Unit 1: The Modeling Process

1.1 What is a model?

Mathematics is an extremely useful tool that can be applied to many interesting problems in the real world. It is almost impossible to turn on the television news or open a newspaper and not find some article in which mathematical analysis played a crucial role. Reports about the economy, health issues, the weather, and so on, all involve mathematics. Advertising is based on market surveys, another mathematical tool. Arguments about political issues (land use, taxes, death penalty, etc.) have to be based on real data. However:

Theorem: The real world is complicated.

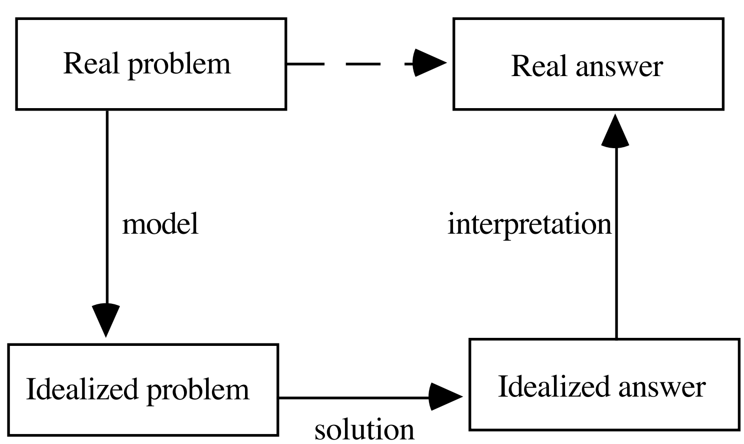
Corollary: In order to analyze the real world, we must simplify it.

It is impossible to capture all the features of the real world in a mathematical analysis. Many variables, with complex interactions, affect the outcome. We must decide which features are most important and which can be ignored.

The process of simplifying the real world into a mathematically tractable form is called *mathematical modeling*.

Definition: A *mathematical model* is a set of assumptions used to simplify a real world problem in order to permit mathematical analysis.

The diagram below shows the basic principle of mathematical modeling. We start with a real world problem. Our goal is to get to the real answer, but the direct path is too difficult. Therefore, we create a model that is an idealization of the real problem but which can be solved mathematically. This gives us an idealized answer which we then have to interpret in the context of the original problem.



- Modeling is an iterative process.

We start with as simple a model as possible. If, after we solve the model, the answer does not agree with observed data or with our intuition, then we start again with a more complicated model. This process continues either until we get a model that gives us reasonable answers or becomes too complicated to solve.

1.2 Types of Models

Models can be roughly divided into two types:

1. Deterministic

These models apply to situations in which the behavior being modelled is predictable and repeatable under identical conditions. For example, if we drop a brick from the top of a building, it will hit the ground in a certain time at a certain speed. If we then drop another brick from the same building under the same conditions, it will hit the ground in the same amount of time and at the same speed as the first brick. Or, if we mix hydrochloric acid (HCl) and sodium hydroxide (NaOH) in precise amounts, we always get predictable amounts of sodium chloride (NaCl) and water (H_2O).

2. Probabilistic

These models apply to situations in which the behavior is not predictable. For instance, if we flip a coin, there is no way to know in advance if it will land on heads or tails. Even if we flip one and note that it lands on heads, the next coin, flipped under "identical conditions" (whatever that means) is just as likely to land on heads or tails.

It is much harder to know when a probabilistic model is correct. If we assume the coin is fair, meaning that heads and tails are equally likely to occur, then the probability of heads is 0.5. What this means is that if we flip the coin many times, then we expect about 50% of the outcomes to be heads. This is a statement about *long-term* behavior so, in order to assess the model's validity, we have to repeat the experiment many times. This differs from the deterministic case where the model predicts an answer that is either right or wrong each time the experiment is performed.

In practice, it may be expensive or otherwise infeasible to repeat the experiment many times. For instance, if we want to claim that a new medicine is 95% effective in treating a certain disease, we have to run clinical trials during which some people will not be cured, or worse. Or if we want to determine the probability that the occupants of a new car can survive a crash of 30 mph into a brick wall, we are going to have to seriously damage a few cars to get meaningful results.

The dividing line between deterministic and probabilistic models is not always well-defined. Many situations have both deterministic and probabilistic features. Suppose we want to model the growth of a population over time. If we are only interested in trends and approximations to the population size, we can view the situation deterministically. However, random factors such as climate, disease, and temporary food shortages affect the rate of growth of the population. One way to think about this is that there is a deterministic part of the model that predicts what we *expect* to happen and a random part that accounts for deviations between what we expect and what we actually see.

There is also a distinction between *discrete* and *continuous* models. In a discrete model, either the variables of interest are integer-valued or the system is observed at discrete points in time. In a continuous model, the variables can take on any (perhaps only positive) real value or the system is observed continuously in time. Discrete models often require linear algebra and difference equations to solve. Continuous models often require calculus and differential equations. Sometimes discrete models are approximated by continuous models because the continuous model is easier to solve. Sometimes it's the other way around.

1.3 Approaches to modeling

There are three basic approaches we can use to create a model.

1. Theoretical:

In this approach, we need to have some inherent theory about how the system behaves. For example, suppose we want to study the motion of a drag racer. If we assume that the acceleration of the racer is constant and that its initial velocity is 0, then its position is given by $y = \frac{1}{2}at^2$, where a is the acceleration. We can check whether this is reasonable by comparing the actual position of the drag racer to that predicted by this model. If they don't agree, then we can try to create a more complicated theoretical model. For instance, the acceleration may depend on the mass of the car which changes during the race as fuel is used up. Or, the acceleration may depend on air resistance, which may be a function of the velocity.

2. Empirical:

Another approach is to collect data giving the position of the drag racer at various points in time and then try to find a curve that fits the data. That is, find a function f such that the position at any time t is given by $s = f(t)$. This is an *empirical* approach to modeling and is used when no adequate theory can be developed. The disadvantage to empirical modeling is that it is only as good as the data collected. *Bad data will give bad results*. Furthermore, any results we get are only valid under the conditions used to collect the data. If we want to see what effect changing the conditions has on the results, we have to collect another set of data. Sometimes an empirical model can give us a clue about a theoretical model.

3. Simulation:

There is another approach to modeling that has become feasible since the advent of modern computers. It is possible to create a *computer simulation* of the system being modeled. Inherently, the simulation is based on a set of assumptions about the behavior of the system. However, the computer allows us to inexpensively repeat the experiment thousands of times and also makes it easier to change the parameters of the model. For example, suppose we wanted to design a system of elevators for a high-rise building. We could gather information about the number of people expected to use the elevator at various times of day and their intended destinations. Then we could have the computer experiment with many different configurations and evaluate their effectiveness, perhaps measured by the average amount of time riders need to wait for an elevator that goes to their destination. We could then choose the best of these configurations. (This would not necessarily be the optimal solution; it is just the best of the options we considered.)

1.4 Steps in the modeling process

1. Identify the problem

What are you trying to model? What is the response variable? On what factors does it depend? Sometimes this is the hardest part.

2. Decide on the important variables

Pick a subset of the explanatory variables that is likely to have the most influence on the response. Ignore the others or treat them as constants. You may have to modify your selection in a future iteration.

3. Make assumptions

Hypothesize a relationship between the response and explanatory variables. This may be expressed as one or more equations, or as a graph.

4. Solve the model

Do the math. Solve the equations, draw graphs, make statistical inferences. Approximations may be necessary. (**Corollary**: The real world is nonlinear. Nonlinear systems are usually harder to solve analytically.)

5. Validate the model

Do the conclusions of the model agree with observations and intuition? If not, change the variables, change the assumptions, check the math.

6. Implement the model

The modeler is often not the user of the results. You must be able to explain the results in a language that the user understands. Two of Murphy's Laws apply:

- You can make it foolproof, but you can't make it damnfoolproof.
- Anyone can make a mistake, but it takes an expert to really screw things up.

7. Maintain the model

The results of the model often depend on factors that change over time. The model should be flexible enough to accommodate changing data without starting over.

1.5 Two examples

1.5.1 Population growth

Suppose we wanted to predict the population of a certain species of fish as a function of time. (This is an important problem in wildlife management and affects decisions about fishing regulations such as licensing, catch limitations, etc.).

Let $Y(t)$ be the population of fish at time t. The rate at which the population changes is equal to the birth rate – death rate.

Assumption: The birth and death rates are proportional to the size of the population. In other words, a constant fraction of the population gives birth in a given time period and a constant fraction dies in a given time period.

Under this assumption, $Y(t)$ is governed by the first-order differential equation $\dfrac{dY}{dt} = rY$, where $r > 0$ is a constant that represents the difference between the birth and death rates.

The solution to this equation is $Y = Y_0 e^{rt}$, where Y_0 is the initial population.

This model implies that the population increases without bound. (See graph below.) In the short term, this is reasonable but, in the long term, it may not be. The environment cannot support an infinite number of fish; sooner or later, the food supply and living space will run out. So the model is probably too simple.

A more complicated model assumes that the net growth rate r is not constant. While the birth rate may be constant, the death rate increases as the population increases due to competition for scarce resources. Thus, r is a decreasing function of $Y(t)$. The simplest decreasing function is linear.

New assumption: $r = k\left(1 - \dfrac{Y}{L}\right)$, where L is the maximum sustainable population (called the *carrying capacity*).

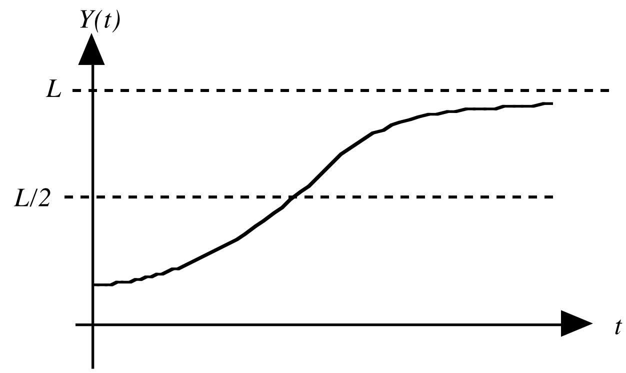
Then $\dfrac{dY}{dt} = kY\left(1 - \dfrac{Y}{L}\right)$.

This is a nonlinear equation which can be solved by separation of variables. However, we can say much about the behavior of the solution to this equation without actually solving it.

- $\dfrac{dY}{dt} = 0$ when $Y = 0$ and $Y = L$, and $\dfrac{dY}{dt} > 0$ iff $0 < Y < L$. Hence, if the initial population is between 0 and L, the population grows until it approaches $Y = L$. If the initial population is greater than L, then the population decreases until it reaches $Y = L$.

- $\frac{d^2Y}{dt^2} = k\left(1-\frac{2Y}{L}\right)\frac{dY}{dt} = k^2Y\left(1-\frac{Y}{L}\right)\left(1-\frac{2Y}{L}\right)$, so the graph is concave up for

$0 < Y < \frac{L}{2}$ and concave down for $\frac{L}{2} < Y < L$. In other words, the population grows at an increasing rate until it reaches half its carrying capacity; beyond that, the rate of increase slows down until the population stops growing. This behavior is called *logistic growth*.



We'll talk a lot about population models later in the course.

Now suppose we allow the fish to be harvested. How does this affect the population? Let h = rate at which fish are harvested. If fishing is done by casting out a big net, then it makes sense that the more fish that are in the area, the more you will catch. So:

Assumption: h is proportional to $Y(t)$.

Then $\frac{dY}{dt} = kY\left(1-\frac{Y}{L}\right) - hY = Y\left(k-h-\frac{kY}{L}\right)$. Now $\frac{dY}{dt}=0$ when $Y=0$ or $Y=(k-h)\frac{L}{k}$. If $k > h$, then solution behaves as above, but levels off at slightly lower value. If $k < h$ then $\frac{dY}{dt} < 0$ for all $Y > 0$. This means harvesting level is too great for population to survive.

1.5.2 Vehicular braking distance

A car is traveling at velocity v_0 when the driver receives a signal to apply the brakes. How far does the car travel before it comes to a complete stop? A number of factors influence the answer: mass of car, road conditions, tire conditions, and so on. We'll ignore everything except the velocity.

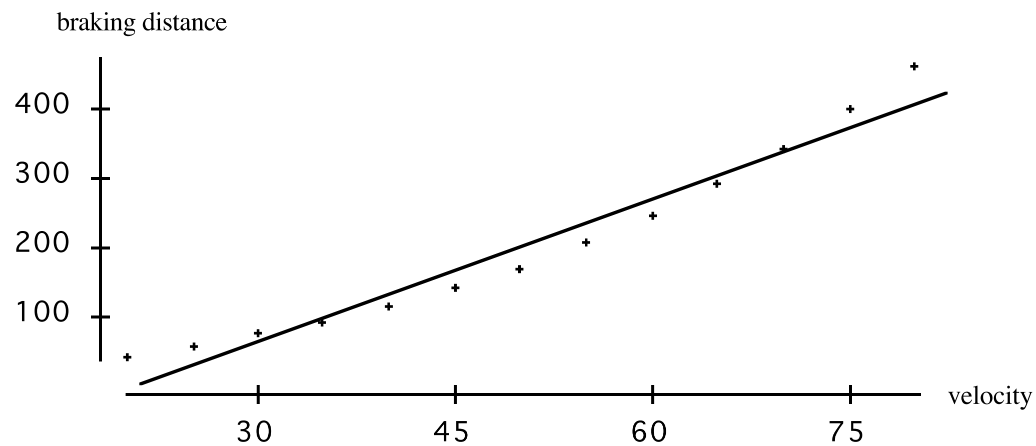
Analytical approach: The total stopping distance consists of two parts – a reaction distance and a braking distance. Assuming the reaction time is constant t_r, then the distance travelled by the car during the reaction time is proportional to the velocity; that is, $D_r = t_r v_0$. Assuming the brakes apply a constant force F, then by Newton's Law of

Motion, $-F = m\frac{dv}{dt}$, where $v(t)$ is the velocity of the car t seconds after the driver applies the brakes and m is the mass of the car. Solving this simple differential equation subject to the initial condition $v(0) = v_0$ gives $v(t) = \frac{-F}{m}t + v_0$.

Let t_b be the time required to stop the car; i.e. $v(t_b) = 0$. Then $t_b = \frac{m}{F}v_0$. The braking distance $D_b = \int_0^{t_b} v(t)dt = \frac{-F}{2m}t_b^2 + v_0 t_b = \frac{m}{2F}v_0^2$. Hence, D_b is proportional to v_0^2 So, $D = D_r + D_b = t_r v_0 + kv_0^2$.

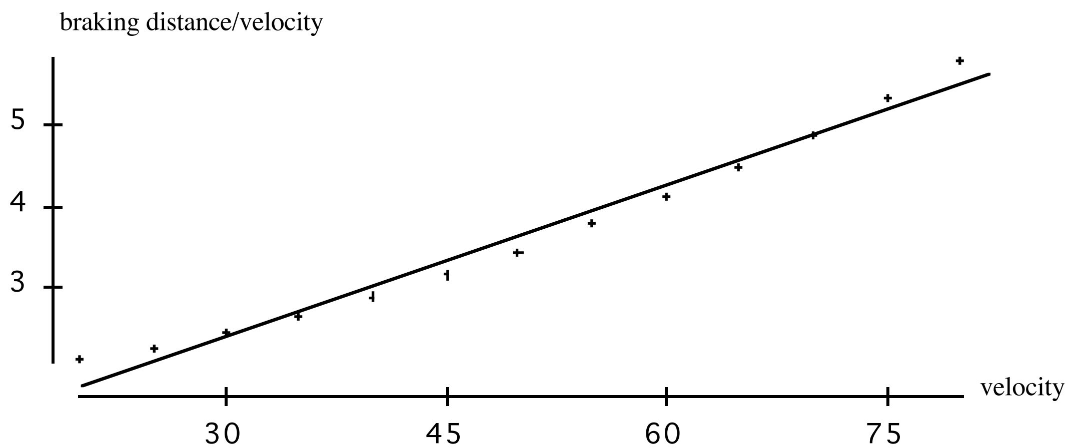
Does this agree with actual data?

vel.(mph)	20	25	30	35	40	45	50	55	60	65	70	75	80
dist.(ft.)	42	56	74	92	116	143	173	210	248	293	343	401	464



The scatterplot shows a distinctly nonlinear trend. But is it quadratic? There are several ways of answering this question.

Since $D = 0$ when $v_0 = 0$, then the quadratic function must pass through the origin. That is, $D = av_0 + bv_0^2$ (no constant term). This means $\frac{D}{v_0} = a + bv_0$, a linear function. Here's a graph of $\frac{D}{v_0}$ versus v_0. It does not appear to be linear. So our analytical model is probably not adequate.



Here's another way to confirm the inadequacy of the quadratic model. Suppose $y = f(x) = ax + bx^2$. Then $\Delta f(x) = f(x+h) - f(x) = ah + 2bhx + bh^2$. So, for equally-spaced x-values, the difference between y-values should be a linear function of x.

Here, the differences between consecutive distances are:

$$\Delta D = \{14, 18, 18, 24, 27, 30, 37, 38, 45, 50, 58, 63\}.$$

If this were a linear function, then the difference between consecutive values in this sequence should be constant. Clearly they aren't. So our quadratic model is not particularly good.

At this point, we could either try to come up with a better analytical model (perhaps by accounting for air resistance, which depends on velocity) and see if it fits the data better. Or we could try to fit a more complicated function (perhaps cubic or exponential) to the given data and see if we can come up with a plausible analytical explanation.

Unit 2: Empirical Modeling

2.1 Overview

Assume we want to model a response variable y as a function of a single explanatory variable x. We collect a sample of data $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$. The objective is to find a function $y = f(x)$ that "fits the data well."

For n data points, it is always possible to find a polynomial of degree $n - 1$ that passes exactly through all the points (provided the x-values are all distinct). The reason is that a polynomial of degree $n - 1$ has n coefficients. Substituting the given data gives a system of n equations in n variables, which will have a unique solution.

Example: Find a cubic polynomial that fits $\{(1, 2), (2, 5), (3, 9), (4, 6)\}$.

Let $y = a + bx + cx^2 + dx^3$. Substituting, we get four equations:

$2 = a + b + c + d$, $5 = a + 2b + 4c + 8d$, $9 = a + 3b + 9c + 27d$ and $6 = a + 4b + 16c + 64d$. The solution to this system is:

$$a = 8,\ b = \frac{-79}{6},\ c = \frac{17}{2},\ d = \frac{-4}{3}$$

so $y = 8 - \frac{79}{6}x + \frac{17}{2}x^2 - \frac{4}{3}x^3$ is the desired polynomial.

The problem is that if n is large, we get a very high-degree polynomial, which probably cannot be explained analytically. So we relax the requirement that the curve must pass exactly through all the points. We restate the objective as:

- Find a "reasonable" model that passes "close" to the given data points.

By "reasonable", we mean a model that is relatively simple to evaluate, has some intuitively plausible properties, and has some chance of being explained analytically.

Definition: The process of fitting curves to data is called *regression*.

What do we mean by "close"? Let $r_i = y_i - f(x_i)$ = difference between the observed y-value and the one predicted by the model. r_i is called a *residual*. There are several possible criteria we can use:

- <u>Minimize the largest absolute residual</u>. In other words, find f such that $\max(|r_i|)$ is as small as possible. This is called the *Chebyschev criterion*. It ensures that no individual point is too far from the fitted curve. Implementing this criterion usually leads to a linear (or nonlinear) optimization problem which may be hard to solve. We won't consider it further.

- <u>Minimize the sum of the absolute values of the residuals</u>. Let $S = \sum_{i=1}^{n} |r_i|$. S is a

nondifferentiable function, so calculus can't be used to minimize it. For simple functions, it is possible to solve graphically.

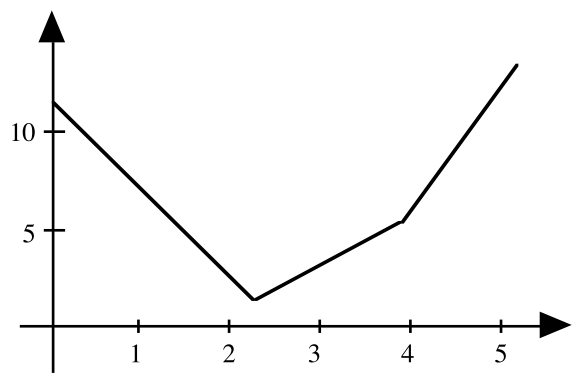
- <u>Minimize the sum of the squares of the residuals</u>. Let $Q = \sum_{i=1}^{n} r_i^2$. Q is a

differentiable function of the parameters (coefficients) of f and, in principle, can be minimized by setting the partial derivatives of Q with respect to each parameter equal to 0. This is called the *least squares criterion* and is usually the one we use.

Example: Suppose we wish to fit a line of the form $y = bx$ through the points (1, 4) and (3, 7). Determine the value of b that minimizes: (i) $S = \sum_{i=1}^{n} |r_i|$ and (ii) $Q = \sum_{i=1}^{n} r_i^2$.

The residuals are $r_1 = 4 - b$ and $r_2 = 7 - 3b$. Then:

(i) $$S = |4-b| + |7-3b| = \begin{cases} 11 - 4b, & \text{if } b < \frac{7}{3} \\ -3 + 2b, & \text{if } \frac{7}{3} < b < 4 \\ 4b - 11, & \text{if } b > 4 \end{cases}$$



The minimum occurs at $b = \frac{7}{3}$.

(ii) $Q = (4-b)^2 + (7-3b)^2 = 65 - 50b + 10b^2$

$\frac{dQ}{db} = -50 + 20b$ which equals 0 when $b = \frac{5}{2}$. This is the value that minimizes Q.

Note that the answers to (i) and (ii) are not the same.

2.2 Types of functions

The first step in empirical modeling is to decide on an appropriate form for the function. The decision is based on the desired properties of the function. Here are a list of some possibilities and their properties:

- **Linear:** $f(x) = a + bx$

  Slope is constant; therefore, $\dfrac{y_{i+1} - y_i}{x_{i+1} - x_i}$ (the slope of the line segments joining consecutive data points) ought to be roughly constant. Also $\lim\limits_{x \to \pm\infty} f(x) = \pm\infty$, meaning the function increases or decreases without bound.

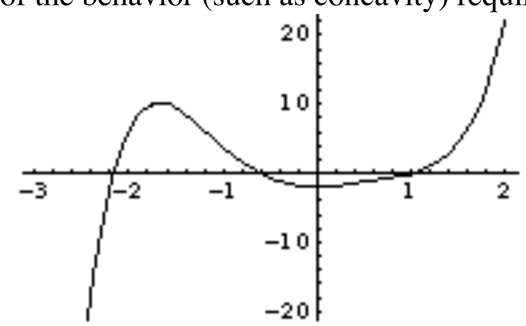
- **Quadratic:** $f(x) = a + bx + cx^2$

  Function is symmetric about $x = \dfrac{-b}{2a}$. Graph is concave up if $c > 0$ and concave down if $c < 0$. There is a local minimum (maximum) at $x = \dfrac{-b}{2a}$ if $c > 0$ ($c < 0$).

- **Polynomials:** $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$

  Graph has at most $n - 1$ turning points (local extrema). If n is odd and $a_n > 0$, then $\lim\limits_{x \to \infty} f(x) = \infty$ and $\lim\limits_{x \to -\infty} f(x) = -\infty$. If n is odd and $a_n < 0$, then $\lim\limits_{x \to \infty} f(x) = -\infty$ and $\lim\limits_{x \to -\infty} f(x) = \infty$. If n is even and $a_n > 0$, then $\lim\limits_{x \to \pm\infty} f(x) = \infty$. If n is even and $a_n < 0$, then $\lim\limits_{x \to \pm\infty} f(x) = -\infty$.

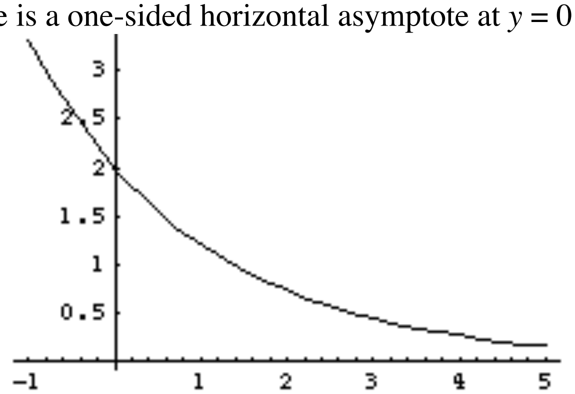
  Further analysis of the behavior (such as concavity) requires calculus.



$f(x) = x^5 - 3x^3 + 4x^2 - 2$ (2 turning points)

- **Exponential:** $f(x) = ae^{bx}, a > 0$

  If $b > 0$, the graph is increasing and concave up. If $b < 0$, the graph is decreasing and concave up. There is a one-sided horizontal asymptote at $y = 0$.

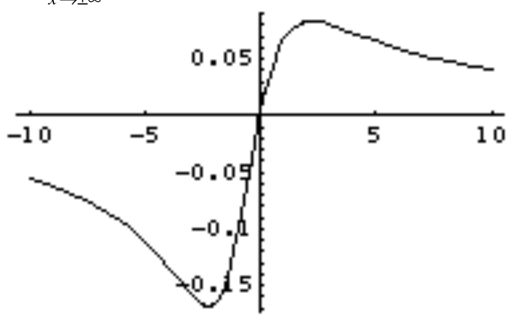


$f(x) = 2e^{-0.5x}$

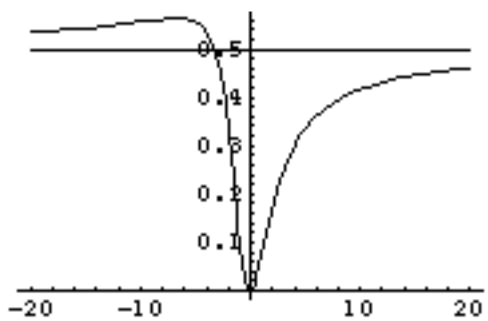
- **Rational:** $f(x) = \dfrac{g(x)}{h(x)}$, where g and h are polynomials

Has a vertical asymptote wherever $h(x) = 0$. These are unusual in the real world, so we usually just consider rational functions in which $h(x)$ is never 0 for any feasible x.

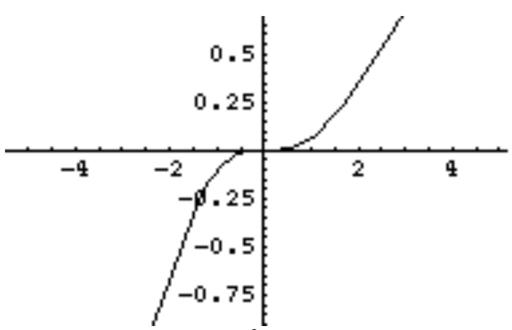
Rational functions may have horizontal asymptotes. If degree of $h(x) >$ degree of $g(x)$, then $y = 0$ is a horizontal asymptote. If the degree of $h(x)$ = degree of $g(x)$, then the horizontal asymptote is at $y = \dfrac{a_n}{b_n}$, where a_n is the leading coefficient (i.e. coefficient of highest power term) of g and b_n is the leading coefficient of h. If the degree of $h(x) <$ degree of $g(x)$, then $\lim_{x \to \pm\infty} f(x) = \pm\infty$.



$f(x) = \dfrac{x}{2x^2 + 3x + 10}$ (horizontal asymptote $y = 0$)



$f(x) = \dfrac{x^2}{2x^2 + 3x + 10}$ (horizontal asymptote $y = 1/2$)



$f(x) = \dfrac{x^3}{2x^2 + 3x + 10}$ (no horizontal asymptote)

- **Trigonometric:** $f(x) = a\sin bx$ or $f(x) = a\cos bx$
  These are useful when the behavior is periodic (repeats itself after a fixed amount of time). The period (length of one wave) is $\dfrac{2\pi}{b}$; the amplitude (maximum height) is a.

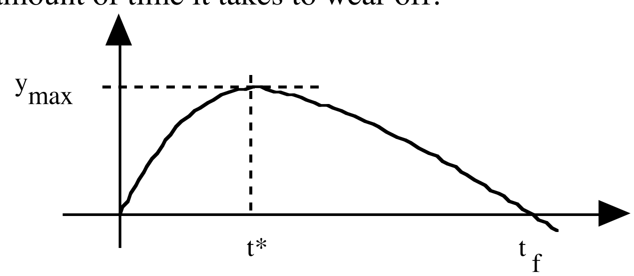
The first step in choosing an appropriate model is to draw a *scatterplot* which will give some indication of the general properties (linear vs. nonlinear, symmetry, asymptotes, etc.) of the model.

Once we choose the model and estimate the parameters, it is important to look at the residuals. A good model will yield residuals that are randomly scattered between positive and negative values. If the residuals show a distinct pattern, then the model is probably insufficient.

Example: You take a certain medication orally. Initially the concentration in your blood is 0. After a while, the concentration builds to a maximum and then gradually wears off. Let $y(t)$ = concentration (in mg/l) at time t. The maximum concentration $y_{\max}$ occurs at time t^*.

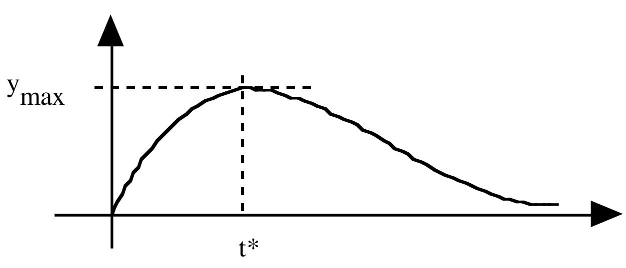
The phrase "gradually wears off" can either mean that (i) there is some finite time t_f such that $y(t_f) = 0$, or that (ii) $\lim_{t\to\infty} y(t) = 0$.

Case (i): Presumably the amount of time until maximum concentration is reached is less than the amount of time it takes to wear off.



Here, a polynomial might be an appropriate model. Since the graph is not symmetric, a quadratic polynomial is not adequate. A cubic polynomial might work.

Case (ii):



An ordinary exponential function $y(t) = ae^{-bt}$ doesn't work since that has $y(0) = a$ and is strictly decreasing. In order to make $y(0) = 0$, we could try $y(t) = ate^{-bt}$. This has a local maximum at $t = \dfrac{1}{b}$ and also satisfies $\lim_{t\to\infty} y(t) = 0$.

Other possibilities are $y(t) = at^2 e^{-bt}$ or a rational function with the degree of the numerator less than the degree of the denominator; e.g. $y(t) = \dfrac{at}{t^2 + b}$.

If we knew the values of t^* and $y_{\max}$, we could solve for some of the parameters. For example, in the model $y(t) = ate^{-bt}$, since the maximum occurs at $t = \dfrac{1}{b}$, then $b = \dfrac{1}{t^*}$. Furthermore, the maximum value is $y\left(\dfrac{1}{b}\right) = \dfrac{a}{b}e^{-1} = at^* e^{-1}$. Thus, $a = \dfrac{ey_{\max}}{t^*}$.

In practice, we are unlikely to know these values exactly, so we will have to estimate them from collected data.

2.3 Estimating parameters

2.3.1 One parameter model

The model that we choose will depend on one or more parameters. As stated earlier, we'll determine the values of those parameters in such a way that the sum of the squares of the residuals is minimized. We call this *least squares estimation*.

Consider the one-parameter model $y = bx$. The goal is to find the value of b that minimizes $Q = \sum_{i=1}^{n} \left(y_i - bx_i\right)^2$. (Note Q is a function of b. The x's and y's are given data.)

$$\frac{dQ}{db} = -2\sum_{i=1}^{n}(y_i - bx_i)x_i = -2\left[\sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2\right].$$

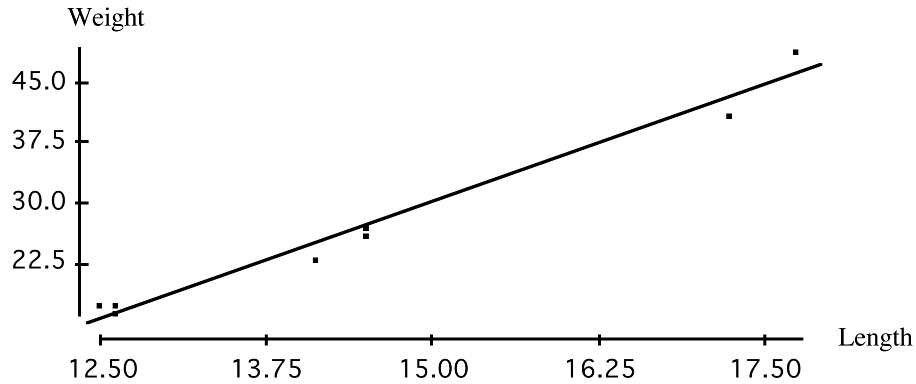
Thus, $\dfrac{dQ}{db} = 0$ when $b = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$. This is the least squares estimate.

Previously, we considered the following example: Fit a line of the form $y = bx$ through the points (1, 4) and (3, 7) that minimizes Q. Using the formula above, we get $b = \dfrac{1(4) + 3(7)}{1^2 + 3^2} = \dfrac{25}{10} = \dfrac{5}{2}$, as before.

This formula holds for any model of the form $y = bf(x)$, by replacing x_i with $f(x_i)$ throughout.

Example: The length x (in inches) and weight w (in ounces) of a sample of bass fish are given below.

x	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
w	27	17	41	26	17	49	23	16



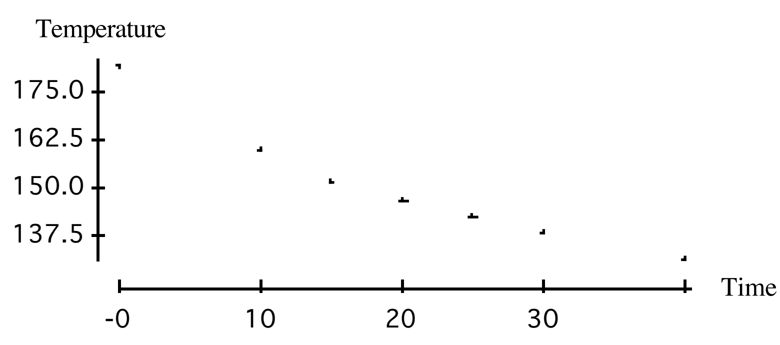
The scatterplot shows that the relationship is nonlinear, so a simple model $w = bx$ is not adequate. A better choice is $w = bx^3$, which makes sense since weight should be proportional to volume which is measured in cubic inches. The constant b may be interpreted as the density (in ounces/cu. in.) of the fish.

The least squares estimate of b is $\dfrac{\sum_{i=1}^{8} x_i^3 w_i}{\sum_{i=1}^{8} x_i^6} = \dfrac{810481}{9.61\times 10^7} \approx .0084$. Hence, $w = .0084x^3$.

The corresponding residuals are $\{1.279, 0.521, -2.305, 0.279, 0.022, 1.818, -0.776, -0.977\}$, which are sufficiently random to indicate a good model. The model predicts fish weights which are within 5% of the observed values.

Example: A turkey is heated to an internal temperature of approximately 180° and allowed to cool in a room whose temperature is 70°. The temperature is recorded at periodic time intervals.

Time (t)	0	10	15	20	25	30	40
Temperature (z)	182	160	152	147	143	139	131



Since the graph of z vs. t should have a horizontal asymptote at $z = 70$, an exponential model might be appropriate. (You might also try a rational model.) Let's try $z = a + be^{-kt}$. Since $z(0) = 182 = a + b$ and $\lim_{t\to\infty} z(t) = 70 = a$, then $b = 112$. Hence,

$$z = 70 + 112e^{-kt}.$$

This is not a linear model as written, but can be transformed into one.

$$e^{-kt} = \frac{z-70}{112} \qquad \text{so} \quad -kt = \ln\left(\frac{z-70}{112}\right).$$

Now we can use the results for one-parameter models with $x_i = t_i$ and $y_i = \ln\left(\frac{z_i - 70}{112}\right)$.

Thus, $-k = \dfrac{\sum_{i=1}^{7} t_i \ln\left(\frac{z_i-70}{112}\right)}{\sum_{i=1}^{7} t_i^2} = -.0166$, so $k = .0166$.

The model is $z = 70 + 112e^{-.0166t}$.

The corresponding residuals are $\{0, -4.872, -5.317, -3.364, 0.965, 0.924, 3.334\}$, also on the order of 3 – 5% of the observed values. Note, however, that the first 3 residuals (after the 0 which results from the way in which we determined the coefficients) are negative, while the remaining ones are positive. This is a sign that the model may need some tweaking.

There is an analytical justification for the model used here. *Newton's Law of Cooling* says that the rate of change of temperature is proportional to the difference between the current temperature and the ambient (surrounding) temperature. Thus:

$$\frac{dz}{dt} = -k(z-70).$$

The solution to this equation, subject to the initial condition $z(0) = 182$ is $z = 70 + 112e^{-kt}$.

There is some evidence that Newton's Law of Cooling isn't quite right; perhaps $\frac{dz}{dt} = -k(z-70)^{\alpha}$, where $\alpha > 1$ is another parameter. The solution to this equation is a rational function, which might give us a better fit. (We have to be a bit careful. This model may only apply if the initial temperature is greater than 70 since $(z-70)^{\alpha}$ may not be defined for $z < 70$ for some α, such as $\alpha = 1.5$. Or for $\alpha = 2$, an initial temperature less than 70 would have $\frac{dz}{dt} < 0$, which makes no sense. So we might need one equation for temperatures greater than the ambient temperature and another equation, with the sign changed, for temperatures less than the ambient temperature.)

2.3.2 Two-parameter model

Now consider the two-parameter linear model $y = a + bx$. The sum of the squares of the residuals is $Q = \sum_{i=1}^{n}(y_i - a - bx_i)^2$, a function of a and b. To minimize, set the partial derivatives equal to 0.

(i) $\dfrac{\partial Q}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = -2\left[\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i\right]$

(ii) $\dfrac{\partial Q}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = -2\left[\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2\right]$

Setting (i) = 0, we get: $\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$.

From (ii), we have: $\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$.

The solution to this system of equations is given by:

Theorem 2.1: The least squares estimates of a and b in the model $y = a + bx$ are

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},\ a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}.$$

If we let $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$ be the average (or *sample mean*) x- and y-values,

then the least squares estimates can be rewritten as $b = \dfrac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$, $a = \bar{y} - b\bar{x}$.

Note that this implies the regression equation always passes through the point $(\bar{x}, \bar{y})$. Any line that passes through $(\bar{x}, \bar{y})$ has the sum of the residuals equal to 0. The line through $(\bar{x}, \bar{y})$ with slope given above has the minimum sum of squares of residuals.

As in the one-parameter case, these results can be used for any model of the form $g(y) = a + bf(x)$. For example:

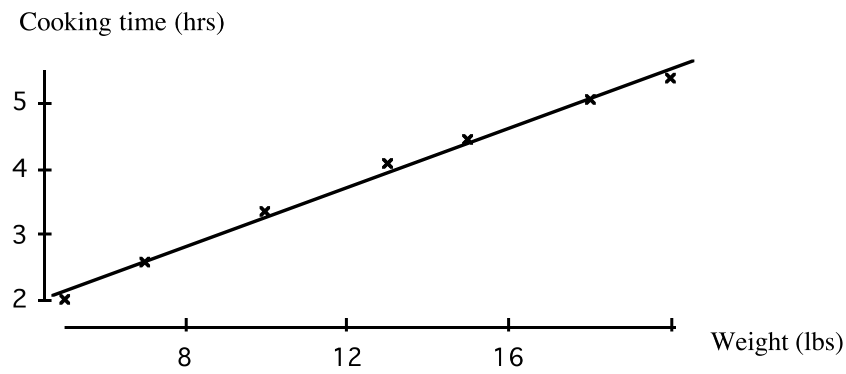
Power model: $y = ax^b \Rightarrow \ln y = \ln a + b\ln x$

Exponential model: $y = ae^{bx} \Rightarrow \ln y = \ln a + bx$

Rational model: $y = \dfrac{x}{ax+b} \Rightarrow \dfrac{1}{y} = a + \dfrac{b}{x}$.

Example: A turkey is considered to be "done" when its internal temperature is approximately 180 degrees Fahrenheit. Clearly, the bigger the turkey, the longer it needs to be cooked until it reaches this temperature. The data below gives the weight (in pounds) and cooking time (in hours) for seven turkeys.

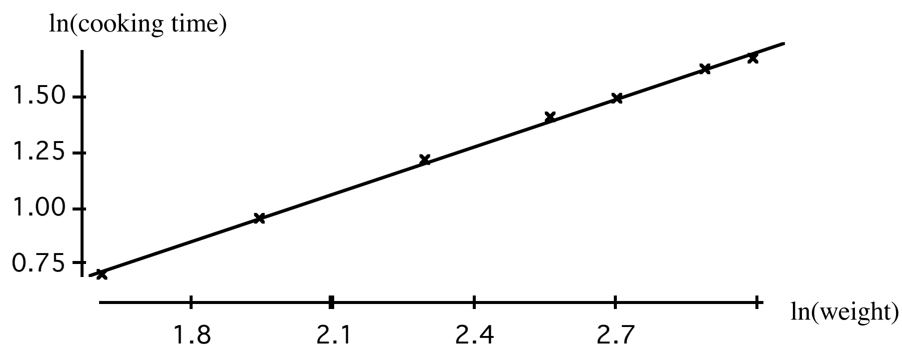
Weight (lbs.)	5	7	10	13	15	18	20
Cook time (hrs.)	2	2.6	3.4	4.1	4.5	5.1	5.4



The graph is non-linear, increasing and concave down. Since we have no reason to believe the graph has a horizontal asymptote, we might try a power function model of the form $t = aw^b$, where $a > 0$ and $0 < b < 1$.

In order to use least squares, we first linearize the model by taking the logarithm:

$$\ln t = \ln a + b \ln w$$



Now use Theorem 2.1 with $y_i = \ln t_i$ and $x_i = \ln w_i$. Upon doing so, we get $b = .719$ and $\ln a = -.448$, from which $a = .638$.

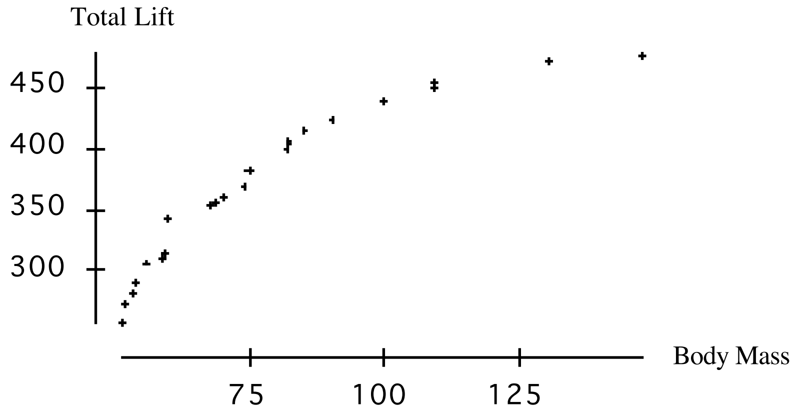
Hence, $t = .638w^{.719}$.

Unfortunately, this is neither explainable nor usable. Exponents of .719 do not often occur in nature. But, it does give us a clue. Cooking time is likely to depend on the thickness of the turkey (rather than its weight), which is proportional to $V^{2/3}$, where V is the volume. If we assume constant turkey density, then V is proportional to w. Thus, t is proportional to $w^{2/3}$. (Note that 2/3 is not far from .719). Now we can fit a one-parameter model $t = bw^{2/3}$. The least squares estimate of b is .734. Hence, $t = .734w^{2/3}$.

Using this model, the predicted cooking times for the turkeys in the data set are {2.14, 2.69, 3.40, 4.05, 4.46, 5.04, 5.40}. The corresponding residuals are {–.14, –.09, 0, .05, .04, .06, 0}, all quite small.

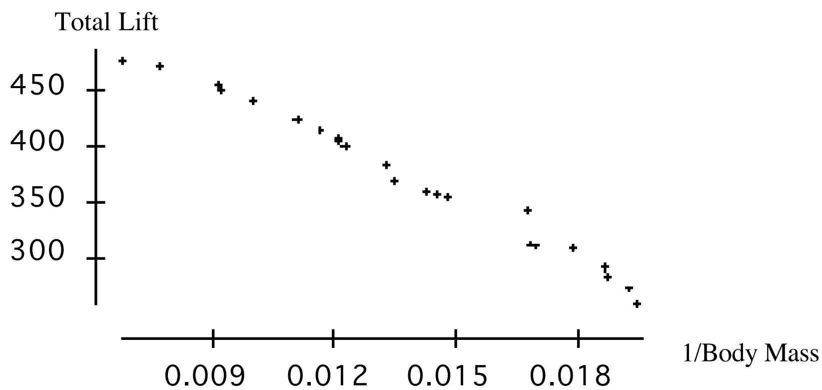
Many cookbooks recommend cooking turkey 20 minutes per pound (a one-parameter linear model). This gives cooking times of {1.67, 2.33, 3.33, 4.33, 5, 6, 6.67}. This is not bad for moderate size turkeys, but large turkeys will be overcooked by this rule.

Example: Here is a scatterplot of the absolute world records in weightlifting as a function of the body mass of the lifter. (An absolute record means that no lifter of smaller body mass has ever lifted more than the record weight.) The data recorded is the sum of the weights lifted in two events – the *snatch* (where the lifter must lift the weight overhead in one motion) and the *jerk* (a two-part motion in which the lifter rests the weight on his chest and then lifts it overhead). Both weights and body mass are measured in kilograms.



The y-values appear to level off at approximately 500 kg, so any model should have a horizontal asymptote at or near 500.

We'll try the rational model $y = a + \frac{b}{x}$, where x = body mass and y = weight lifted. The graph of y vs. $1/x$ appears to be approximately linear.



The least-squares equation is $y = 609.9 - \frac{17315.4}{x}$. This levels off at 609.9 kg, which might be a little high. Perhaps a slight change in the exponent of x will give a more realistic value.

It is possible to extend the results of this section to models with more than 2 parameters such as $y = a + bx + cx^2$ or $z = a + bx + cy$. The calculations of the least-squares estimates in these cases are rather messy and best left to a computer.

Problems

1. (a) Given two points (x_1, y_1) and (x_2, y_2), there is a unique straight line that passes through them. Show that an equation of this line can be written in the form

$$y = p_1(x) = \frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2 .$$

[Hint: Just show that the two given points are on the line.]

(b) Rewrite the equation in (a) in the form $y = mx + b$ and show that it gives the correct slope and y-intercept.

(c) Given three points, we claimed in class that there was a unique parabola (second-degree polynomial) that passes through them exactly. Assume that the equation of this parabola can be written in the form $p_2(x) = g_1(x)y_1 + g_2(x)y_2 + g_3(x)y_3$. Determine the quadratic functions $g_1(x)$, $g_2(x)$ and $g_3(x)$.

(d) Use your result to (c) to find a second-degree polynomial that passes through the points (1, 4), (3, 8) and (6, –2).

(e) Extend the results to third-degree polynomials and show that it gives the correct results for the data in the Example on page 9 of these notes.

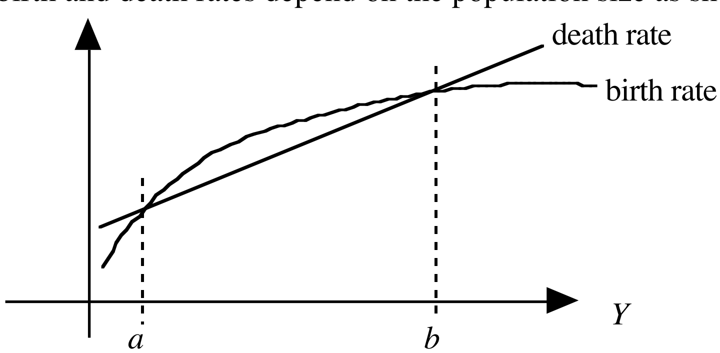
2. A new brand of widget is introduced on the market. At first, the market share for this brand (i.e. percentage of widget buyers who buy this brand) is 0 and increases very gradually. Then, after an advertising campaign, the popularity of the brand increases more rapidly. When the campaign is over, the widget market settles down with the new brand representing about 30 percent of widget sales.

(a) Draw a graph of market share vs. time.

(b) What function might be used to describe this behavior? Give a specific functional form which may depend on one or more parameters. Describe how those parameters affect the behavior of the function.

[Hint: You might try thinking rationally.]

3. A population grows according to the model $\frac{dY}{dt} = rY$, where r = birth rate – death rate. Suppose the birth and death rates depend on the population size as shown below.



(a) Sketch a graph of r vs. Y, where $r = b - d$.

(b) Sketch a graph of $\frac{dY}{dt}$ vs. Y.

(c) Sketch a graph of Y vs. t. You should have several different graphs, depending on the initial population size, Y_0.

4. In Section 2.3.1, we predicted the weight of a bass as a function of its length. We determined that the model $w = bx^3$ was appropriate. An alternative model takes into account the girth (circumference around the widest part) of the fish. The data below contains the weight, length and girth of 8 bass. Use least squares to fit the model $w = bx^2y$, where b is a constant. Compute the predicted values and residuals.

Length (x)	14.5	12.5	17.25	14.5	12.625	17.75	14.125	12.625
Girth (y)	9.75	8.375	11.0	9.75	8.5	12.5	9.0	8.5
Weight (w)	27	17	41	26	17	49	23	16

5. Suppose we want to fit the model $y = e^{bx}$ to the data $\{(0,1), (1,4), (2,8)\}$.

(a) Find the least-squares estimate of b by taking the logarithm of the model and using the results of Section 2.3.1.

(b) Let $Q = \sum_{i=1}^{3} \left(y_i - e^{bx_i}\right)^2$. Substitute the given data and express Q in terms of b alone. Then find the value of b that minimizes Q. [Note: You will have to use numerical or graphical techniques to approximate the answer.]

6. Suppose we want to fit the model $y = bx$ to the data $\{(1,4), (5,10)\}$. For this problem, we'll use the Chebyschev criterion of minimizing the largest absolute residual. Let $r_1 = |4 - b|$, $r_2 = |10 - 5b|$ be the absolute residuals for the two data points.

(a) Plot r_1 and r_2 as functions of b on the same axes.

(b) Plot $M = \max\{r_1, r_2\}$ as a function of b.

(c) Determine the value of b that minimizes M.