Math 132

Population Growth: Raleigh and Wake County

S. R. Lubkin

Application

Ask anyone who's been living in Raleigh more than a couple of years what the biggest issue is here, and if the answer has nothing to do with the weather, it'll be something to do with the population growing too fast. But how fast *is* Raleigh growing? And how big is it going to get before it stops growing? Are we the next Atlanta? The next Los Angeles?

We will take data from the US Census for Raleigh and for Wake County, analyze the past growth, and try to predict the future of the population here.

Mathematical Models

We have already seen some basic population models. The simplest is the **exponential model**, written as a word equation

$$\text{(rate of change of population)} = \text{(growth rate)*(population)} \qquad (1)$$

or as a differential equation

$$\frac{dy}{dt} = ry \qquad (2)$$

where *y(t)* is the population as a function of time, and *r* is the **growth rate**. Some people prefer to write

$$r = b - d \qquad (3)$$

where *b* is the birth rate and *d* is the death rate, or

$$r = b - d + I - E \qquad (4)$$

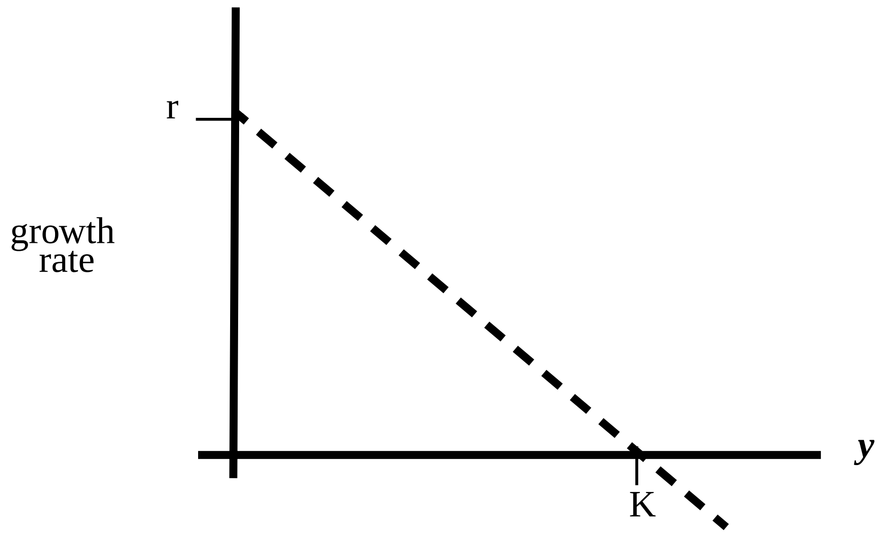
where *I* and *E* are the immigration and emigration rates. All that equations (3)-(4) are saying is that the overall growth rate includes components of births, deaths, immigrations, and emigrations. (Question: How would you write the word equations describing eqs. (3) and (4)?)

The second simplest model we have seen is the **logistic model**, whose word equation is

$$\text{(rate of change of population)} = \text{(growth rate)*(population)} \qquad (5)$$

$$\text{(growth rate)} = \text{(maximum growth rate)(1 – fraction of carrying capacity)} \qquad (6)$$

so that we can think of the growth rate in the logistic model as declining as the population gets larger, until when the population reaches its **carrying capacity** *K*, the growth rate (births - deaths + immigrations + emigrations) goes to zero. The carrying capacity could be determined in a natural population by the food or water available, or in a human population by the space, economic opportunity, and infrastructure available. We can think of the logistic model as having a growth rate which is a decreasing function of the population:



This means that as the population gets larger, the growth rate declines. Contrast this with the exponential model, where no matter what the population is, the growth rate remains the same. Later in the course we will see an example of a growth rate which is an *increasing* function of the population, meaning that as things get crowded, the population grows even faster.

The logistic model is written as a differential equation as

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right) \tag{7}$$

Do this: To solve the exponential and logistic differential equations (2) and (7), open up a Maple session. Type the following two lines, to define your two differential equations:

```
> exponential:=diff(y(t),t) = r*y(t);
> logistic:=diff(y(t),t) = r*y(t)*(1-y(t))/K;
```

Remember to press return in Maple every time you change a command. To solve the differential equations, with the initial conditions $y(0) = y_0$, enter the following:

```
> expsol:=dsolve({exponential,y(0)=y0},y(t));
> logisticsol:=dsolve({logistic,y(0)=y0},y(t));
```

These give the solutions for the exponential and logistic models:

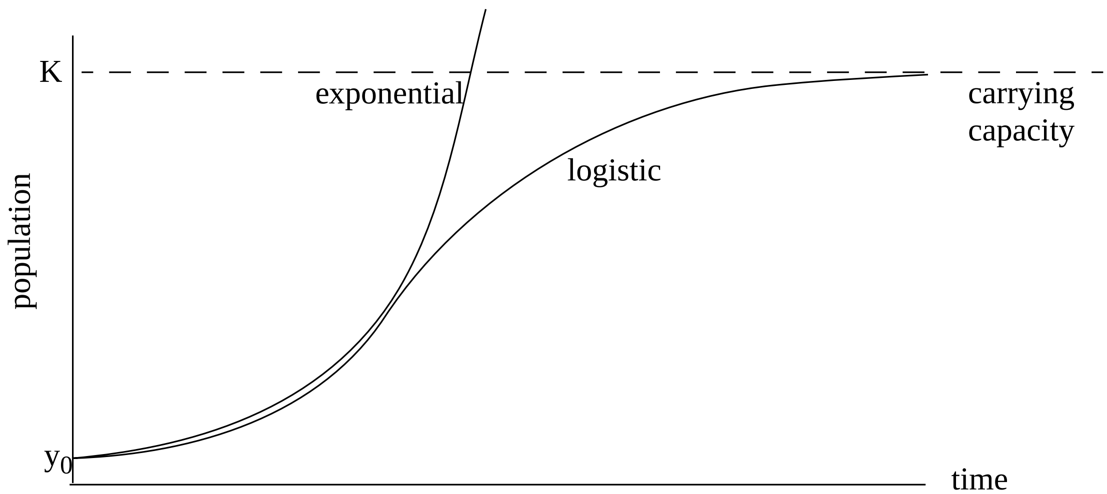
exponential:

$$y(t) = y_0 \exp(rt) \tag{8}$$

logistic:

$$y(t) = \frac{K}{\left(1 - \left(1 - \frac{K}{y_0}\right)\exp(-rt)\right)} \tag{9}$$

These two solutions have these general shapes:



Here is the US Census data for Raleigh and Wake County through 1990.

year	Raleigh	Wake
1920	24,418	75,155
1930	37,379	94,757
1940	46,897	109,544
1950	65,679	136,450
1960	93,931	169,082
1970	122,830	228,453
1980	150,255	301,327
1990	212,092	423,380

It looks like the Raleigh population in 1900 must have been about 20,000 and in 2000 it may be around 250,000, but we can't be sure without doing some parameter fitting and analysis.
Our goal is to be able to reliably predict the population of Raleigh and Wake County, both for 2000 and well into the future.

The models (8) and (9) contain two or three **parameters**, r, K, and y_0. A parameter is a number which is constant as the independent variable (time) changes, but may vary from "experiment" to "experiment". We will do **least squares data fitting** to find the parameter values r, K, and y_0, which make (8) or (9) fit as closely as possible to the data series in the table. We will then use our best-fit model to make predictions of the population at later times.

The task of **data fitting** is to determine the parameter values *r, K*, and y_0 which will make all the measured data points $y_i(t_i)$ lie as close as possible to the curve *y(t)*. **"Close" means that the sum of the squares of the distances between each point $y_i(t_i)$ and the curve *y(t)* are as small as possible, within the constraint of the form of the function *y(t)*.**

We form the sum of squares *S* as

$$S = \sum_{i=1}^{n} (y_i(t_i) - y(t_i))^2 \tag{10}$$

or, in the case of the exponential model (8),

$$S_e = \sum_{i=1}^{n} (y_i(t_i) - y_0 \exp(rt))^2 \tag{11}$$

and in the case of the logistic model (9),

$$S_l = \sum_{i=1}^{n} \left(y_i(t_i) - \frac{K}{\left(1 - \left(1 - \frac{K}{y_0}\right)\exp(-rt)\right)} \right)^2 \tag{12}$$

We can then think of *S* as a function of the parameters *r, K*, and y_0 . If we can minimize *S* by changing *r, K*, and y_0 , we will have the version of each model which fits the data best.

Methods of Solution

We'll use the spreadsheet to organize our data and to form the sum of squares. Excel's Solver function will minimize *S* for us. The Data Table will have the following columns:

- times of data points, t_i
- Population at those times, $y_i(t_i)$
- value of model function at those times, $y(t_i)$
- square of difference, $S_i(t_i)$

We will also construct a Parameter Table keeping track of our parameter values and the sum of squares. We can easily plot both the data and the fitted curve using the ChartWizard, and visually assess how good the fit is for different parameter values.

We will allow Solver to change the parameter values we are using until the least squares solution is found (the parameters which minimize *S*).

Implementation (step-by-step instructions)

1. Open the spreadsheet.
2. Copy the table headings and data in the first two columns below. (Wait till steps (5) and (7) to

enter the formulas in the last 2 columns). (Be sure to use the times exactly as listed below, or

t_i = year since 1900	y_i = Raleigh pop.	$y(t_i)$ of model	$(y_i-y(t_i))$^2
20	24,418	= $?$?*exp($?$?*a2)	= (c2-b2)^2
30	37,379		
40	46,897		
50	65,679		
60	93,931		
70	122,830		
80	150,255		
90	212,092		

the numbers will be very difficult to interpret. Strange but true; trust me.)

3. Make a graph of the data in the first 2 columns using the ChartWizard and type “scatter” without the curve connecting the points. (By the way, Trendline can’t help you on this assignment. Why?)
4. Make a second table on the same spreadsheet, with the headings below, and initial guesses for the three parameters.[1]

r	y0	K	S
.01	20000	1000000	= sum(d2:d9)

5. In the third column of the Data Table, evaluate the exponential model (8) at the time indicated in the first column. (The addresses of the parameters are given as $?$? in the table above, but you will have the cell addresses of your parameters there instead.) Pull down the formula to all cells in the column, to evaluate *y(t)* at each time. How good a guess did we use? Can you change the guesses a little to get column C to be closer to column B?
6. Change your chart to add the third column in the Data Table, or start a new chart. It is nicest if the actual data is in dots and the fitted function is in a line with no dots. Notice that the initial guess does not give a perfect fit.
7. In the last column of the Data Table, form the square of the difference of the previous two columns.
8. In the Parameter Table, for the value of S, put the sum of the entries of the last column of the

1. The guesses are from common-sense. Most human populations grow from 0-5% per year, so we guess r to be 1%. The table looks like the Raleigh population in 1900 was around 20,000 so we guess y_0 to be 20,000. And a big city would be around 1 million, so that is our guess for the carrying capacity K. You can start with your own guess, and the least-squares fit should still be the same.

Data Table.

9. Select Solver from the Tools menu at the top of Excel[1]. In Solver's window, enter the address of the objective function *S* (the thing we want to optimize), tell it we want to *minimize* it, and enter the locations of the parameters that Solver is allowed to vary (of the Parameter Table). Remember, for the exponential model, only *r* and y_0 can be changed.
10. Click Solve. Watch as the parameter values in the Parameter Table change. Watch as the graph changes. Watch as the value of *S* decreases. Assuming there was no error message, the best-fitting parameter values for the initial population (in year 1900) and growth rate have just appeared in the parameter table.
11. Repeat the process *in a separate table* for the logistic model (9). Be sure to tell Solver that it can change *r*, *K*, and y_0 . Which model gives a lower least sum of squares *S*?

Assessment

Raleigh is growing at a very high rate, apparently. The exponential model gives a growth rate *r* of 2.8% per year, and the logistic model gives the maximum growth rate *r* of 3.0%.

The models predict that the population in 2000 is 280,858 and 274,673 respectively. Which is a better model? One answer is the best model is the one which makes the best predictions. We can compare the predictions with the actual census data: In 2000, Raleigh had 276,093 people, and Wake had 627,846. But this doesn't help us if we have to act now in planning for a future that we haven't measured yet! The models predict that the Raleigh population in 2050 will be 1,190,997 and 864,510 respectively. We'd have to wait an awfully long time to test the models' predictions! One obvious drawback to the exponential model is that it predicts that the population of Raleigh will never level off. The logistic model predicts that Raleigh will reach a carrying capacity of 2,131,020, which is large and frightening, but not inconceivable.

Another point of comparison of the two models is the sum of squares S. The logistic model gives a slightly smaller S than the exponential model (1.26×10^8 versus 1.29×10^8). That means that the logistic model fits the data we have slightly better than the exponential model does.

But is the logistic model the best possible model? Can we come up with another model which fits the data even better? That is the subject of the last lesson in the course....

---

1. If Solver isn't there under Tools, you need to go to Tools -> Add-Ins, scroll down, and click that Solver should be added. Then you can use Tools -> Solver. If Solver isn't in the Add-Ins, your installation of Excel will not work for this project, and you will either need to install Solver or find another computer.