This document provides some suggestions for possible final projects for MA 137. You can select any other project of your choice provided it has both a substantial mathematical component (related to what we learned in MA 137) and an adequate biological and/or medical interest. You can freely draw your projects from the World Wide Web. The following website maintained by the author of our textbook is also a valuable source of information:

`http://bioquest.org/numberscount/`

Here are some guidelines with which your paper must comply:

- You have to turn in a typewritten paper, at least four (4) pages long. It should be double spaced, it should use 12pt fonts (Times New Roman, Helvetica, or Arial);

- Your paper should not be in an itemized form but it should be written in a narrative/expository form. You can use sections and subsections. There must be an abstract, an introductory preamble and a final conclusion;

- You must quote, at the end of your paper, all the references that you used for your work;

- Your paper must provide historical and contextual facts and/or preliminary background material on the topic of your paper;

- Your paper must contain all the steps of your calculations (say, derivatives and/or limits). You must also indicate which properties you are using while doing your calculations (say, product rule, chain rule, etc...);

- Your paper must contain illustrative graphs and tables (Maple or Mathematica outputs, Excel spreadsheets, etc...);

- Your paper can be written by yourself or by a group of at most two students.

- Your project paper will be worth at most 20 points.

Please seek help from your instructor if you are uncertain on what to do. You can also seek a preliminary opinion from your instructor to determine whether you are doing satisfactory work. DO NOT WAIT until the last minute to complete your project and/or seek help.

Project Ideas/Suggestions

1. **(U.S. Population Growth Analysis)**
   a) Describe the properties of the discrete logistic equation and the stability of its equilibrium points.

   b) The following table provides population data for the United States, 1790-2000, measured in millions.

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890
Pop.	3.93	5.31	7.24	9.64	12.87	17.07	23.19	31.44	39.82	50.16	62.95

Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Pop.	75.99	91.97	105.71	122.78	131.67	151.33	179.32	203.21	226.50	249.63	281.42

   Which values of the parameters N_0, R and K of the discrete logistic equation give a reasonable fit of the U.S. population data given above? Make a graph with both the actual data and the discrete logistic model fit.

2. **(State/county population Analysis)** Do a similar analysis for the state of Kentucky (or your home state) or your home county. Indicate where you obtained your population data and how reliable the data is, or might be. The population data should go back at least 150 years.

3. **(Model of selection and mutation)** Bacterial growth is usually exponential. Let the population of a bacterium at time t be denoted by b_t. Suppose that a mutant wild type with population m_t appears and begins competing. Assume the original type (or wild type) satisfies the difference equation $b_{t+1} = rb_t$ and the mutant type satisfies the difference equation $m_{t+1} = sm_t$.

   If $s > r$ then the mutant type will grow faster than the wild type, while if $r > s$ then the wild type will outperform the mutants. The establishment of this mutant is an example of selection. Selection occurs when the frequency of a genetic type changes over time.

   If we observe this population over time, counting all of the bacteria each hour would be impossible. Nonetheless, we could sample the population and measure the fraction of the mutant type by counting or using a specific stain. If the fraction grows larger and larger, then we know that the mutant type was taking over.

   We are interested in the growth of this fraction of the population. Let p_t be the fraction of mutants at time t. What is this fraction?

   $$p_t = \frac{\text{number of mutants}}{\text{total number}}$$
   $$= \frac{\text{number of mutants}}{\text{number of mutants} + \text{number of wild type}}$$
   $$= \frac{m_t}{m_t + b_t}$$

2

The fraction of the wild type is

$$\text{fraction of wild type} = \frac{\text{number of wild type}}{\text{total number}}$$

$$= \frac{\text{number of wild type}}{\text{number of mutants} + \text{number of wild type}}$$

$$= \frac{b_t}{m_t + b_t}$$

Now, note that these two fractions must add to 1.

$$\frac{b_t}{m_t + b_t} + \frac{m_t}{m_t + b_t} = \frac{m_t + b_t}{m_t + b_t} = 1.$$

Therefore, the fraction of wild type is $1 - p_t$.

Now, what we are after is the fraction of the mutants in the next generation, p_{t+1}. This should be easy.

$$p_{t+1} = \frac{m_{t+1}}{m_{t+1} + b_{t+1}}.$$

Using the growth rate for these two strains we have

$$p_{t+1} = \frac{sm_t}{sm_t + rb_t}.$$

This would be fine, except that we would have to have exact values for m_t and b_t, and these were not possible to count. Instead we would like to have p_{t+1} in terms of p_t. How do we do this?

Divide numerator and denominator by $b_t + m_t$. This gives

$$p_{t+1} = \frac{s\dfrac{m_t}{m_t + b_t}}{s\dfrac{m_t}{m_t + b_t} + r\dfrac{b_t}{m_t + b_t}}.$$

Recalling that $1 - p_t = \dfrac{b_t}{m_t + b_t}$ and $p_t = \dfrac{m_t}{m_t + b_t}$, we have

$$p_{t+1} = \frac{sp_t}{sp_t + r(1 - p_t)}.$$

a) Show that the three equilibrium points for this difference equation are $s = r$, $p^* = 0$ or $p^* = 1$. Interpret these equilibria as they pertain to the model.

b) Using the stability criterion, what is the condition on r and s that makes $p^* = 0$ a locally stable solution?
   Illustrate this situation using cobwebbing for a couple of values of r and s and starting values for the two populations.

c) Using the stability criterion, what is the condition on r and s that makes $p^* = 1$ a locally stable solution?
   Illustrate this situation using cobwebbing for a couple of values of r and s and starting values for the two populations.

4. (Hematopoietic cell replication and control) Periodic hematopoiesis is a disease characterized by large oscillations in the red blood cell count. Red blood cells are generated by a feedback mechanism that approximately obeys the difference equation

$$x_{t+1} = \frac{\tau}{1 + \gamma\tau} F(x_t) + \frac{1}{1 + \gamma\tau} x_t.$$

The function $F(x_t)$ describes production as a function of number of cells and takes the form

$$F(x) = F_0 \frac{\theta^n}{\theta^n + x^n}.$$

The terms here are:

Parameter	Meaning	Normal Value
x_t	Number of cells at time t (in billions)	About 330
τ	Time for cell development	5.7 days
γ	Fraction of cells that die each day	0.0231
F_0	Maximum production of cells	76.2 billion
θ	Value where cell production is halved	247 billion
n	Shape parameter	7.6

a) Graph and explain $F(x)$ with these parameter values. Does this sort of feedback system make sense?

b) Use a computer to find the equilibrium points with the normal parameters.

c) Graph and cobweb the difference equation. If you start near the equilibrium point, do values remain nearby? What might a solution look like?

d) Certain autoimmune diseases increase γ, the death rate of cells. Study what happens to the equilibrium and the solution as γ increases. Can you explain your results in biological terms?

e) Explore what would happen if the value of n were decreased. Does the equilibrium become more sensitive to small changes in γ?

5. (Bees collecting pollen and nectar—Optimization)
Many bees collect both pollen and nectar. Pollen is used for protein and nectar is used for energy. Suppose the amount of nectar harvested during t seconds on a flower is

$$F(t) = \frac{t}{1 + t}$$

and that the amount of pollen harvested during t seconds on a flower is

$$G(t) = \frac{t}{2 + t}.$$

The bee collects pollen and nectar simultaneously. Travel time between flowers is $\tau = 1.0$ seconds.

a) What is the optimal time for the bee to leave one flower for the next in order to collect nectar at the maximum rate?

b) What is the optimal time for the bee to leave one flower for the next in order to collect pollen at the maximum rate? Why are the two times different?

c) Suppose that the bee values pollen twice as much as nectar. Find a single function $V(t)$ that give the value of resources collected by time t. What is the optimal time for the bee to leave? [You may need to use a computer or calculator to solve this equation.]

d) Suppose that the bee first collects nectar and then switches to pollen. Assume it spends 1.0 second collecting nectar. How long should it spend on pollen?

e) Suppose that the bee values pollen twice as much as nectar. Experiment to try to find a solution that is best when nectar and pollen are harvested sequentially.

6. **(Small birds flapping and gliding—Optimization)** Some small song birds intermittently flap their wings and glide with wings folded between flapping sessions. Why? R. M. Alexander (*Optima for Animals*, Princeton University Press) suggests the following analysis:
   The power required to propel an airplane at a speed u is

$$P = Au^3 + \frac{BL^2}{u}$$

where A and B are constants specific to the airplane and L is the upward force that lifts the plane. Au^3 represents the drag on the airplane due largely to the air striking the front of the plane.

a) For what speed is the required power the smallest?

b) The energy required to propel the airplane is $E = P/u$. For what speed is the energy required to propel the plane the smallest?

c) How does the speed of minimum energy compare with the speed of minimum power?

d) Suppose a bird has drag coefficient A_b with wings folded and $A_b + A_w$ with wings extended and flapping, and let x be the fraction of time the bird spends flapping its wings. Suppose that the speed of the bird while flapping its wings is the same as the speed when the wings are folded and that all the lift is provide when the wings are flapping. The lift over one complete cycle should be

$$(1 - x)L_{\text{folded}} + xL_{\text{flapping}} = xL_{\text{flapping}} = mg$$

where m is the mass of the bird. Then the power required while flapping is

$$P_{\text{flapping}} = (A_b + A_w)u^3 + B\left(\frac{mg}{x}\right)^2 \frac{1}{u}.$$

Write an expression for P_{folded}.

e) The average power over a whole cycle should be

$$\bar{P} = (1 - x)P_{\text{folded}} + xP_{\text{flapping}} = A_b u^3 + xA_w u^3 + B\frac{m^2 g^2}{xu}$$

Find the value of x for which the average power over the whole cycle is a minimum.

5

f) The average energy over a whole cycle is $\bar{E} = \dfrac{\bar{P}}{u}$. For what value of x is the average energy a minimum?

Alexander further notes that it may be necessary also to consider the efficiency of muscle contraction at different flapping rates.

7. **(Harvesting of whales)** The sei whale (pronounced 'say') is a well-studied example of over-exploitation of a natural resource. Because they were fast swimmers, not often found near shore, and usually sank when killed, they were not hunted until modern methods of hunting and processing at sea were developed. Reasonably accurate records of sei whale harvests have been kept. "sei whale catches increased rapidly in the late 1950s and early 1960s (Mizroch et al., 1984c). The catch peaked in 1964 at over 20,000 sei whales, but by 1976 this number dropped to below 2,000 and the species received IWC protection in 1977."
(`http://spo.nwr.noaa.gov/mfr611/mfr6116.pdf`)
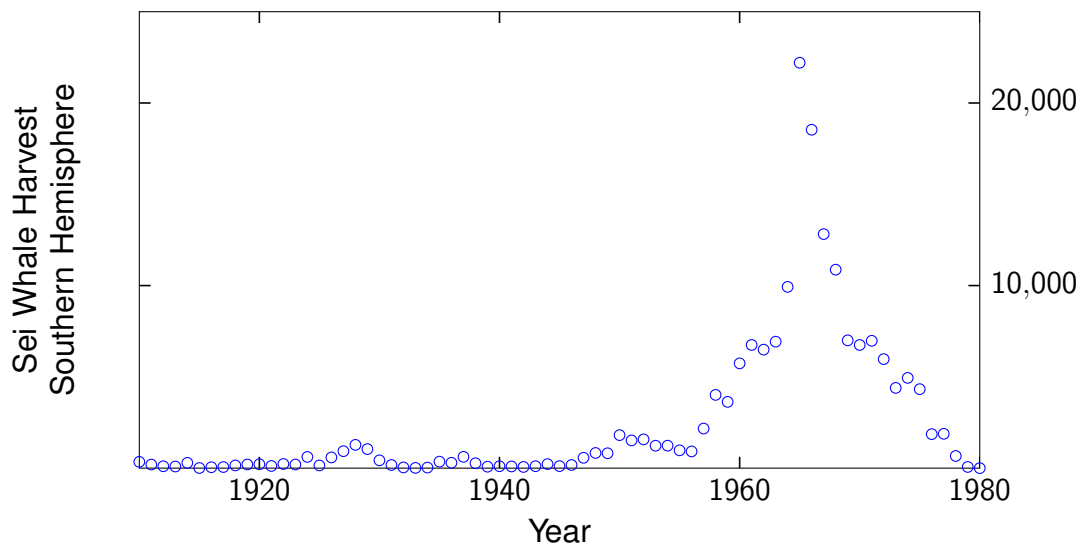
Data shown in the figure above are from Joseph Horwood, *The Sei Whale: Population Biology, Ecology & Management*, Croom Helm, London, 1987.

The International Convention for Regulating Whaling was convened in 1946 and gradually became a force so that by 1963/64 effective limits on catches of blue and fin whales were in place because of depletion of those populations. As is apparent from the data, whalers turned to the sei whale in 1964-65, catching 22,205 southern hemisphere sei whales. Catches declined in the years 1965-1979 despite continued effort to harvest them, indicating depletion of the southern hemisphere stock. A moratorium on sei whale harvest was established in 1979. The Maximum Sustainable Yield (MSY) is an important concept in population biology, being an amount that can be harvested without eliminating the population.

a) Would you estimate that the southern hemisphere sei whale would sustain a harvest of 6000 whales per year?
   The actual size of a whale population is difficult to measure. A standard technique is the *Mark and Recapture Method*, in which a number, N, of the whales is marked at a certain time, and during a subsequent time interval the number m of marked whales among a total of M whales sited is recorded.

b) Suppose 100 fish in a lake are caught, marked, and returned to the lake. Suppose that ten days later 100 fish are caught, among which 5 were marked. How many fish would you estimate were in the lake?

Another method to estimate population size is the *catch per unit effort method*, based on the number of whales caught per day of hunting. As the population decreases, the catch per day of hunting decreases.

The table and graph in below illustrate estimates of southern hemisphere sei whale population based on the mark and recapture method, as shown in Table 13.11 of Horwood. Current estimates of that population are about 50,000.

Southern Hemisphere	
Years	Total Whale Population
1963-65	393,000
1965-67	428,000
1967-69	237,000
1969-71	204,000
1971-73	105,000
1973-75	96,000
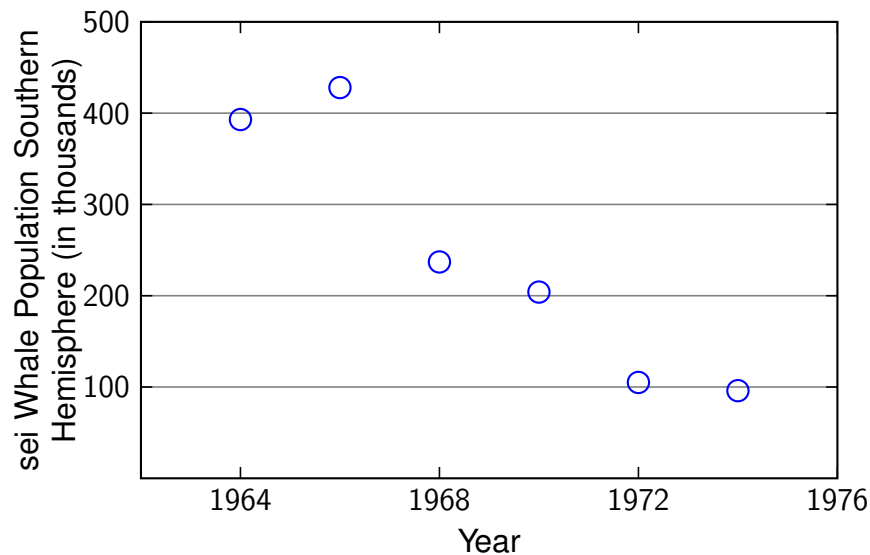
Estimates of southern hemisphere sei whale population
based on mark and recapture data.

In the Report of the International Whaling Commission (1978), J. R. Beddington refers to the following model of Sei whale populations.

$$N_{t+1} = 0.94N_t + N_{t-8} \left[0.06 + 0.0567 \left(1 - \left(\frac{N_{t-8}}{N_*} \right)^{2.39} \right) \right] - 0.94C_t. \tag{1}$$

N_t, N_{t+1} and N_{t-8} represent the adult female whale population subjected to whale harvesting in years t, $t + 1$, and $t - 8$, respectively. C_t is the number of female whales harvested in year t. There is an assumption that whales reach sexual maturity and are able to reproduce at eight years of age and become subject to harvesting the same year that they reach sexual

maturity. The whales of age less than 8 years are not included in N_t. N_* is the number of female whales that the environment would support with no harvesting taking place.

c) Show that if there is no harvest ($C_t = 0$) and both N_t and N_{t-8} are equal to N_* then $N_{t+1} = N_*$. If we divide all terms of Equation 1 by N_*, we get

$$\frac{N_{t+1}}{N_*} = 0.94\frac{N_t}{N_*} + \frac{N_{t-8}}{N_*}\left[0.06 + 0.0567\left(1 - \left(\frac{N_{t-8}}{N_*}\right)^{2.39}\right)\right] - 0.94\frac{C_t}{N_*}.$$

We define new variables, $D_t = \dfrac{N_t}{N_*}$ and $E_t = \dfrac{C_t}{N_*}$ and have the equation

$$D_{t+1} = 0.94D_t + D_{t-8}\left[0.06 + 0.0567\left(1 - D_{t-8}^{2.39}\right)\right] - 0.94E_t \tag{2}$$

This equation is simpler by one parameter (N_*) than Equation 1 and yet illustrates the same properties. Rather than use new variables, it is customary to simply rewrite Equation 1 with new interpretations of N_t and C_t and obtain

$$N_{t+1} = 0.94N_t + N_{t-8}\left[0.06 + 0.0567\left(1 - N_{t-8}^{2.39}\right)\right] - 0.94C_t \tag{3}$$

N_t now is a fraction of N_*, the number supported without harvest, and C_t is a fraction of N_* that is harvested.

Our original equation tells us that the population in year $t+1$ is affected by three things: the number of female whales in the previous year (N_t), the *recruitment* of eight year old female whales into the population subject to harvest, and the harvest during the previous year (C_t).

d) Consider the following
   i) Suppose $N_* = 250,000$ and a quota of 500 female whales harvested each year is established. Change both Equations 1 and 2 to reflect these parameter values.
   ii) Suppose 2% of the adult female whale population is harvested each year. Change both Equations 1 and 2 to reflect this procedure.
   iii) What is the meaning of 0.94 at the two places it enters Equation 3?
   The term
   $$N_{t-8}\left[0.06 + 0.0567\left(1 - N_{t-8}^{2.39}\right)\right]$$
   represents the number of eight year old females recruited into the adult population in year, t. The factor,
   $$\left[0.06 + 0.0567\left(1 - N_{t-8}^{2.39}\right)\right]$$
   represents the fecundity of the females in year $t - 8$. It has been observed that as whale numbers decrease, the fecundity increases. This term was empirically determined by Beddington.

e) Draw the graph of fecundity versus N_{t-8}.

f) Suppose the harvest level, C_t is set at a constant level, C, for a number of years. Then the whale population should reach an equilibrium level, Ne, and approximately
   $$N_{t+1} = N_t = N_{t-8} = N_e.$$
   We would like to have N_e as a function of the harvest level, C. This is a little messy (actually quite a bit messy), but it is rather easy to write the inverse function, C as a function of N_e.
   i) Let $C_t = C$ and $N_{t+1} = N_t = N_{t-8} = N_e$ in Equation 3 and solve the resulting equation for C. Draw the graph of C versus N_e.

8

    ii) Find a point on the graph of $C(N_e)$ at which the tangent to the graph is horizontal. You may wish to graph $C(N_e)$ on your calculator and use its capabilities, or use a computer program.

    iii) If you did not do it in the previous step, compute $C'(N_e) = \dfrac{dC}{dN_e}$ and find a value of N_e for which $C'(N_e)$ is zero.

    iv) Give an interpretation of the point $(N_e, C(N_e)) = (0.60001, 0.025507)$. What do you suppose happens if the constant harvest level is set to $C = 0.03$?

    v) N_* for the southern hemisphere sei whale has been estimated to be 250,000. If N_* is 250,000, what is the maximum harvest that will lead to equilibrium? You should find that the maximum supportable yield of the southern hemisphere sei whale is about 6,000 whales per year.

g) In this problem, you will gain some experience with the solution to Equation 3 for several values of the parameters by computing the solutions on your calculator. You should record the behavior of the solutions. If you want to use a computer algebra system, the logic of the program is the same, but the particular commands will depend on the program. This is a program for a TI-84.
On your calculator, press PRGM EDIT NAME = (TYPE IN SEI) ENTER. Then type in the following program.

```
:PROGRAM:SEI
:0.0255 -> C              | -> is the STO-> key
:0.61 -> S                | S is the initial population size.
:9 -> dim N               | Store population sizes N_(t-8),
:For(K,1,9):S->N(K):End   | N(t-7),...,Nt in N(1),N(2),...,N(9)
:For(I,1,30)              |
   :For(J,1,5)            | Display only every fifth iterate.
      :NEW = 0.94*N(9)+N(1)*(0.06+0.0567*(1-N(1)^2.39))-0.94*C
      For(K,1,8):N(K+1)->N(K):End
      :NEW -> N(9)
   :END
   :Disp NEW
:END
```

    i) Press EXIT. Type SEI. Press ENTER repeatedly. What happens to the whale population?

    ii) Change $0.61 \to S$ to $0.59 \to S$, and run again. What happens to the whale population?

    iii) Change $0.0255 \to C$ to $0.023 \to C$ and run again. What happens to the whale population?

    iv) Find the values of N_e that correspond to $C = 0.023$
. Suggestion: a. Iterate $x_0 = 0.5$, $x_{n+1} = 0.3813 + (x^n)^{3.39}$
                b. Iterate $x_0 = 0.5$, $x_{n+1} = (x_n - 0.3813)^{1/3.39}$.

    v) Use $0.0230 \to C$ and $0.43 \to S$ and run again. What happens to the whale population?

    vi) Change $0.43 \to S$ to $0.45 \to S$ and run again. What happens to the whale population?

    vii) Change $0.45 \to S$ to $0.78 \to S$ and run again. What happens to the whale population?

8. **(Invasion of the Toads)** This comes from the Boston University *Differential Equations Project*. As is often the case in mathematical modeling, there is no "best answer" for this question. There

are, however, "better answers": the correctness of a mathematical model is determined by how well it agrees with the reality of experimental measurements.

From 1935-37, the American marine toad (*Bufo marinus*) was introduced into Queensland, Australia in eight coastal sugar cane districts. Due to lack of natural predators and an abundant food supply, the population grew and the poisonous toads began to be found far from the region in which they were originally introduced. Survey data presented by J. Covacevich and M. Archer ("The distribution of the cane toad, *Bufo marinus*, in Australia and its effects on indigenous vertebrates," *Mem. Queensland Mus*, **17**: 305-310) shows how the toads expanded their territorial bounds within a forty-year period. This data is reproduced in the table below; it was mathematically analyzed by M. Sabath, W. Boughton, and S. Easteal ("Cumulative Geographical Range of *Bufo marinus* in Queensland, Australia from 1935 to 1974", *Copeia*, no. 3, 1981, pp. 676-680).

Year	Area Occupied (square km)
1939	32,800
1944	55,800
1949	73,600
1954	138,000
1959	202,000
1964	257,000
1969	301,000
1974	584,000

Cumulative geographical range of Bufo Marinus in Queensland, Australia

Our goal is to construct a mathematical model that best fits the given data. Note that the data is not given to us as "number of toads at five year intervals," and, in fact, this is often the case. For the toads in question, this would be virtually impossible data to obtain, although statistical methods may be used to estimate this value.

a) What assumptions could one make in order to convert the given data into population data? How realistic are these assumptions?

b) Suppose you are writing a small grant to study the population of this toad. You want to ask for money to hire two research assistants to gather additional data for three months. What additional data would you propose to obtain that would give you additional insight into the toad's total population?

c) For the ease of computation, we will assume that, on the average, there is one toad per square kilometer. (Of course, some fields are more densely populated, the middle of cities and lakes don't have any toads, and so on.) We will also count the toads in units of thousands, and time in units of years, beginning with 1939 as "time zero."
In your report, please include sketches of four solution curves, for differing values of the birth rate, as described below. You can generate these solution curves by using the calculator and putting

$$Y1=32.8 \, e\hat{\ }(kx)$$

and plotting this along with the above data in L1 and L2.

i) Start with $k = 0.1$ and vary the birth-rate k until the solution to the difference equation appears to fit the data well over the time period $[0, 35]$. Record this value of k.

ii) It also is possible to solve analytically for a value of k that will guarantee that the curve passes through exactly two of the data points. If $P(0) = 32.8$ find and record a value of k so that $P(5) = 55.8$. Find a different value of k that will give $P(35) = 584$.

iii) Plot this data on a log-linear plot.
If we suspect that data fits an exponential model, we can proceed as follows:
- Take the natural log of the dependent quantity (in our case, the population, P) so that we get a new data set of the form $(t, \ln(P(t)))$. If you have put the population data into L2, then put L3=ln(L2).
- Find the line of least squares that fits this data. This gives us an equation of the form $\ln(P(t) = mt + b$ where m and b are the slope and intercept corresponding to the line of best fit. Record the value of the correlation coefficient that indicates how well the data is approximated by a line (a correlation coefficient of 1 or -1 means perfect correlation).
- Exponentiating both sides allows us to find an initial population and value of the birth rate that best fits the given data.

iv) Do your four values of k agree closely with each other? Should they? Which of the four values do you think is the best model for the growth of *Bufo marinus* during the years for which we have data? Use this birth-rate to predict the toad's range in the year 2039. Given that the area of Australia is 7,619,000 square kilometers, how confident are you of this prediction? Explain your reasoning.

d) You may question the validity of the previous question's assumption that there is an average of one toad per square kilometer. Suppose we were wrong and there were actually an average of **two** toads per square kilometer.
As before, solve analytically for a value of k that will guarantee that the curve passes through exactly two of the data points. In particular, if we now assume that $P(0) - 65.6$, find and record a value of k so that $P(5) = 111.6$, and a different value of k so that $P(35) = 1168$. How do these values of k compare with the values you found in the previous question? What does this tell us? Comment on the importance of knowing the exact average density of the toad population.

For more recent information about the ever growing population of *Bufo marinus*, see the December 1995 issue of FROGLOG (number 15), the Newsletter of the World Conservation Union (IUCN), Species Survival Commission Declining Amphibian Populations Task Force (DAPTF).

9. **(Invasion of white pine)** We will examine data that indicates the prevalence of white pine (*Pinus strobus*) in the vicinity of the Lake of the Clouds, a lake in the Boundary Waters Canoe Area of northeastern Minnesota. The lake is deep (31 meters), calm, sheltered from wind, and devoid of inflowing streams. Consequently, the lake's bottom is covered with layers of annual sedimentary deposits. Each layer contains a sampling of pollen, and by counting the pollen belonging to each species of tree and herb, it is possible to estimate the ratio of one plant species to another.
White pine became extinct in northern Minnesota during the last period of glaciation, although it remained in southern climates such as Virginia. Once the glaciers began to retreat, the white pine began to expand northward again; it reappeared in northern Minnesota about

9400 years ago (H. E. Wright, "The roles of pine and spruce in the forest history of Minnesota and adjacent areas," *Ecology*, **49**, 937-55, 1968).

A. J. Craig (*Absolute pollen analysis of laminated sediments: a pollen diagram from northeastern Minnesota*, M.S. thesis, University of Minnesota, 1970) counted pollen in a phenomenal 9400 sedimentary laminæ from a core at Lake of the Clouds. According to his data, as white pine invaded the region surrounding the lake, it competed with entrenched populations of jack pine (*Pinus banksiana*) and red pine (*Pinus resinosa*), which occupy essentially the same coarse, sandy soil as *P. strobus*. The combined pine tree pollen accounted for about 60-70% of the pollen during the period of interest; other plant species remained essentially constant during the time period (with the exception of spruce (*Picea*) which decreased).

Craig's data is condensed and analyzed by W. A. Watts ("Rates of change and stability in vegetation in the perspective of long periods of time," *Quaternary Plant Ecology*, H.J.B. Birks and R.G. West, eds, Blackwell Scientific, 1973), and portions of this data are reproduced in the table below. Note that in the second column, time is measured in units of thousands of years.

Years Ago	Years since 9131 (Thousands)	P.bank/P.resin percentage	P.strobus percentage
9131	0.0	53.4	3.2
8872	0.259	65.5	0.0
8491	0.640	61.8	3.7
8121	1.010	55.2	3.4
7721	1.410	60.4	1.7
7362	1.769	59.4	1.8
7005	2.126	50.6	10.6
6699	2.432	51.6	7.0
6444	2.687	40.0	21.2
5983	3.148	29.7	34.2
5513	3.618	25.0	40.4
5022	4.109	32.5	29.8
4518	4.613	22.7	46.2
4102	5.029	31.6	33.0
3624	5.507	32.5	37.6
3168	5.963	27.1	39.5

Percentages of pollen for red/jack pine and white pine
for sedimentary layers at Lake of the Clouds, MN.

By scanning the data, it is clear that percentages of red and jack pine decreased during the time period indicated, whereas white pine pollen increased. If we assume that these pollen counts are representative of the relative populations of these species, then we have a basis for examining the population growth of *P. strobus* and the simultaneous decline of *P. banksiana* and *P. resinosa*.
 a) Enter the *P. strobus* data set and plot the data to model the growth of the white pine population near Lake of the Clouds. In particular,
   – Find and record an initial population, a net birth rate, and a coefficient of over-crowding that qualitatively matches the growth curve for this species of pine.

- Identify the three phases of logistic growth for this species of pine (if applicable). How long was the expansion phase?

- Estimate the equilibrium value for this species.

b) Repeat the above analysis for the *P. banksiana* and *P. resinosa* data set. Again, identify phases of growth, if applicable.

c) Using log-linear plots find equations for the growth and decline of these species of pines.

The next three models come from *Mathematical Models in Biology* by Leah Edelstein-Keshet.

10. **(Population Genetics: Hardy-Weinberg Law)** The genetic material in eukaryotes is made up of units called chromosomes. Organisms (such as humans) that are diploid have two sets of chromosomes, one obtained from each parent. A locus (a given location on a chromosome) may contain the blueprint instructions for some physical trait (such as eye color), which is determined by the combination of genes derived from each of the parents. A given gene may have one of several forms, called alleles.

Suppose that there are two alleles, denoted by a and A, and that these are passed down in the population from one generation to the next. A given individual could then have one of three combinations: AA, aa, or Aa. (The first two combinations are called homozygous, the last one heterozygous.) It is of interest to follow the distribution of genes in a population over the course of many generations.

A question we might explore is whether the relative frequencies of genes will change, and, if so, whether some new stable distribution will emerge. Until 1914 it was believed that any rare allele would gradually disappear from a population. After a more rigorous treatment of the problem it was shown that if mating is random and all genotypes (combinations of alleles, which in this case are aa, AA, and Aa) are equally fit (have an equal likelihood of surviving to produce offspring), then gene frequencies do not change. This fact is now known as the Hardy-Weinberg law.

Let

$$p = \text{frequency of allele } A = \frac{\text{total number of } A \text{ alleles}}{2N},$$

$$q = \text{frequency of allele } a = \frac{\text{total number of } a \text{ alleles}}{2N},$$

where $N =$ the population size.

We also make the following assumptions:

(1) Mating is random.
(2) There is no variation in the number of offspring from parents of different genotypes.
(3) Offspring are equally likely to survive.
(4) There are no mutations at any step.

We define the genotype frequencies in the population to be:

$u =$ frequency of AA genotype
$v =$ frequency of Aa/aA genotype
$w =$ frequency of aa genotype

So, for example, $u \times 100 = \%$ of population with the AA genotype. And similarly for the other genotypes.

The next step is to calculate the probability that parents of particular genotypes will mate. If mating is random, the mating likelihood depends only on the likelihood of encounter. This in turn depends on the product of the frequencies of the two parents. The mating table summarizes these probabilities. (Note that some entries are missing, and need to be determined.)

Mating Frequency Table

			Males		
Genotype			AA	Aa	aa
		Frequency	u	v	w
	AA	u	u^2	uv	uw
Females	Aa	v	—	v^2	—
	aa	w	—	—	—

Based on the likelihood of mating, one can determine the probability of a given match resulting in offspring of a given genotype. To do this, it is necessary to take into account the four possible combinations of alleles derived from a given pair of parents. The offspring table is a convenient way to summarize the information. (Again note that some entries are intentionally left blank.)

Offspring Frequency Table

Type of Parents	Frequency	Offspring Genotype Frequencies		
		AA	Aa	aa
$AA \times AA$	u^2	u^2	—	—
$AA \times Aa$	$2uv$	uv	uv	—
$AA \times aa$	—	0	—	—
$Aa \times Aa$	v^2	$v^2/4$	$v^2/2$	$v^2/4$
$Aa \times aa$	—	0	—	—
$aa \times aa$	—	0	—	—
Total		$(u^2 + uv + v^2/4)$	—	—

By counting up the total frequency of offspring of each type, we get values for u, v, and w for the next generation, which we shall call u_{n+1}, v_{n+1}, and w_{n+1}. It can be shown that these are related to the nth-generation frequencies by the equations

$$u_{n+1} = u_n^2 + u_n v_n + \frac{1}{4}v_n^2 \tag{4}$$

$$v_{n+1} = u_n v_n + 2u_n w_n + \frac{1}{2}v_n^2 + v_n w_n \tag{5}$$

$$w_{n+1} = \frac{1}{4}v_n^2 + v_n w_n + w_n^2 \tag{6}$$

As an example, consider equation 4 and note that the cumulative number of offspring of genotype AA (in the vertical column of the Offspring Frequency Table) leads to the given relationship.

(a) From the definitions of u, v, and w, it is clear that $u + v + w = 1$. Express p and q in terms of the frequencies u, v, and w and show that $p + q = 1$. Give an intuitive explanation why this is true.

(b) Complete the mating frequency table in the background information.

(c) Determine the offspring genotype frequencies for each possible paring of parents. (e.g. AA mother with AA father, Aa father, etc.) Use this information to complete the offspring table.

(d) Verify that the offspring genotype frequencies are governed by equations 4, 5, and 6.

(e) Show that $u_{n+1} + v_{n+1} + w_{n+1} = 1$. (i.e. that u, v, and w will always sum up to 1.)

(f) Suppose that (u^*, v^*, w^*) is a fixed point of the equations 4, 5, and 6. Use the information in part (e) to show that $u^* = \dfrac{(v^*)^2}{4w^*}$.

(g) Since w_{n+1} is related to u_{n+1} and v_{n+1} (similarly for w_n, u_n, and v_n) by the identity in (e), we can eliminate one variable from the equations 4, 5, and 6. Rewrite these equations in terms of u and v.

(h) Using part (g), show that

$$u_{n+1} = \left(u_n + \frac{1}{2} v_n \right)^2$$

$$v_{n+1} = \left(u_n + \frac{1}{2} v_n \right) \left[2 - 2 \left(u_n + \frac{1}{2} v_n \right) \right]$$

(i) Now show that

$$\left(u_{n+1} + \frac{1}{2} v_{n+1} \right) = \left(u_n + \frac{1}{2} v_n \right)$$

(j) Show that this implies that the frequencies p and q do not change from one generation to the next; i.e. that

$$p_{n+1} = p_n, \qquad q_{n+1} = q_n$$

Hint: you will need to use the fact that $p^2 + 2pq + q^2 = 1$.

11. **(Population Genetics: Positive Assortative Mating)** This uses much of the same background information as problem 10. Begin by reading the description of that problem.

   Now consider the following modification of the random mating assumption: Suppose that individuals mate only with those of like genotype (e.g., Aa with Aa, AA with AA, and so forth). This is called *positive assortative mating*. How would you set up this problem, and what conclusions do you reach?

12. **(Population Genetics: Negative Assortative Mating)** This uses much of the same background information as problem 10. Begin by reading the description of that problem.

   In *negative assortative matings*, like individuals do not mate with each other. Different types of models may be obtained, depending on assumptions made about the permissible matings. In the questions that follow it is assumed that homozygous females mate only with homozygous males of opposite type and that heterozygous females (Aa) mate with AA and aa males depending on their relative prevalence. The permitted matings are then as follows:

Females		Males
AA	$\times$	aa
aa	$\times$	AA
Aa	$\times$	AA
Aa	$\times$	aa

   Assuming that a single male can fertilize any number of females results in a mating table that depends largely on the female frequencies, as shown in the mating table.

Mating Frequency Table

Genotype			Males		
			AA	Aa	aa
		Frequency	u	v	w
Females	AA	u	0	0	u
	Aa	v	$\frac{vu}{u+w}$	0	$\frac{vw}{u+w}$
	aa	w	w	0	0

It is assumed that $u + v + w = 1$.

(a) Explain the entries in the table.

(b) Derive an offspring table by accounting for all possible products of the matings shown in the above mating table.

(c) Show that the fractions of AA, Aa, and aa offspring denoted by u_{n+1}, v_{n+1}, and w_{n+1} satisfy the equations

$$u_{n+1} = \frac{1}{2}v_n \frac{u_n}{u_n + w_n}$$

$$v_{n+1} = u_n + w_n + \frac{1}{2}v_n$$

$$w_{n+1} = \frac{1}{2}v_n \frac{w_n}{u_n + w_n}$$

(d) Show that $u + v + w = 1$ in the $(n + 1)$st generation. Use this fact to eliminate w from the equations.

(e) Show that the equations you obtain have a fixed point with $v^* = \frac{2}{3}$. Is there a unique value of (u^*, w^*) for this fixed point? What is the stability of this fixed point?

(f) Show that the ratio u/w does not change from one generation to the next.

The next 5 models come from Gerda deVries, Thomas Hillen, et al., *A Course in Mathematical Biology: Quantitative Modeling with Mathematical and Computational Methods*, 2006, Society for Industrial and Applied Mathematics.

13. **(Epidemic Models: HIV)** The table below shows data on the HIV epidemic in Cuba from 1986 until 2000. Design a model which describes the epidemic spread of HIV in Cuba, and fit the data in the table. Which are the relevant parameters of your model? Try to introduce control mechanisms to lower the number of AIDS cases. Compare your control mechanism with the data of the given time period. You need to look into facts about HIV transmission and Cuba's control strategy via the Internet or in appropriate textbooks.

Year	HIV cases	AIDS cases	Death from AIDS
1986	99	5	2
1987	75	11	4
1988	93	14	6
1989	121	13	5
1990	140	28	23
1991	183	37	17
1992	175	71	32
1993	102	82	59
1994	122	102	62
1995	124	116	80
1996	234	99	92
1997	364	121	99
1998	362	150	98
1999	493	176	122
2000	545	251	142

HIV data from Cuba 1986-2000 (data from [42]).

14. **(Epidemic Models: Smallpox)** An outbreak of smallpox in Abakaliki in southeastern Nigeria in 1967 was reported by Bailey and Thomas. People living there belong to a religious group that is quite isolated and declines vaccination. Overall, there were 30 cases of infection in a population of 120 individuals. The time (in days) between newly reported pox cases is given in the following sequence:

13, 7, 2, 3, 0, 0, 1, 4, 5, 3, 2, 0, 2, 0, 5, 3, 1, 4, 0, 1, 1, 1, 2, 0, 1, 5, 0, 5, 5.

Develop a model which describes these data and analyze the epidemic outbreak.

15. **(Epidemic Models: Influenza)** In the *British Medical Journal* in 1978, a report was published with detailed statistics of a flu epidemic in a British boarding school[1]. The school had 733 pupils[2], all but 30 of whom were boarders, and all boys. Of these, 512 were confined to bed during the epidemic, which lasted from January 22nd to February 4th, 1978. It seems that one infected boy initiated the epidemic. The school dealt with the epidemic by putting a boy to bed as soon as it was discovered that he was infected. Detailed data are shown in the table below. Model the epidemic outbreak.

---

[1]Anonymous. Influenza in a boarding school. *Brit. Med. J.*, 1:587, 1978

[2]Ampleforth Abbey (Publisher). Tashkent influenza ('red flu') January 1978. *Ampleforth Journal*, 83: 110–111

Date	In bed	Convalescent
Jan.22nd	3	0
Jan. 23rd	8	0
Jan. 24th	26	0
Jan. 25th	76	0
Jan. 26th	225	9
Jan. 27th	298	17
Jan. 28th	258	105
Jan. 29th	233	162
Jan. 30th	189	176
Jan. 31st	128	166
Feb. 1st	68	150
Feb. 2nd	29	85
Feb. 3rd	14	47
Feb. 4th	4	20

Influenza in a boarding school

16. **(Epidemic Models: Yellow Fever in Senegal in 2002)** Yellow fever is a viral hemorrhagic fever transmitted by infected mosquitoes. Yellow fever is spread into human populations in three stages:

   (1) **Sylvatic (or jungle)**. Yellow fever occurs in tropical rain forests where mosquitoes, which feed on infected monkeys, pass the virus to humans who work in the forest.

   (2) **Intermediate**. Yellow fever occurs as infected individuals bring the disease into rural villages, where it is spread by mosquitoes among humans (and also monkeys).

   (3) **Urban**. Yellow fever occurs as soon as an infected individual enters urban areas. This can lead to an explosive epidemic in densely inhabited regions. Domestic mosquitoes carry the virus from person to person.

   The epidemic can be controlled by vaccination. The yellow fever vaccine is safe and effective, and provides immunity within one week in 95% of those vaccinated.

   The table below shows a data set of yellow fever cases and yellow fever deaths during an outbreak in Senegal in 2002, collected from the Internet archives of the World Health Organization (WHO)[3]. As soon as the virus was identified, a vaccination program was started (Oct. 1, 2002). On Oct. 11, 2002, the disease was reported in Touba, a city of 800,000 residents. More information can be found on the WHO website[4].

   a) Develop a model for the three stages of yellow fever as outlined above.

   b) Include a fourth stage which describes vaccination in urban areas.

   c) Fit your model to the data.

   d) What would have happened without vaccination?

   e) Would you expect the disease to die out, or to become persistent?

   [3] http://www.who.int/research/en/
   [4] http://www.who.int/research/en/

Report date	Cases (total)	Deaths (total)
Jan. 18th	18	0
Oct. 4th	12	0
Oct. 1ith	15	2
Oct. 17th	18	2
Oct. 24th	41	4
Oct. 31st	45	4
Nov. 20th	57	10
Nov. 28th	60	ll

Yellow Fever in Senegal, 2002 (data from the disease outbreak news archives of the WHO `http://www.who.int/research/en/`)

17. **(Epidemic Models: Cholera in South Africa 2000–2001)** The seventh cholera pandemic began in Indonesia in 1961. Over the following 40 years, the virus *Vibric cholerae* O1 spread around the world, mainly into underdeveloped countries. In South Africa, the cholera epidemic arrived in mid-August 2000. In the table below there is data on the number of cholera cases and cholera-related death cases, taken from the disease outbreak news archives of the WHO.[5]

Date	Cases (total)	Deaths (total)
Oct. 13, 2000	2175	22
Oct. 18, 2000	3075	26
Oct. 19, 2000	3279	27
Oct. 26, 2000	3806	33*
Nov. 02, 2000	4270	32
Nov. 09, 2000	4583	33
Nov. 19, 2000	5285	35
Nov. 27, 2000	5876	35
Dec. 05, 2000	6548	35
Dec. 19, 2000	8137	41
Dec. 29, 2000	11183	51
Jan. 09, 2001	15983	60
Jan. 14, 2001	19499	66
Jan. 25, 2001	27431	74
Feb. 04, 2001	37204	85
Feb. 14, 2001	48647	108
Feb. 22, 2001	56092	120
Mar. 03, 2001	62607	131
Mar. 14, 2001	69761	139
Mar. 28, 2001	78140	163
Apr. 16, 2001	86107	181

Cholera outbreak in South Africa 2000-2001 (data from the WHO archives `http://www.who.int/research/en/`). The number highlighted with an asterisk was later corrected by the WHO to 31; to show that data can be inconsistent, the original number is given here.

---

[5] `http://www.who.int/research/en/`

Large cholera outbreaks are usually related to contaminated water. The cholera virus is present in brackish water through algae blossom and through human feces. Only 10–20% of infected individuals suffer from severe symptoms. Many individuals do not show symptoms at all, but their feces are infectious. Cholera is a serious disease since the progress of symptoms can be very fast if not treated.

The WHO recommends four major control mechanisms, namely,

(1) hygienic disposal of human feces,
(2) adequate supply of safe drinking water,
(3) good food hygiene and cooking, and.
(4) washing hands after defecation and before meals.

More information about this disease, control mechanisms, and vaccination can be found at the website of the WHO.[6]

Develop a model for the outbreak of cholera in South Africa.

a) First, model the epidemic without any control mechanism.

b) Then include the recommended control mechanisms in the model and see if you can obtain a better fit to the data.

c) Use your model to determine which of the above control mechanisms is most effective.

d) Can you predict the further development of the disease, provided that all control measures are in place?

---

[6]http://www.who.int/research/en/