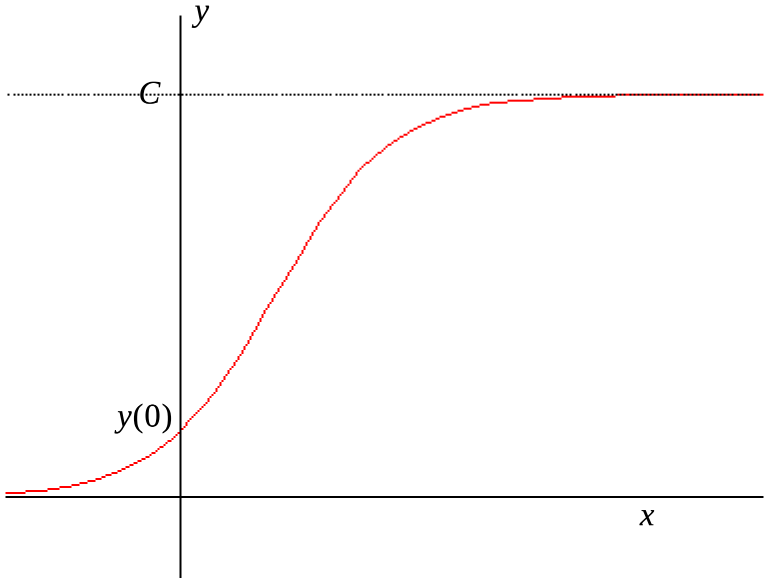
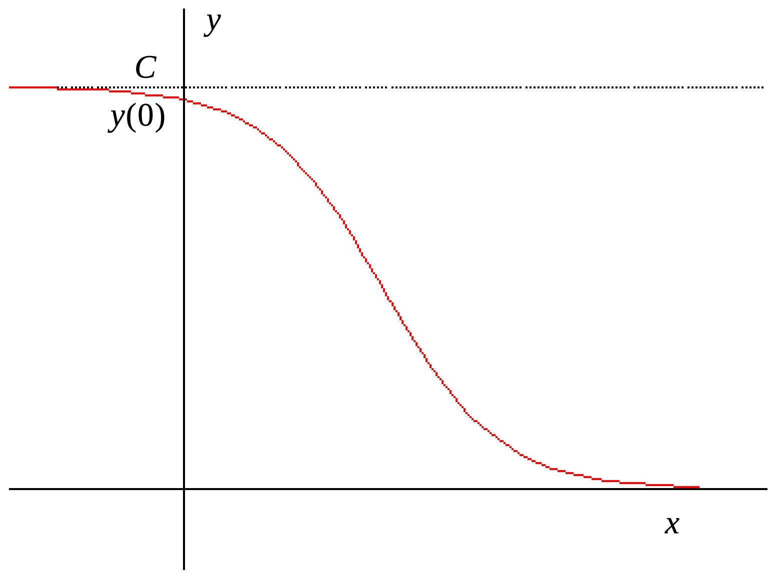
In the past weeks, we have considered the use of linear, exponential, power and polynomial functions as mathematical models in many different contexts. Another type of function, called the **logistic function**, occurs often in describing certain kinds of growth. These functions, like exponential functions, grow quickly at first, but because of restrictions that place limits on the size of the underlying population, eventually grow more slowly and then level off.



As is clear from the graph above, the characteristic S-shape in the graph of a logistic function shows that initial exponential growth is followed by a period in which growth slows and then levels off, approaching (but never attaining) a maximum upper limit.

Logistic functions are good models of biological population growth in species which have grown so large that they are near to saturating their ecosystems, or of the spread of information within societies. They are also common in marketing, where they chart the sales of new products over time; in a different context, they can also describe demand curves: the *decline* of demand for a product as a function of increasing price can be modeled by a logistic function, as in the figure below.



The formula for the logistic function,

$$y = \frac{C}{1 + Ae^{-Bx}}$$

involves three parameters A, B, C. (Compare with the case of a quadratic function $y = ax^2 + bx + c$ which also has three parameters.) We will now investigate the meaning of these parameters.

First we will assume that the parameters represent positive constants. As the input x grows in size, the term $-Bx$ that appears in the exponent in the denominator of the formula becomes a larger and larger negative value. As a result, the term e^{-Bx} becomes smaller and smaller (since raising any number bigger than 1, like e, to a negative power gives a small positive answer). Hence the term Ae^{-Bx} also becomes smaller and smaller. Therefore, the entire denominator $1 + Ae^{-Bx}$ is always a number larger than 1 and decreases to 1 as x gets larger. Finally then, the value of y, which equals C divided by this denominator quantity, will always be a number smaller than C and increasing to C. It follows therefore that the parameter

C represents the limiting value of the output

past which the output cannot grow (see the figures above).

On the other hand, when the input x is near 0, the exponential term Ae^{-Bx} in the denominator is a value close to A so that the denominator $1 + Ae^{-Bx}$ is a value near $1 + A$. Again, since y is computed by dividing C by this denominator, the value of y will be a quantity much smaller than C. Looking at the graph of the logistic curve in Figure 1, you see that this analysis explains why y is small near $x = 0$ and approaches C as x increases.

To identify the exact meaning of the parameter A, set $x = 0$ in the formula; we find that

$$y(0) = \frac{C}{1 + Ae^{-B \cdot 0}} = \frac{C}{1 + A}$$

Clearing the denominator gives the equation $(1 + A)y(0) = C$. One way to interpret this last equation is to say that

the limiting value C is $1 + A$ times larger than the initial output $y(0)$

An equivalent interpretation is that

A is the number of times that the initial population must grow to reach C

The parameter B is much harder to interpret exactly. We will be content to simply mention that

if B is positive, the logistic function will always increase, while if B is negative, the function will always decrease

(see Exercise 9).

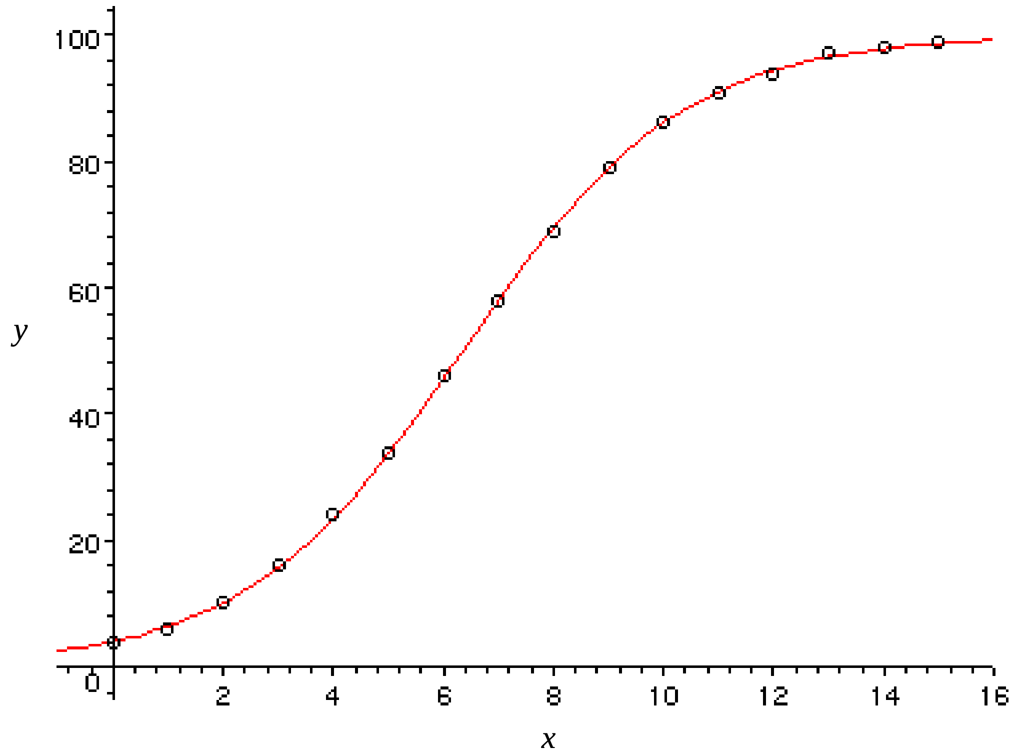
Let us illustrate these ideas with an example.

Example 1. Construct a scatterplot of the following data:

x	0	1	2	3	4	5	6	7
y	4	6	10	16	24	34	46	58

x	8	9	10	11	12	13	14	15
y	69	79	86	91	94	97	98	99

The scatterplot should clearly indicate the appropriateness of using a logistic model to fit this data. You will find that using the parameters $A = 24$, $B = 0.5$, $C = 100$ produce a model with a very good fit.



Notice that no matter how large you take x, y will never exceed 100, illustrating that $C = 100$ represents the limiting upper bound for the output. Also, from the data, $y(0) = 4$, and C has a value $1 + A = 25$ times larger than $y(0)$. Finally, B is positive and we observe that y is always increasing.

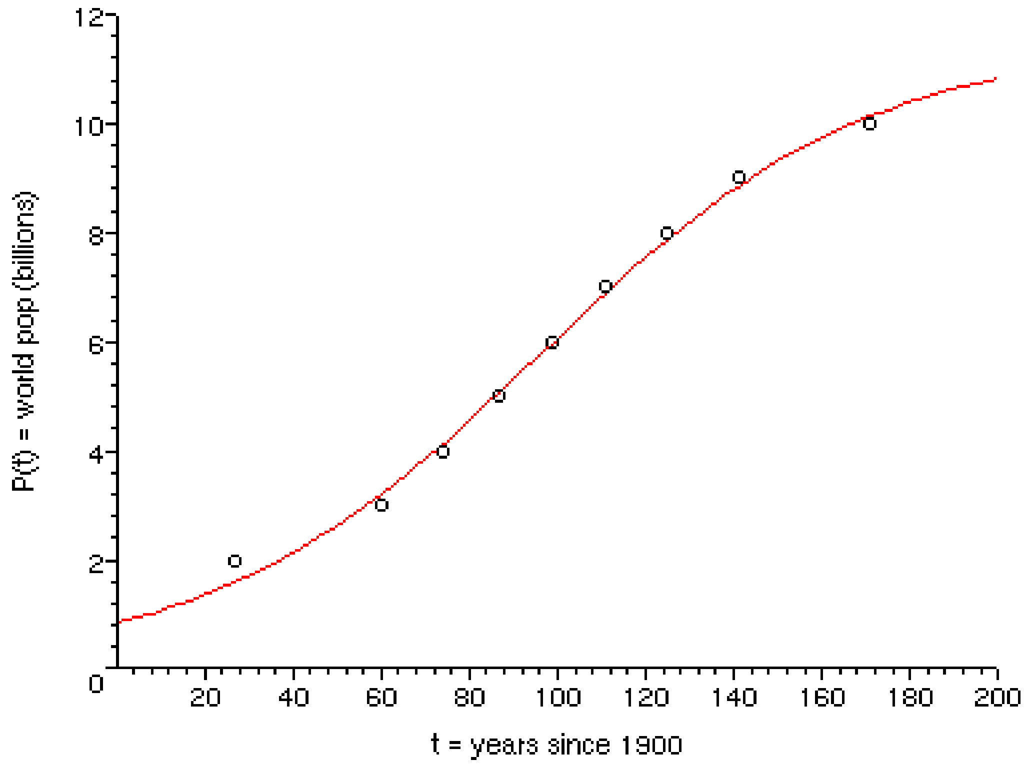
Example 2. The following table shows that results of a study by the United Nations (*New York Times*, November 17, 1995) which has found that world population growth is slowing. It indicates the year in which world population has reached a given value:

Year	1927	1960	1974	1987	1999	2011	2025	2041	2071
Billions	2	3	4	5	6	7	8	9	10

Construct a scatterplot of the data, using the input variable t = # years since 1900 and output variable P = world population (in billions). You will notice the characteristic S-shape typical of logistic functions. It turns out that $A = 12.8$, $B = 0.0266$, $C = 11.5$ are parameter values that yield a logistic function with a good fit to this data:

$$P(t) = \frac{11.5}{1 + 12.8e^{-0.0266t}}$$

Confirm this by graphing the curve atop the scatterplot. Notice, by the way, that the function is increasing and that, as we expect, B is positive.



Build a table of values of the function for $t = 0, 50, 100, 150$, etc. Observe how the values of the function increase with $C = 11.5$ as limiting value. In fact, by the time $x = 300$, the value of P already equals 11.5 (to one decimal place accuracy). Also note that the initial population is $P(0) = 0.833$; that is, in 1900, the world population is 833 million. We also confirm that

$$(1 + A)P(0) = (1 + 12.8)(0.833) = 11.5 = C$$

We have seen that the parameter C refers to the long run behavior of the function and that A describes the relation between the initial and limiting output values. Another important feature of any logistic curve is related to its shape: **every logistic curve has a single inflection point which separates the curve into two equal regions of opposite concavity**. It is easy to identify the precise coordinates of this inflection point: because of the symmetry of the curve about this point, it must occur halfway up the curve at height $y = C/2$. Setting y equal to $C/2$ in the formula yields

$$\frac{C}{2} = \frac{C}{1 + Ae^{-Bx}}$$

$$\frac{2}{C} = \frac{1 + Ae^{-Bx}}{C}$$

$$2 = 1 + Ae^{-Bx}$$

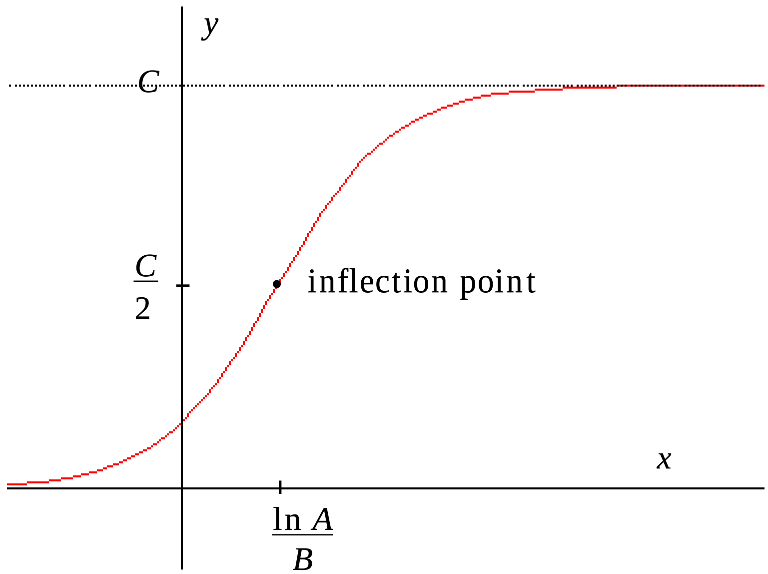
$$1 = Ae^{-Bx}$$

$$e^{Bx} = A$$

$$Bx = \ln A$$

$$x = \frac{\ln A}{B}$$

This shows that **the inflection point has coordinates** $\left(\frac{\ln A}{B}, \frac{C}{2}\right)$. Another consequence of this result is that we can determine the value of *B* if we know the coordinates of the inflection point.



Example 2, again. The inflection point on the world population curve occurs when $t = \ln A / B = \ln 12.8 / 0.0266 = 95.8$. In other words, according to the model, in 1995 world population attained 6.75 billion, half its limiting value of 11.5 billion. From this year on, population will continue to increase but at a slower and slower rate.

Example 3. The percentage of U.S. households that own a VCR has risen steadily since their introduction in the late 1970s. We will measure time with the input variable t = # years since 1980, and percentages of U.S. households that own a VCR with the output variable $p = f(t)$.

Year	1978	1979	1980	1981	1982	1983	1984
%	0.3	.05	1.1	1.8	3.1	5.5	10.6

Year	1985	1986	1987	1988	1989	1990	1991
%	20.8	36	48.7	58	64.6	71.9	71.9

A look at a scatterplot of these data makes clear that a logistic function provides a good model. It also appears that the percentage of households with VCRs seems to be quickly approaching its limiting value in the low-seventy percent range. It makes sense for us to estimate this limiting value as $C = 75$.

Next, since our data give us that $f(0) = 1.1$, we have the following equation for A: $(1 + A)f(0) = C$, or $(1 + A)1.1 = 75$. Therefore, $1 + A = 68.2$, or $A = 67.2$. Finally, the data point in the middle of the plot, (6, 36), is a good candidate for our inflection point since its y-coordinate is roughly one-half the limiting value. Consequently, we can set $6 = \ln A / B = \ln 67.2 / B$ to get an equation we can solve for B: $B = \ln 67.2 / 6 = 0.701$. We have thus created a set of values for our parameters, giving the logistic function

$$p = f(t) = \frac{75}{1 + 67.2e^{-0.701t}}$$

A graph of the function over the scatterplot shows the nice fit.

On the other hand, your calculator will also provide a logistic regression function with different values for the parameters (in this case, it should give $A = 115.1$, $B = 0.769$, $C = 73.7$) but it, too, provides a nice fit.

1. The Federal Bureau of Alcohol, Tobacco and Firearms recorded the following annual statistics on the number of bombings that have taken place in the U.S. (taken from the Cincinnati Enquirer, July 31, 1996):

Year	1989	1990	1991	1992	1993	1994
# of Bombings	1699	2062	2490	2989	2980	3163

(a) Construct a scatterplot for the data and graph with it the logistic function

$$y = \frac{3350}{1 + 1.025e^{-0.5814x}}$$

Here, x = # years since 1989 and y = # of bombings in year x. Do you think this function provides a good fit for the data? What features of the data are modeled by the curve?

(b) According to the model, how many bombings should be expected in the U.S. in the year 1996?

(c) What will happen in the long run to the number of bombings?

2. Of a group of 200 college men the number N who are taller than x inches is given below:

x:	65	66	67	68	69	70	71	72	73	74	75	76
N:	197	195	184	167	139	101	71	43	25	11	4	2

(a) Explain why a logistic model is appropriate for this data. (Construct a scatterplot.)

(b) It turns out that the function

$$N = \frac{200}{1 + 0.015e^{0.8x}}$$

is a good fit to the data. Identify the parameters A, B, C and explain why the given value of C is appropriate.

(c) Determine the value of A directly from the data, and compare it to the parameter in the function in (b).

(d) Identify from your scatterplot the approximate coordinates of the inflection point. Use this and the value of A you calculated in (b) to determine B; compare with the value in the function in (b).

3. The Ohio Department of Health released the following data (*Cincinnati Enquirer*, December 11, 1994) tallying the number of newly diagnosed cases of AIDS in the state:

Year	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Cases	2	8	27	58	121	209	394	533	628	674	746	725

(a) Build a scatterplot of the data. Use the input variable t = # years since 1981. Explain why a logistic model is appropriate here.
(b) Construct a formula for $c(t)$ = # new cases of AIDS reported in Ohio in year t as follows. First, explain why $C = 750$ is a reasonable choice for this parameter.
(c) Determine a corresponding value of A.
(d) Estimate from the plot the x-coordinate of the inflection point for the model. Use this estimate to determine a corresponding value for B.
(e) Write down the final formula for $c(t)$.

4. The National Foundation for Infantile Paralysis reported the following statistics on the total number of polio cases during the 1949 epidemic in the U.S.:

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
cases	494	759	1016	1215	1619	2964	8489	22377	32618	38153	41462	42375

Find a logistic function to fit this data. Compare with the model obtained by using the logistic regression feature on your calculator.

5. There is deep concern about world oil reserves: some geologists are finding that there is less oil to be had from drilling worldwide. Here is data reporting the number of trillions of barrels available worldwide.

Year	1935	1945	1955	1965	1975	1985
Reserves	177.9	337.8	616.3	1022.1	1233.5	1427.4

(a) Find a logistic model for the data.
(b) According to your model, what is the upper limit for world oil reserves?
(c) In what year, according to the model, will we have obtained 95% of world oil reserves?

6. The following table gives the size of Toronto's Jewish population through the 20th century

Year	1901	1911	1921	1931	1941	1951	1961	1971	1981	1991
Pop	3103	18294	34770	46751	52798	66773	85000	97000	128650	162605

(a) Find a logistic model for the data.

(b) What is the limiting value of the size of the Jewish population of Toronto, according to the model?
(c) Give the coordinates of the inflection point on the curve. Interpret this in the context of the data.

7. Like most states, South Carolina has struggled with the health crisis of AIDS. Here is data giving the annual number of reported cases in the state differentiated by sex.

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Males	435	418	382	483	585	819	613	450	399	285
Females	274	288	354	418	488	716	661	472	406	287

(a) Construct a table that records $c(t)$ = the *cumulative* number of AIDS cases in South Carolina t years since 1986 for *both* sexes.
(b) Fit a logistic function to the data for $c(t)$.
(c) What is the limiting value of the number of cases of AIDS in South Carolina, according to the model?
(d) Find the coordinates of the inflection point on the curve. Interpret this in the context of the data.

8. (a) Consider a human female 20 inches long at birth who grows logistically to a mature height of 68 inches. If she reaches 95% of her mature height at age 15, at what age is she growing most quickly?
(b) Repeat part (a) in the case of a horse which is 36 inches tall at birth, grows to a height of 58 inches at maturity, and reaches 95% of its mature height at age 2.

9. Describe the long run behavior of a logistic function in the case where B is negative.