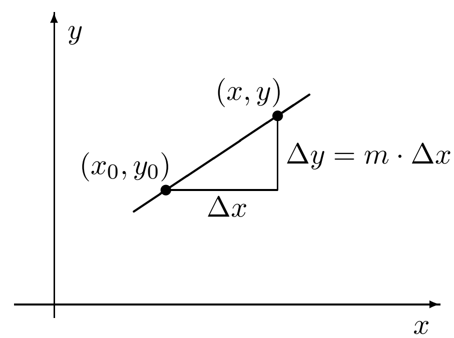
The way Δy depends on Δx can be simple or it can be complex, depending on the function involved. The simplest possibility is that Δy and Δx are **proportional**:

The defining property of a linear function

$$\Delta y = m \cdot \Delta x, \quad \text{for some constant } m.$$

Thus, if Δx is doubled, so is Δy; if Δx is tripled, so is Δy. A function whose input and output are related in this simple way is called a **linear function**, because the graph is a straight line. Let's take a moment to see why this is so.

The graph of a linear function. The graph consists of certain points (x, y) in the x, y-plane. Our job is to see how those points are arranged. Fix one of them, and call it (x_0, y_0). Let (x, y) be any other point on the graph. Draw the line that connects this point to (x_0, y_0), as we have done in the figure at the right. Now set



$$\Delta x = x - x_0, \qquad \Delta y = y - y_0.$$

By definition of a linear function, $\Delta y = m \cdot \Delta x$, as the figure shows, so the slope of this line is $\Delta y / \Delta x = m$. Recall that m is a constant; thus, if we pick a new point (x, y), the slope of the connecting line won't change.

Since (x, y) is an arbitrary point on the graph, what we have shown is that **every point on the graph lies on a line of slope m through the point (x_0, y_0)**. But there is only one such line—and all the points lie on it! That line must be the graph of the linear function.

A linear function is one that satisfies $\Delta y = m \cdot \Delta x$; its graph is a straight line whose slope is m.

Rates, slopes, and multipliers. The interpretation of m as a slope is just one possibility; there are two other interpretations that are equally important. To illustrate them we'll use Mark Twain's vivid description of the shortening of the Lower Mississippi River (see page 21). This will also give us the chance to see how a linear function emerges in context.

Twain says "the Lower Mississippi has shortened itself ... an average of a trifle over a mile and a third per year." Suppose we let L denote the length of the river, in miles, and t the time, in years. Then L depends on t, and Twain's statement implies that L is a *linear* function of t—in the sense in which we have just defined a linear function. Here is why. According to our definition, there must be some number m which makes $\Delta L = m \cdot \Delta t$. But notice that Twain's statement has exactly this form if we translate it into mathematical language. Convince yourself that it says

Stop and do the translation

$$\Delta L \text{ miles} = -1\tfrac{1}{3}\,\frac{\text{miles}}{\text{year}} \times \Delta t \text{ years}.$$

Thus we should take m to be $-1\frac{1}{3}$ miles per year.

The role of m here is to convert one quantity (Δt years) into another (ΔL miles) by multiplication. All linear functions work this way. In the defining equation $\Delta y = m \cdot \Delta x$, multiplication by m converts Δx into Δy. Any change in x produces a change in y that is m times as large. For this reason we give m its second interpretation as a **multiplier**.

It is easier to understand why the usual symbol for *slope* is m—instead of s—when you see that a slope can be interpreted as a multiplier.

It is important to note that, in our example, m is not simply $-1\frac{1}{3}$; it is $-1\frac{1}{3}$ *miles per year*. In other words, m is the **rate** at which the river is getting shorter. All linear functions work this way, too. We can rewrite the equation $\Delta y = m \cdot \Delta x$ as a ratio

$$m = \frac{\Delta y}{\Delta x} = \text{the rate of change of } y \text{ with respect to } x.$$

For these reasons we give m its third interpretation as a **rate of change**.

For a linear function satisfying $\boldsymbol{\Delta y = m \cdot \Delta x}$, the coefficient $\boldsymbol{m}$ is rate of change, slope, and multiplier.

We already use y' to denote the rate of change of y, so we can now write $m = y'$ when y is a linear function of x. In that case we can also write

$$\Delta y = y' \cdot \Delta x.$$

This expression should recall a pattern very familiar to you. (If not, change y to S and x to t!) It is the fundamental formula we have been using to calculate future values of S, I, and R. We can approach the relation between y and x the same way. That is, if y_0 is an "initial value" of y, when $x = x_0$, then *any* value of y can be calculated from

$$y = y_0 + y' \cdot \Delta x \quad \text{or} \quad y = y_0 + m \cdot \Delta x.$$

If x and y have units, so does m

Units. Suppose x and y are quantities that are measured in specific units. If y is a linear function of x, with $\Delta y = m \cdot \Delta x$, then m must have units too. Since m is the multiplier that "converts" x into y, the units for m must be chosen so they will convert x's units into y's units. In other words,

$$\text{units for } y = \text{units for } m \times \text{units for } x.$$

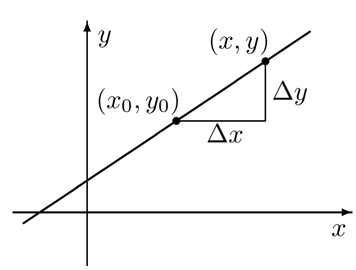
This implies

$$\text{units for } m = \frac{\text{units for } y}{\text{units for } x}.$$

For example, the multiplier in the Mississippi River problem converts years to miles, so it must have units of miles per year. The rate equation $R' = bI$ in the S-I-R model is a more subtle example. It says that R' is a linear function of I. Since R' is measured in persons per day and I is measured in persons, we must have

$$\text{units for } b = \frac{\text{units for } R'}{\text{units for } I} = \frac{\dfrac{\text{persons}}{\text{days}}}{\text{persons}}$$

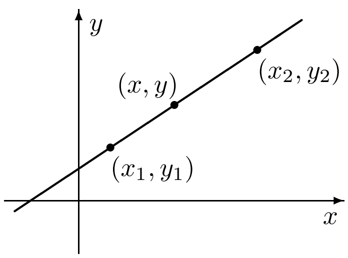
Formulas for linear functions. The expression $\Delta y = m \cdot \Delta x$ declares that y is a linear function of x, but it doesn't quite tell us what y itself *looks like* directly in terms of x. In fact, there are several equivalent ways to write the relation $y = f(x)$ in a formula, depending on what information we are given about the function.



• **The initial-value form**. Here is a very common situation: we know the value of y at an "initial" point—let's say $y_0 = f(x_0)$—and we know the rate of change—let's say it is m. Then the graph is the straight line of slope m that passes through the point (x_0, y_0). The formula for f is

$$y = y_0 + \Delta y = y_0 + m \cdot \Delta x = y_0 + m(x - x_0) = f(x).$$

What you should note particularly about this formula is that it expresses y in terms of the initial data x_0, y_0, and m—as well as x. Since that data consists of a point (x_0, y_0) and a slope m, the initial-value formula is also referred to as the **point-slope form** of the equation of a line. It may be more familiar to you with that name.



• **The interpolation form**. This time we are given the value of y at *two* points—let's say $y_1 = f(x_1)$ and $y_2 = f(x_2)$. The graph is the line that passes through (x_1, y_1) and (x_2, y_2), and its slope is therefore

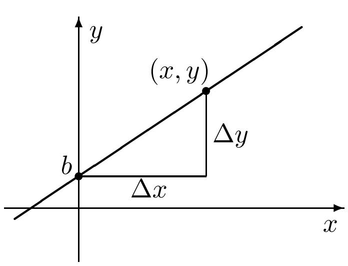
$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Now that we know the slope of the graph we can use the point-slope form (taking (x_1, y_1) as the "point", for example) to get the equation. We have

$$y = y_1 + m(x - x_1) = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) = f(x).$$

Notice how, once again, y is expressed in terms of the initial data—which consists of the two points (x_1, y_1) and (x_2, y_2).

The process of finding values of a quantity between two given values is called **interpolation**. Since our new expression does precisely that, it is called the interpolation formula. (Of course, it also finds values outside the given interval.) Since the initial data is a pair of points, the interpolation formula is also called the **two-point formula** for the equation of a line.



• **The slope-intercept form**. This is a special case of the initial-value form that occurs when the initial $x_0 = 0$. Then the point (x_0, y_0) lies on the y-axis, and it is frequently written in the alternate form $(0, b)$. The number b is called the ***y*-intercept**. The equation is

$$y = mx + b = f(x).$$

In the past you may have thought of this as *the* formula for a linear function, but for us it is only one of several. You will find that we will use the other forms more often.

Functions of Several Variables

Language and notation. Many functions depend on more than one variable. For example, sunrise depends on the day of the year but it also depends on the latitude (position north or south of the equator) of the observer. Likewise, the crop yield from an acre of land depends on the amount of fertilizer used, but it also depends on the amount of rainfall, on the composition of the soil, on the amount of weeding done—to mention just a few of the other variables that a farmer has to contend with.

A function can have several input variables

The rate equations in the S-I-R model also provide examples of functions with more than one input variable. The equation

$$I' = .00001\,SI - I/14$$

says that we need to specify both S and I to find I'. We can say that

$$F(S, I) = .00001\,SI - I/14$$

is a function whose input is the **ordered pair** of variables (S, I). In this case F is given by an algebraic formula. While many other functions of several variables also have formulas—and they are extremely useful—not all functions do. The sunrise function, for example, is given by a two-way table (see page 167) that shows the time of sunrise for different days of the year and different latitudes.

As a technical matter it is important to note that the input variables S and I of the function $F(S, I)$ above appear in a particular *order*, and that order is part of the definition of the function. For example, $F(1, 0) = 0$, but $F(0, 1) = -1/14$. (Do you see why? Work out the calculations yourself.)

Parameters. Suppose we rewrite the rate equation for I', replacing .00001 and 1/14 with the general values a and b:

$$I' = aSI - bI.$$

This makes it clear that I' depends on a and b, too. But note that a and b are not variables in quite the same way that S and I are. For example, a and b will vary if we switch from one disease to another or from one population to another. However, they will stay fixed while we consider a particular disease in a particular population. By contrast, S and I will *always* be treated as variables. We call a quantity like a or b a **parameter**.

To emphasize that I' depends on the parameters as well as S and I, we can write I' as the output of a new function

$$I' = I'(S, I, a, b) = aSI - bI$$

Some functions depend on parameters

whose input is the set of *four* variables (S, I, a, b), in that order. The variables S, I, and R must also depend on the parameters, too, and not just on t. Thus, we should write $S(t, a, b)$, for example, instead of simply $S(t)$. We implicitly used the fact that S, I, and R depend on a and b when we discovered there was a threshold for an epidemic (page 18). In exercise 22 of section 1 (page 25), you made the relation explicit. In that problem you show I will simply decrease over time (i.e., there will be no "burst" of infection) if

$$S < \frac{b}{a}.$$

There are even more parameters lurking in the S-I-R problem. To uncover them, recall that we needed *two* pieces of information to estimate S, I, and R over time:

1) the rate equations;
2) the initial values S_0, I_0, and R_0.

We used $S_0 = 45400$, $I_0 = 2100$, and $R_0 = 2500$ in the text, but if we had started with other values then S, I, and R would have ended up being different functions of t. Thus, we should really write

$$S = S(t, a, b, S_0, I_0, R_0)$$

to tell a more complete story about the inputs that determine the output S. Most of the time, though, we do *not* want to draw attention to the parameters; we usually write just $S(t)$.

Further possibilities. Steps I, II, and III on page 16 are also functions, because they have well-defined input and output. They are unlike the other examples we have discussed up to this point because they have more than one output variable. You should see, though, that there is nothing more difficult going on here.

In our study of the S-I-R model it was natural not to separate functions that have one input variable from those that have several. This is the pattern we shall follow in the rest of the course. In particular, we will want to deal with parameters, and we will want to understand how the quantities we are studying depend on those parameters.

The Beginnings of Calculus

While functions, graphs, and computers are part of the general fabric of mathematics, we can also abstract from the S-I-R model some important aspects of the calculus itself. The first of these is the idea of a **rate of change**. In this chapter we just assumed the idea was intuitively clear. However, there are some important questions not yet answered; for example, how do you deal with a quantity whose rate of change is itself always changing? These questions, which lead to the fundamental idea of a **derivative**, are taken up in chapter 3.

Rate equations—more commonly called **differential equations**—lie at the very heart of calculus. We will have much more to say about them, because many processes in the physical, biological, and social realms can be modelled by rate equations. In our analysis of the S-I-R model, we used rate equations to estimate future values by assuming that rates stay fixed for a whole day at a time. The discussion called "there and back again" on page 14 points up the shortcomings of this assumption. In chapter 2 we will develop a procedure, called Euler's method, to address this problem. In chapter 4 we will return to differential equations in a general way, equipped with Euler's method and the concept of the derivative.

How the next three chapters are connected

Exercises

Functions and Graphs

1. Sketch the graph of each of the following functions. Label each axis, and mark a scale of units on it. For each line that you draw, indicate

i) its slope;
ii) its y-intercept;
iii) its x-intercept (where it crosses the x-axis).

a) $y = -\frac{1}{2}x + 3$

b) $y = (2x - 7)/3$

c) $5x + 3y = 12$

2. Graph the following functions. Put labels and scales on the axes.

a) $V = .3Z - 1$;

b) $W = 600 - P^2$.

3. Sketch the graph of each of the following functions. Put labels and scales on the axes. For each graph that you draw, indicate

i) its y-intercept;
ii) its x-intercept(s).

For part (d) you will need the **quadratic formula**

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

for the roots of the **quadratic equation** $ax^2 + bx + c = 0$.

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

The next four questions refer to these functions:

$c(x, y) = 17$	a constant function
$j(z) = z$	the identity function
$r(u) = 1/u$	the reciprocal function
$D(p, q) = p - q$	the difference function
$s(y) = y^2$	the squaring function
$Q(v) = \dfrac{2v + 1}{3v - 6}$	a rational function

$$H(x) = \begin{cases} 5 & \text{if } x < 0 \\ x^2 + 2 & \text{if } 0 \leq x < 6 \\ 29 - x & \text{if } 6 \leq x \end{cases}$$

$$T(x, y) = r(x) + Q(y)$$

4. Determine the following values:

$c(5, -3)$	$j(17)$	$c(a, b)$	$j(u^2 + 1)$
$j(c(3, -5))$	$s(1.1)$	$r(1/17)$	$Q(0)$
$Q(2)$	$Q(3/7)$	$D(5, -3)$	$D(-3, 5)$
$H(1)$	$H(7)$	$H(4)$	$H(H(H(-3)))$
$r(s(-4))$	$r(Q(3))$	$Q(r(3))$	$T(3, 7)$

5. True or false. Give reasons for your answers: if you say true, explain why; if you say false, give an example that shows why it is false.

a) For every non-zero number x, $r(r(x)) = j(x)$.

b) If $a > 1$, then $s(a) > 1$.

c) If $a > b$, then $s(a) > s(b)$.

d) For all real numbers a and b, $s(a + b) = s(a) + s(b)$.

e) For all real numbers a, b, and c, $D(D(a, b), c)) = D(a, D(b, c))$.

6. Find all numbers x for which $Q(x) = r(Q(x))$.

7. The **natural domain** of a function f is the largest possible set of real numbers x for which $f(x)$ is defined. For example, the natural domain of $r(x) = 1/x$ is the set of all non-zero real numbers.

a) Find the natural domains of Q and H.

b) Find the natural domains of $P(z) = Q(r(z))$; $R(v) = r(Q(v))$.

c) What is the natural domain of the function $W(t) = \sqrt{\dfrac{1 - t^2}{t^2 - 4}}$?

Computer Graphing

The purpose of these exercises is to give you some experience using a "graphing package" on a computer. This is a program that will draw the graph of a function $y = f(x)$ whose formula you know. You must type in the formula, using the following symbols to represent the basic arithmetic operations:

to indicate	type
addition	`+`
subtraction	`-`
multiplication	`*`
division	`/`
an exponent	`^`

The caret " ^ " appears above the "6" on a keyboard (Shift-6).
Here is an example:

to enter:	type:
$\dfrac{7x^5 - 9x^2}{x^3 + 1}$	`(7*x^5 - 9*x^2)/(x^3 + 1)`

The parentheses you see here are important. If you do not include them, the computer will interpret your entry as

$$7x^5 - \frac{9x^2}{x^3} + 1 = 7x^5 - \frac{9}{x} + 1 \neq \frac{7x^5 - 9x^2}{x^3 + 1}.$$

In some graphing packages, you do not need to use `*` to indicate a multiplication. If this is true for the package you use, then you can enter the fractional expression above in the somewhat simpler form

```
(7x^5 - 9x^2)/(x^3 + 1).
```

To do the following exercises, follow the specific instructions for the graphing package you are using.

8. Graph the function $f(x) = .6x + 2$ on the interval $-4 \leq x \leq 4$.

a) What is the y-intercept of this graph? What is the x-intercept?

b) Read from the graph the value of $f(x)$ when $x = -1$ and when $x = 2$. What is the difference between these y values? What is the difference between the x values? According to these differences, what is the slope of the graph? According to the *formula*, what is the slope?

9. Graph the function $f(x) = 1 - 2x^2$ on the interval $-1 \leq x \leq 1$.

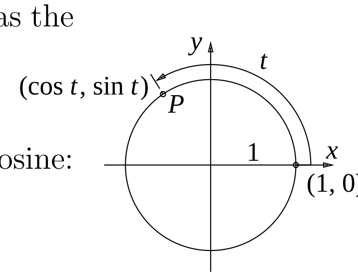
a) What is the y-intercept of this graph? The graph has two x-intercepts; use algebra to find them.

You can also find an x-intercept using the computer. The idea is to **magnify** the graph near the intercept until you can determine as many decimal places in the x coordinate as you want. For a start, graph the function on the interval $0 \leq x \leq 1$. You should be able to see that the graph on your computer monitor crosses the x-axis somewhere around .7. Regraph $f(x)$ on the interval $.6 \leq x \leq .8$. You should then be able to determine that the x-intercept lies between .70 and .71. This means $x = .7\ldots$; that is, you know the location of the x-intercept to one decimal place of accuracy.

b) Regraph $f(x)$ on the interval $.70 \leq x \leq .71$ to get two decimal places of accuracy in the location of the x-intercept. Continue this process until you have at least 7 places of accuracy. What is the x-intercept?

The circular functions Graphing packages "know" the familiar functions of trigonometry. Trigonometric functions are qualitatively different from the functions in the preceding problems. Those functions are defined by algebraic formulas, so they are called **algebraic functions**. The trigonometric functions are defined by explicit "recipes," but *not* by algebraic formulas; they are called **transcendental functions**. For calculus, we usually use the definition of the trigonometric functions as **circular functions**. This definition begins with a unit circle centered at the origin. Given the input number t, locate a point P on the circle by tracing an arc of length t along the circle from the point $(1, 0)$. If t is positive, trace the arc counterclockwise; if t is negative, trace it clockwise. Because the circle has radius 1, the arc of length t subtends a central angle of **radian** measure t.

The circular (or trigonometric) functions $\cos t$ and $\sin t$ are defined as the coordinates of the point P,



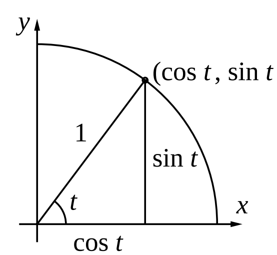
$$P = (\cos t, \sin t).$$

The other trigonometric functions are defined in terms of the sine and cosine:

$$\tan t = \sin t / \cos t, \qquad \sec t = 1/\cos t,$$
$$\cot t = \cos t / \sin t, \qquad \csc t = 1/\sin t.$$

Notice that when t is a positive acute angle, the circle definition agrees with the right triangle definitions of the sine and cosine:

$$\sin t = \frac{\text{opposite}}{\text{hypotenuse}} \qquad \text{and} \qquad \cos t = \frac{\text{adjacent}}{\text{hypotenuse}}.$$



However, the circle definitions of the sine and cosine have the important advantage that they produce functions whose domains are the set of *all* real numbers. (What are the domains of the tangent, secant, cotangent and cosecant functions?)

In calculus, angles are also always measured in radians. To convert between radians and degrees, notice that the circumference of a unit circle is 2π, so the radian measure of a semi-circular arc is half of this, and thus we have

$$\pi \text{ radians } = 180 \text{ degrees}.$$

As the course progresses, you will see why radians are used rather than degrees (or mils or any other unit for measuring angles)—it turns out that the formulas important to calculus take their simplest form when angles are expressed in radians.

Simplicity determines the choice of radians for measuring angles

How do we get values for the circular functions?

Graphing packages "know" the trigonometric functions in exactly this form: circular functions with the input variable given in radians. You might wonder, though, how a computer or calculator "knows" that $\sin(1) = .017452406\ldots$. It certainly isn't drawing a very accurate circle somewhere and measuring the y coordinate of some point. While the circular function approach is a useful way to think about the trigonometric functions conceptually, it isn't very helpful if we actually want values of the functions. One of the achievements of calculus, as you will see later in this course, is that it provides effective methods for computing values of functions like the circular functions that aren't given by algebraic formulas.

The following exercises let you review the trigonometric functions and explore some of the possibilities using computer graphing.

10. Graph the function $f(x) = \sin(x)$ on the interval $-2 \le x \le 10$.

a) What are the x-intercepts of $\sin(x)$ on the interval $-2 \le x \le 10$? Determine them to two decimal places accuracy.

b) What is the largest value of $f(x)$ on the interval $-2 \le x \le 10$? Which value of x makes $f(x)$ largest? Determine x to two decimal places accuracy.

c) Regraph $f(x)$ on the very small interval $-.001 \le x \le .001$. Describe what you see. Can you determine the slope of this graph?

11. Graph the function $f(x) = \cos(x)$ on the interval $0 \le x \le 14$. On the same screen graph the *second* function $g(x) = \cos(2x)$.

a) How far apart are the x-intercepts of $f(x)$? How far apart are the x-intercepts of $g(x)$?

b) The graph of $g(x)$ has a pattern that repeats. How wide is this pattern? The graph of $f(x)$ also has a repeating pattern; how wide is *it*?

c) Compare the graphs of $f(x)$ and $g(x)$ to one another. In particular, can you say that one of them is a stretched or compressed version of the other? Is the compression (or stretching) in the vertical or the horizontal direction?

d) Construct a *new* function $f(x)$ whose graph is the same shape as the graph of $g(x) = \cos(2x)$, but make the graph of $f(x)$ twice as tall as the graph of $g(x)$. [A suggestion: either deduce what $f(x)$ should be, or make a guess. Then test your choice on the computer. If your choice doesn't work, think how you might modify it, and then test your modifications the same way.]

12. The aim here is to find a solution to the equation $\sin x = \cos(3x)$. There is no purely *algebraic* procedure to solve this equation. Because the sine and cosine are not defined by *algebraic* formulas, this should not be particularly surprising. (Even for algebraic equations, there are only a few very special cases for which there are formulas like the quadratic formula. In chapter 5 we will look at a method for solving equations when formulas can't help us.)

a) Graph the two functions $f(x) = \sin(x)$ and $g(x) = \cos(3x)$ on the interval $0 \le x \le 1$.

b) Find a solution of the equation $\sin(x) = \cos(3x)$ that is accurate to six decimal places.

c) Find *another* solution of the equation $\sin(x) = \cos(3x)$, accurate to four decimal places. Explain how you found it.

13. Use a graphing program to make a sketch of the graph of each of the following functions. In each case, make clear the domain and the range of the function, where the graph crosses the axes, and where the function has a maximum or a minimum.

a) $F(w) = (w-1)(w-2)(w-3)$

b) $Q(a) = \dfrac{1}{a^2+5}$

c) $E(x) = x + \dfrac{1}{x}$

d) $e(x) = x - \dfrac{1}{x}$

e) $g(u) = \sqrt{\dfrac{u-1}{u+1}}$

f) $M(u) = \dfrac{u^2-2}{u^2+2}$

14. Graph on the same screen the following three functions:

$$f(x) = 2^x, \qquad g(x) = 3^x, \qquad h(x) = 10^x.$$

Use the interval $-2 \le x \le 1.2$.

a) Which function has the largest value when $x = -2$?

b) Which is climbing most rapidly when $x = 0$?

c) Magnify the picture at $x = 0$ by resetting the size of the interval to $-.0001 \le x \le .0001$. Describe what you see. Estimate the slopes of the three graphs.

Proportions, Linear Functions, and Models

15. Go back to the three functions given in problem 1. For each function, choose an initial value x_0 for x, find the corresponding value y_0 for y, and express the function in the form $y - y_0 = m \cdot (x - x_0)$.

16. You should be able to answer all parts of this problem without ever finding the equations of the functions involved.

a) Suppose $y = f(x)$ is a linear function with multiplier $m = 3$. If $f(2) = -5$, what is $f(2.1)$? $f(2.0013)$? $f(1.87)$? $f(922)$?

b) Suppose $y = G(x)$ is a linear function with multiplier $m = -2$. If $G(-1) = 6$, for what value of x is $G(x) = 8$? $G(x) = 0$? $G(x) = 5$? $G(x) = 491$?

c) Suppose $y = h(x)$ is a linear function with $h(2) = 7$ and $h(6) = 9$. What is $h(2.046)$? $h(2 + a)$?

17. In Massachusetts there is a sales tax of 5%. The tax T, in dollars, is proportional to the price P of an object, also in dollars. The constant of proportionality is $k = 5\% = .05$. Write a formula that expresses the sales tax as a linear function of the price, and use your formula to compute the tax on a television set that costs \$289.00 and a toaster that costs \$37.50.

18. Suppose $W = 213 - 17\,Z$. How does W change when Z changes from 3 to 7; from 3 to 3.4; from 3 to 3.02? Let ΔZ denote a change in Z and ΔW the change thereby produced in W. Is $\Delta W = m\,\Delta Z$ for some constant m? If so, what is m?

19. a) In the following table, q is a linear function of p. Fill in the blanks in the table.

p	–3	0		7	13		π
q	7		4	1		0	

b) Find a formula to express Δq as a function of Δp, and another to express q as a function of p.

20. **Thermometers**. There are two scales in common use to measure the temperature, called **Fahrenheit degrees** and the **Celsius degrees**. Let F and C, respectively, be the temperature on each of these scales. Each of these quantities is a linear function of the other; the relation between them in determined by the following table:

physical measurement	C	F
freezing point of water	0	32
boiling point of water	100	212

a) Which represents a larger change in temperature, a Celsius degree or a Fahrenheit degree?

b) How many Fahrenheit degrees does it take to make the temperature go up one Celsius degree? How many Celsius degrees does it take to make it go up one Fahrenheit degree?

c) What is the multiplier m in the equation $\Delta F = m \cdot \Delta C$? What is the multiplier μ in the equation $\Delta C = \mu \cdot \Delta F$? (The symbol μ is the Greek letter *mu*.) What is the relation between μ and m?

d) Express F as a linear function of C. Graph this function. Put scales and labels on the axes. Indicate clearly the slope of the graph and its vertical intercept.

e) Express C as a linear function of F and graph this function. How are the graphs in parts (d) and (e) related? Give a clear and detailed explanation.

f) Is there any temperature that has the same reading on the two temperature scales? What is it? Does the temperature of the air ever reach this value? Where?

21. **The Greenhouse Effect**. The concentration of carbon dioxide (CO_2) in the atmosphere is increasing. The concentration is measured in parts per million (PPM). Records kept at the South Pole show an increase of .8 PPM per year during the 1960s.

a) At that rate, how many years does it take for the concentration to increase by 5 PPM; by 15 PPM?

b) At the beginning of 1960 the concentration was about 316 PPM. What would it be at the beginning of 1970; at the beginning of 1980?

c) Draw a graph that shows CO_2 concentration as a function of the time since 1960. Put scales on the axes and label everything clearly.

d) The *actual* CO_2 concentration at the South Pole was 324 PPM at the beginning of 1970 and 338 PPM at the beginning of 1980. Plot these values on your graph, and compare them to your calculated values.

e) Using the actual concentrations in 1970 and 1980, calculate a new rate of increase in concentration. Using that rate, estimate what the increase in CO_2

concentration was between 1970 and 1990. Estimate the CO_2 concentration at the beginning of 1990.

f) Using the rate of .8 PPM per year that held during the 1960s, determine how many years before 1960 there would have been *no* carbon dioxide at all in the atmosphere.

22. **Thermal Expansion**. Measurements show that the length of a metal bar increases in proportion to the increase in temperature. An aluminum bar that is exactly 100 inches long when the temperature is 40°F becomes 100.0052 inches long when the temperature increases to 80°F.

a) How long is the bar when the temperature is 60°F? 100°F?

b) What is the multiplier that connects an increase in length ΔL to an increase in temperature ΔT?

c) Express ΔL as a linear function of ΔT.

d) How long will the bar be when $T = 0$°F?

e) Express L as a linear function of T.

f) What temperature change would make $L = 100.01$ inches?

g) For a *steel* bar that is also 100 inches long when the temperature is 40° F, the relation between ΔL and ΔT is $\Delta L = .00067\,\Delta T$. Which expands more when the temperature is increased; aluminum or steel?

h) How long will this steel bar be when $T = 80$°F?

23. **Falling Bodies**. In the simplest model of the motion of a falling body, the velocity increases in proportion to the increase in the time that the body has been falling. If the velocity is given in feet per second, measurements show the constant of proportionality is approximately 32.

a) A ball is falling at a velocity of 40 feet/sec after 1 second. How fast is it falling after 3 seconds?

b) Express the change in the ball's velocity Δv as a linear function of the change in time Δt.

c) Express v as a linear function of t.

The model can be expanded to keep track of the *distance* that the body has fallen. If the distance d is measured in feet, the units of d' are feet per second; in fact, $d' = v$. So the model describing the motion of the body is given by the rate equations

$$d' = v \quad \text{feet per second;}$$
$$v' = 32 \quad \text{feet per second per second.}$$

d) At what rate is the distance increasing after 1 second? After 2 seconds? After 3 seconds?

e) Is d a linear function of t? Explain your answer.

In many cases, the rate of change of a variable quantity is proportional to the quantity itself. Consider a human population as an example. If a city of 100,000 is increasing at the rate of 1500 persons per year, we would expect a similar city of 200,000 to be increasing at the rate of 3000 persons per year. That is, if P is the population at time t, then the **net growth rate** P' is proportional to P:

$$P' = k\,P.$$

In the case of the two cities, we have

$$P' = 1500 = k\,P = k \times 100000 \quad \text{so} \quad k = \frac{1500}{100000} = .015.$$

24. In the equation $P' = k\,P$, above, explain why the units for k are

$$\frac{\text{persons per year}}{\text{person}}.$$

The number k is called the **per capita growth rate**. (“Per capita” means “per person”—“per *head*”, literally.)

25. **Poland and Afghanistan**. In 1985 the per capita growth rate in Poland was 9 persons per year per thousand persons. (That is, $k = 9/1000 = .009$.) In Afghanistan it was 21.6 persons per year per thousand.

a) Let P denote the population of Poland and A the population of Afghanistan. Write the equations that govern the growth rates of these populations.

b) In 1985 the population of Poland was estimated to be 37.5 million persons, that of Afghanistan 15 million. What are the net growth rates P' and A' (as distinct from the *per capita* growth rates)? Comment on the following assertion: When comparing two countries, the one with the larger per capita growth rate will have the larger net growth rate.

c) On the average, how long did it take the population to increase by one person in Poland in 1985? What was the corresponding time interval in Afghanistan?

26. a) **Bacterial Growth**. A colony of bacteria on a culture medium grows at a rate proportional to the present size of the colony. When the colony weighed 32 grams it was growing at the rate of 0.79 grams per hour. Write an equation that links the growth rate to the size of the population.

b) What is ΔP if $\Delta t = 1$ minute? Estimate how long it would take to make $\Delta P = .5$ grams.

27. **Radioactivity**. In radioactive decay, radium slowly changes into lead. If one sample of radium is twice the size of a second lump, then the larger sample will produce twice as much lead as the second in any given time. In other words, the rate of decay is proportional to the amount of radium present. Measurements show that 1 gram of radium decays into lead at the rate of 1/2337 grams per year. Write an equation that links the decay rate to the size of the radium sample. How does your equation indicate that the process involves *decay* rather than *growth*?

28. **Cooling**. Suppose a cup of hot coffee is brought into a room at 70°F. It will cool off, and it will cool off *faster* when the temperature difference between the coffee and the room is greater. The simplest assumption we can make is that the rate of cooling is proportional to this temperature difference (this is called Newton's law of cooling). Let C denote the temperature of the coffee, in °F, and C' the rate at which it is cooling, in °F per minute. The new element here is that C' is proportional, not to C, but to the *difference* between C and the room temperature of 70°F.

a) Write an equation that relates C' and C. It will contain a proportionality constant k. How did you indicate that the coffee is *cooling* and not *heating up*?

b) When the coffee is at 180°F it is cooling at the rate of 9°F per minute. What is k?

c) At what rate is the coffee cooling when its temperature is 120°F?

d) Estimate how long it takes the temperature to fall from 180°F to 120°F. Then make a better estimate, and explain why it is better.

1.3 Using a Program

Computers

A computer changes the way we can use calculus as a tool, and it vastly enlarges the range of questions that we can tackle. No longer need we back away from a problem that involves a lot of computations. There are two aspects to the power of a computer. First, it is fast. It can do a million additions in the time it takes us to do one. Second, it can be programmed. By arranging computations into a loop—as we did on page 15—we can construct a program with only a few instructions that will carry out millions of repetitive calculations.

The purpose of this section is to give you practice using a computer program that estimates values of S, I, and R in the epidemic model. As you will see, it carries out the three rounds of calculations you have already done by hand. It also contains a loop that will allow you to do a hundred, or a million, rounds of calculations with no extra effort.

The Program SIR

The program on the following page calculates values of S, I, and R. It is a set of instructions—sometimes called **code**—that is designed to be read by you and by a computer. These instructions mirror the operations we performed by hand to generate the table on page 13. The code here is similar to what it would be in most programming languages. The line numbers, however, are not part of the program; they are there to help us refer to the lines. A computer reads the code one line at a time, starting at the top. Each line is a complete instruction which causes the computer to do something. The purpose of nearly every instruction in this program is to assign a numerical value to a symbol. Watch for this as we go down the lines of code.

Read a program line by line from the top

Lines 1–5

The first line, `t = 0`, is the instruction "Give t the value 0." The next four lines are similar. Notice, in the fifth line, how Δt is typed out as `deltat`. It is a common practice for the name of a variable to be several letters long. A few lines later S' is typed out as `Sprime`, for instance. The instruction on the sixth line is the first that does not assign a value to a symbol. Instead, it causes the computer to print the following on the computer monitor screen:

Line 6

```
0        45400        2100        2500
```

Program: SIR

```
1    t = 0
2    S = 45400
3    I = 2100
4    R = 2500
5    deltat = 1
6    PRINT t, S, I, R
7    FOR k = 1 TO 3
8            Sprime = -.00001 * S * I              }
9            Iprime = .00001 * S * I - I / 14      } Step I
10           Rprime = I / 14                       }
11           deltaS = Sprime * deltat
12           deltaI = Iprime * deltat
13           deltaR = Rprime * deltat
14           t = t + deltat
15           S = S + deltaS
16           I = I + deltaI
17           R = R + deltaR
18           PRINT t, S, I, R
19   NEXT k
```

Line 7

Skip over the line that says `FOR k = 1 TO 3`. It will be easier to understand after we've read the rest of the program.

Lines 8–10

Look at the first three indented lines. You should recognize them as coded versions of the rate equations

$$S' = -.00001\,SI,$$
$$I' = .00001\,SI - I/14,$$
$$R' = I/14,$$

for the measles epidemic. (The program uses `*` to denote multiplication.) They are instructions to assign numerical values to the symbols S', I', and R'. For instance, `Sprime = -.00001 * S * I` (line 8) says

Give S' the value $-.00001\,SI$;
use the current values of S and I to get $-.00001\,SI$.

Now the computer knows that the current values of S and I are 45400 and 2100, respectively. (Can you see why?) So it calculates the product

$-.00001 \times 45400 \times 2100 = -953.4$ and then gives S' the value -953.4. There is an extra step to calculate the product.

Notice that the first three indented lines are bracketed together and labelled "Step I," because they carry out Step I in the flow chart. The next three indented lines carry out Step II in the flow chart. They assign values to three more symbols—namely ΔS, ΔI, and ΔR—using the current values of S', I', R' and Δt. Lines 11–13

The next four indented lines present a puzzle. They don't make sense if we read them as ordinary mathematics. For example, in an expression like `t = t + deltat`, we would cancel the `t`'s and conclude `deltat = 0`. The lines *do* make sense when we read them as computer instructions, however. As a computer instruction, `t = t + deltat` says Lines 14–17

Make the new value of t equal to the current value of $t + \Delta t$.

(To make this clear, some computer languages express this instruction in the form `let t = t + deltat`.) Once again we have an instruction that assigns a numerical value to a symbol, but this time the symbol (t, in this case) already has a value before the instruction is carried out. The instruction gives it a *new* value. (Here the value of t is changed from 0 to 1.) Likewise, the instruction `S = S + deltaS` gives S a new value. What was the old value, and what is the new?

Compare the three lines of code that produce new values of S, I, and R with the original equations that we used to define Step III back on page 16: How a program computes new values

`S = S + deltaS`	new $S =$ current $S + \Delta S$,
`I = I + deltaI`	new $I =$ current $I + \Delta I$,
`R = R + deltaR`	new $R =$ current $R + \Delta R$.

The words "new" and "current" aren't needed in the computer code because they are automatically understood to be there. Why? First of all, a symbol (like `S`) always has a *current* value, but an instruction can give it a *new* value. Second, a computer instruction of the form `A = B` is always understood to mean "new `A` = current `B`."

Notice that the instructions `A = B` and `B = A` mean different things. The second says "new `B` = current `A`." Thus, in `A = B`, `A` is altered to equal `B`, while in `B = A`, `B` is altered to equal `A`. To emphasize that the symbol on the left is always the one affected, some programming languages use a modified equal sign, as in `A := B`. We sometimes read this as "`A` gets `B`".

Line 18

The next line is another `PRINT` statement, exactly like the one on line 6. It causes the current values of t, S, I, and R to be printed on the computer monitor screen. But this time what appears is

```
1      44446.6      2903.4      2650
```

The values were changed by the previous four instructions. It is important to remember that the computer carries out instructions in the order they are written. Had the second `PRINT` statement appeared right after line 13, say, the old values of t, S, I, and R would have appeared on the monitor screen a second time.

Lines 7 and 19: the loop

We will take the last line and line 7 together. They are the instructions for the loop. Consider the situation when we reach the last line. The variables t, S, I, and R now have their "day 1" values. To continue, we need an instruction that will get us back to line 8, because the instructions on lines 8–17 will convert the current (day 1) values of t, S, I, and R into their "day 2" values. That's what lines 7 and 19 do.

Here is the meaning of the instruction `FOR k = 1 TO 3` on line 7:

`FOR k = 1 TO 3`

Give k the value 1, and be prepared later to give it the value 2 and then the value 3.

The variable k plays the role of a **counter**, telling us how many times we have gone around the loop. Notice that k did not appear in our hand calculations. However, when we said we had done three rounds of calculations, for example, we were really saying $k = 3$.

After the computer reads and executes line 7, it carries out all the instructions from lines 8 to 18, arriving finally at the last line. The computer then interprets the instruction `NEXT k` on the last line as follows:

`NEXT k`

Give k the next value that the `FOR` command allows, and move back to the line immediately after the `FOR` command.

After the computer carries out this instruction, k has the value 2 and the computer is set to carry out the instruction on line 8. It then executes that instruction, and continues down the program, line by line, until it reaches line 19 once again. This sets the value of k to 3 and moves the computer back to line 8. Once again it continues down the program to line 19. This time there is no allowable value that k can be given, so the program stops.

How the program stops

The `NEXT k` command is different from all the others in the program. It is the only one that directs the computer to go to a different line. That action causes the program to **loop**. Because the loop involves all the indented instructions between the `FOR` statement and the `NEXT` statement, it is called a **FOR–NEXT loop**. This is just one kind of loop. Computer programs can contain other sorts that carry out different tasks. In the next chapter we will see how a **DO–WHILE** loop is used.

Exercises

The program SIR

The object of these exercises is to verify that the program SIR works the way the text says it does. Follow the instructions for running a program on the computer you are using.

1. Run the program to confirm that it reproduces what you have already calculated by hand (table, page 13).

2. On a copy of the program, mark the instructions that carry out the following tasks:

a) give the input values of S, I, and R;

b) say that the calculations take us 1 day into the future;

c) carry out step II (see page 16);

d) carry out step III;

e) give us the output values of S, I, and R;

f) take us once around the whole loop;

g) say how many times we go around the loop.

3. Delete all the lines of the program from line 7 onward (or else type in the first 6 lines). Will this program run? What will it do? Run it and report what you see. Is this what you expected?

4. Starting with the original SIR program on page 50, delete lines 7 and 19. These are the ones that declare the FOR–NEXT loop. Will this program run? What will it do? Run it and report what you see. Is this what you expected?

5. Using the 17-line program you constructed in the previous question, remove the PRINT statement from the last line and insert it between what appear as lines 13 and 14 on page 50. Will this program run? What will it do? Run it and report what you see. Is this what you expected?

6. Starting with the original SIR program on page 50, change line 7 so it reads `FOR k = 26 TO 28`. Thus, the counter k takes the values 26, 27, and 28. Will this program run? What will it do? Run it and report what you see. Is this what you expected?

Programs to practice on

In this section there are a number of short programs for you to analyze and run.

Program 1

```
A = 2
B = 3
A = B
PRINT A, B
```

Program 2

```
A = 2
B = 3
B = A
PRINT A, B
```

Program 3

```
A = 2
B = 3
A = A + B
B = A + B
PRINT A, B
```

7. When Program 1 runs it will print the values of A and B that are current when the program stops. What values will it print? Type in this program and run it to verify your answers.

8. What will Program 2 do when it runs? Type in this program and run it to verify your answers.

9. After each line in Program 3 write the values that A and B have *after* that line has been carried out. What values of A and B will it print? Type in this program and run it to verify your answers.

The next three programs have an element not found in the program SIR. In each of them, there is a FOR–NEXT loop, and the counter k actually appears in the statements within the loop.

Program 4

```
FOR k = 1 TO 5
     A = k ^ 3
     PRINT A
NEXT k
```

Program 5

```
FOR k = 1 TO 5
     A = k ^ 3
NEXT k
PRINT A
```

Program 6

```
x = 0
FOR k = 1 TO 5
     x = x + k
     PRINT k, x
NEXT k
```

10. What output does Program 4 produce? Type in the code and run the program to confirm your answer.

11. What is the difference between the code in Program 5 and the code in Program 4? What is the output of Program 5? Does it differ from the output of Program 4? If so, why?

12. What output does Program 6 produce? Type in the code and run the program to confirm your answer.

Program 7

```
A = 0
B = 0
FOR k = 1 TO 5
     A = A + 1
     B = A + B
     PRINT A, B
NEXT k
```

Program 8

```
A = 0
B = 1
FOR k = 1 TO 5
     A = A + B
     B = A + B
     PRINT A, B
NEXT k
```

Program 9

```
A = 0
B = 1
FOR k = 1 TO 5
     A = A + B
     B = A + B
NEXT k
PRINT A, B
```

13. Program 7 prints five lines of output. What are they? Type in the program and run it to confirm your answers.

14. What is the output of Program 8? Type in the program and run it to confirm your answers.

15. Describe exactly how the *codes* for Programs 8 and 9 differ. How do the *outputs* differ?

Analyzing the measles epidemic

16. Alter the program SIR to have it calculate estimates for S, I, and R over the first *six* days. Construct a table that shows those values.

17. Alter the program to have it estimate the values of S, I, and R for the first *thirty* days.

a) What are the values of S, I, and R when $t = 30$?

b) According to these figures, on what day does the infection peak? What values do you get for S, I, and R?

c) How can you reconcile the value you just got for S with the value we obtained algebraically on page 18?

18. By adding an appropriate PRINT statement after line 10 you can also get the program to print values for S', I', and R'. Do this, and check that you get the values shown in the table on page 13.

19. According to these estimates, on what day do the largest number new infections occur? How many are there? Explain where you got your information.

20. On what day do you estimate that the largest number of recoveries occurs? Do you see a connection between this question and 17 (b)?

21. On what day do you estimate the infected population grows most rapidly? Declines most rapidly? What value does I' have on those days?

22. a) Alter the original SIR program so that it will go *backward* in time, with time steps of 1 day. Specify the changes you made in the program. Use this altered program to obtain estimates for the values of S, I, and R yesterday. Compare your estimates with those in the text (page 13).

b) Estimate the values of S, I, and R *three* days before today.

23. According to the S-I-R model, when did the infection begin? That is, how many days before today was the estimated value of I approximately 1?

24. **There and back again**. Use the SIR program, modified as necessary, to carry out the calculations described in exercise 18 on page 23. Do your computer results agree with those you obtained earlier?

25. In exercises 22–26 at the end of section 2 (pages 46–48) you set up rate equations to model some other systems. Choose a couple of these and think of some interesting questions that could be answered using a suitably modified SIR program. Make the modifications and report your results.

1.4 Chapter Summary

The Main Ideas

- Natural processes like the spread of disease can often be described by **mathematical models**. Initially, this involves identifying **numerical quantities** and relations between them.

- A relation between quantities often takes the form of a **function**. A function can be described in many different ways; **graphs**, **tables**, and **formulas** are among the most common.

- **Linear functions** make up a special but important class. If y is a linear function of x, then $\boldsymbol{\Delta y = m \cdot \Delta x}$, for some constant m. The constant m is a **multiplier**, **slope**, and **rate of change**.

- If $y = f(x)$, then we can consider the **rate of change** $\boldsymbol{y'}$ of y with respect to x. A mathematical model whose variables are connected by **rate equations** can be analyzed to predict how those variables will change.

- Predicted changes are **estimates** of the form $\boldsymbol{\Delta y = y' \cdot \Delta x}$.

- The computations that produce estimates from rate equations can be put into a **loop**, and they are readily carried out on a **computer**.

- A computer increases the **scope** and **complexity** of the problems we can consider.

Expectations

- You should be able to work with functions given in various forms, to find the output for any given input.

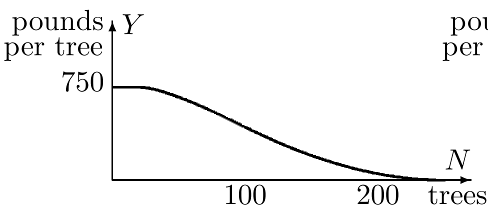
- You should be able to read a graph. You should also be able to construct the graph of a linear function directly, and the graph of a more complicated function using a computer graphing package.
- You should be able to determine the natural domain of a function given by a formula.
- You should be able to express proportional quantities by a linear function, and interpret the constant of proportionality as a multiplier.
- Given any two of these quantities for a linear function—multiplier, change in input, change in output—you should be able to determine the third.
- You should be able to model a situation in which one variable is proportional to its rate of change.
- Given the value of a quantity that depends on time and given its rate of change, you should be able to estimate values of the quantity at other times.
- For a set of quantities determined by rate equations and initial conditions, you should be able to estimate how the quantities change.
- Given a set of rate equations, you should be able to determine what happens when one of the quantities reaches a maximum or minimum, or remains unchanged over time.
- You should be able to understand how a computer program with a FOR–NEXT loop works.

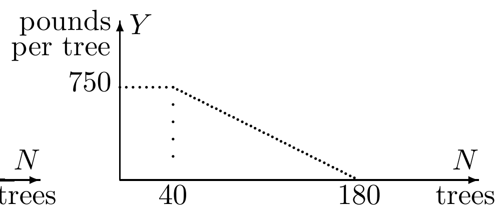
Chapter Exercises

A Model of an Orchard

If an apple orchard occupies one acre of land, how many trees should it contain so as to produce the largest apple crop? This is an example of an **optimization** problem. The word *optimum* means "best possible"—especially, the best under a given set of conditions. These exercises seek an optimum by analyzing a simple mathematical model of the orchard. The model is the function that describes how the total yield depends on the number of trees.

An immediate impulse is just to plant a lot of trees, on the principle: more trees, more apples. But there is a catch: if there are too many trees in a single acre, they crowd together. Each tree then gets less sunlight and nutrients, so it produces fewer apples. For example, the relation between the *yield per tree*, Y, and the *number of trees*, N, may be like that shown in the graph drawn on the left, below.





When there are only a few trees, they don't get in each other's way, and they produce at the maximum level—say, 750 pounds per tree. Hence the graph starts off level. At some point, the trees become too crowded to produce *anything*! In between, the yield per tree drops off as shown by the curved middle part of the graph.

We want to choose N so that the *total yield*, T, will be as large as possible. We have $T(N) = Y(N) \cdot N$, but since we don't know $Y(N)$ very precisely, it is difficult to analyze $T(N)$. To help, let's replace $Y(N)$ by the approximation shown in the graph on the right. Now carry out an analysis using this graph to represent $Y(N)$.

1. Find a formula for the straight segment of the new graph of $Y(N)$ on the interval $40 \leq N \leq 180$. What is the formula for $T(N)$ on the same interval?

2. What are the formulas for $T(N)$ when $0 \leq N \leq 40$ and when $180 \leq N$? Graph T as a function of N. Describe the graph in words.

3. What is the maximum possible total yield T? For which N is this maximum attained?

4. Suppose the endpoints of the sloping segment were P and Q, instead of 40 and 180, respectively. Now what is the formula for $T(N)$? (Note that P and Q are *parameters* here. Different values of P and Q will give different models for the behavior of the total output.) How many trees would then

produce the maximum total output? Expect the maximum to depend on the parameters P and Q.

Rate Equations

Do the following exercises by hand. You may wish to check your answers by using suitable modifications of the program SIR.

5. **Radioactivity**. From exercise 25 of section 2 we know a sample of R grams of radium decays into lead at the rate

$$R' = \frac{-1}{2337}R \quad \text{grams per year.}$$

Using a step size of 10 years, estimate how much radium remains in a 0.072 gram sample after 40 years.

6. **Poland and Afghanistan**. If P and A denote the populations of Poland and Afghanistan, respectively, then their net per capita growth rates imply the following equations:

$$\begin{aligned} P' &= .009\,P \quad \text{persons per year;} \\ A' &= .0216\,A \quad \text{persons per year.} \end{aligned}$$

(See exercise 23 of section 2.) In 1985, $P = 37.5$ million, $A = 15$ million. Using a step size of 1 year, estimate P and A in 1990.

7. **Falling bodies.** If d and v denote the distance fallen (in feet) and the velocity (in feet per second) of a falling body, then the motion can be described by the following equations:

$$\begin{aligned} d' &= v \quad \text{feet per second;} \\ v' &= 32 \quad \text{feet per second per second.} \end{aligned}$$

(See exercise 21 of section 2.) Assume that when $t = 0$, $d = 0$ feet and $v = 10$ feet/sec. Using a step size of 1 second, estimate d and v after 3 seconds have passed.