

Math 132

Population Growth: the World

S. R. Lubkin

Application

If you think growth in Raleigh is a problem, think a little bigger. The population of the world has been growing spectacularly fast in the last few decades. What will the world population be in 10 years? In 50 years? Will the world's population ever level off? If so, at how many billions? What will life be like for your grandchildren?

We will look at world population data and do least-squares fitting to some population models. We will try to find the best model, and use it to make predictions.

Mathematical Models

We have already seen some basic population models. written as a word equation

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

In this lesson, we will explore different models for the **growth rate**.

The simplest is the **exponential model**, with a constant growth rate:

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

where $y(t)$ is the population as a function of time, and r is the growth rate.

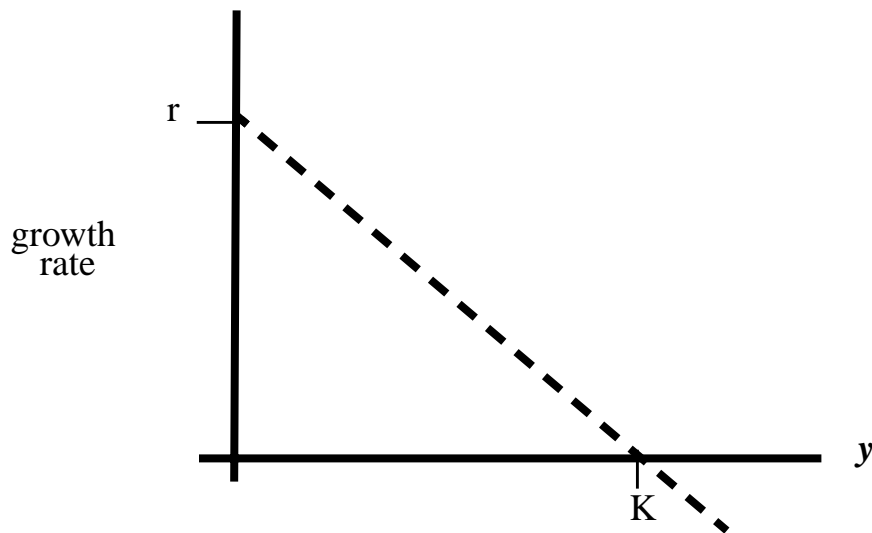
The second simplest model we have seen is the **logistic model**, whose word equation for the growth rate is

$$(\text{growth rate}) = (\text{maximum growth rate})(1 - \text{fraction of carrying capacity}) \quad (3)$$

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 goes to zero.

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. The logistic model is written as a differential equation as

$$\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right) \quad (4)$$

We already found the solutions for the exponential and logistic models:

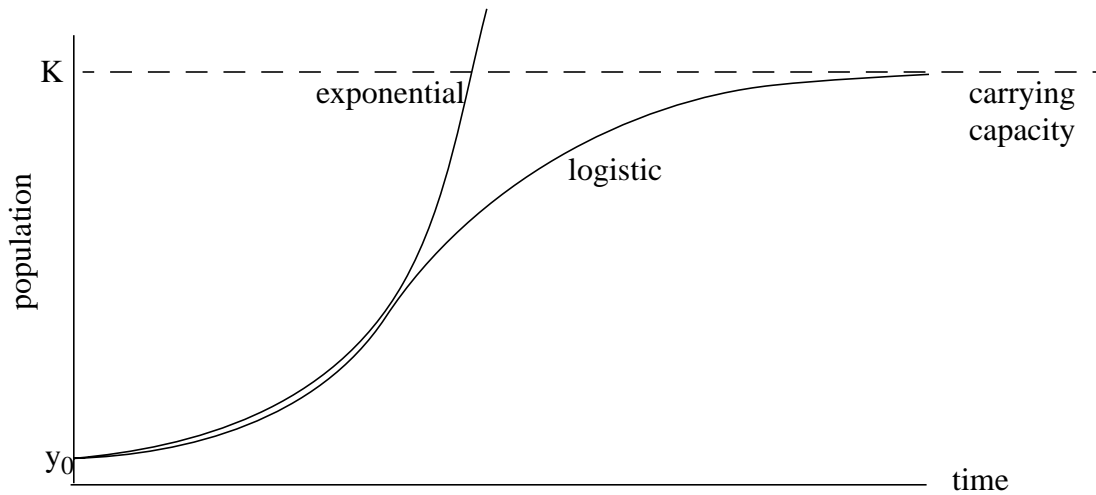
exponential:

$$y(t) = y_0 \exp(rt) \quad (5)$$

logistic:

$$y(t) = \frac{K}{\left(1 - \left(1 - \frac{K}{y_0}\right) \exp(-rt)\right)} \quad (6)$$

These two solutions have these general shapes:



Here is the available UN data for world population¹.

year	population (billions)
1750	0.791
1800	0.978
1850	1.262
1900	1.650
1950	2.521
1999	5.978

If we graph this data we see something pretty interesting.

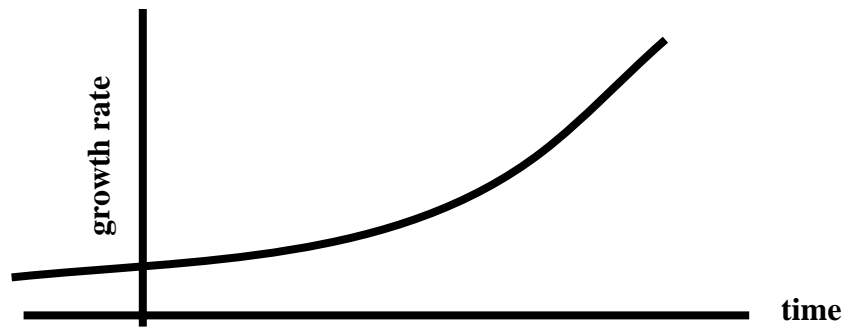
Do this: Enter the data above into two tables in your spreadsheet. Make a scatter chart of each. Add an exponential trendline to each.

Notice that the trendline, the best-fitting exponential model, doesn't fit the data very well. In fact, the growth rate is increasing over time, as the population gets larger. In the logistic model, the growth rate was *decreasing* as the population got larger, not increasing! This means that we can't even try to fit the logistic model to world population data! We need a growth rate which is an *increasing* function of time or population.

One possibility would be the so-called **superexponential model**, where we make the growth rate an exponential in time:

1. <http://www.popin.org/6billion/b3.htm> and <http://www.popin.org/6billion/t02.htm>

$$r = r_0 e^{r_1 t} \quad (7)$$



This corresponds to thinking of the growth rate as proportional to the development of social, scientific, and technological structures, and the discovery of new resources to exploit. The superexponential differential equation then becomes

$$\frac{dy}{dt} = (r_0 e^{r_1 t})y \quad (8)$$

To solve this, we have Maple do the integration:

```
> super:=diff(y(t),t)=r0*exp(r1*t)*y(t);
> dsolve({super,y(0)=y0},y(t));
```

and we get the solution

$$y(t) = y_0 e^{\frac{r_0}{r_1}(e^{r_1 t} - 1)} \quad (9)$$

which can also be written

$$y(t) = y_0 \exp\left(\frac{r_0}{r_1}(\exp(r_1 t) - 1)\right) \quad (10)$$

Our task now will be to do least-squares data fitting for the world population data to the exponential model and the superexponential model:

exponential:

$$y(t) = y_0 \exp(rt) \quad (11)$$

superexponential:

$$y(t) = y_0 \exp\left(\frac{r_0}{r_1}(\exp(r_1 t) - 1)\right) \quad (12)$$

The models (11) and (12) contain **parameters**, r , r_0 , r_1 , 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 , r_0 , r_1 , and y_0 which make (11) or (12) 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 world population at later times.

The task of **data fitting** is to determine the parameter values 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

$$S = \sum_{i=1}^n (y_i(t_i) - y(t_i))^2 \quad (13)$$

or, in the case of the exponential model (11),

$$S_e = \sum_{i=1}^n (y_i(t_i) - y_0 \exp(rt_i))^2 \quad (14)$$

and in the case of the superexponential model (12),

$$S_s = \sum_{i=1}^n \left(y_i(t_i) - y_0 \exp\left(\frac{r_0}{r_1}(e^{r_1 t_i} - 1)\right) \right)^2 \quad (15)$$

We can then think of S as a function of the parameters r , r_0 , r_1 , and y_0 . If we can minimize S by changing r , r_0 , r_1 , 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

$t_i =$ years since 1700	$y_i =$ world pop. (billions)	$y(t_i)$ of model	$(y_i - y(t_i))^2$
50	0.791	$= b * \exp(a * t)$	$= (c - b)^2$
100	0.978		
150	1.262		
200	1.650		
250	2.521		
299	5.978		

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	y_0	S
.01	0.5	$= \text{sum}(d2:d7)$

5. In the third column of the Data Table, evaluate the exponential model (11) 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 Data Table.
9. Select Solver from the menu at the top of Excel². In Solver’s window, enter the address 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 world population in 1700 was around 0.5 billion so we guess y_0 to be 0.5. You can start with your own guess, and the least-squares fit should still be the same.

objective function S (the thing we want to optimize), tell it we want to *minimize* it, and enter the locations of the parameters r and y_0 that Solver is allowed to vary (in the Parameter Table).

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 1700) and growth rate have just appeared in the parameter table.
11. Repeat the process *in a separate pair of tables* for the superexponential model (12). You will

$t_i =$ years since 1700	$y_i =$ world pop. (billions)	$y(t_i)$ of model	$(y_i - y(t_i))^2$
50	0.791	$= c \cdot \exp(a \cdot \exp(b \cdot a^2) - 1)$	$= (c - b)^2$
100	0.978		
150	1.262		
200	1.650		
250	2.521		
299	5.978		

need a very good guess for the starting values for the parameters, or Solver will be starting too far from the correct answer. You might try using, for r_0 and y_0 , the values you found from the fitting of the exponential, since the superexponential is very close to the exponential if r_1 is very small. Be sure to tell Solver that it can change r_0 , r_1 , and y_0 . Which model gives a lower

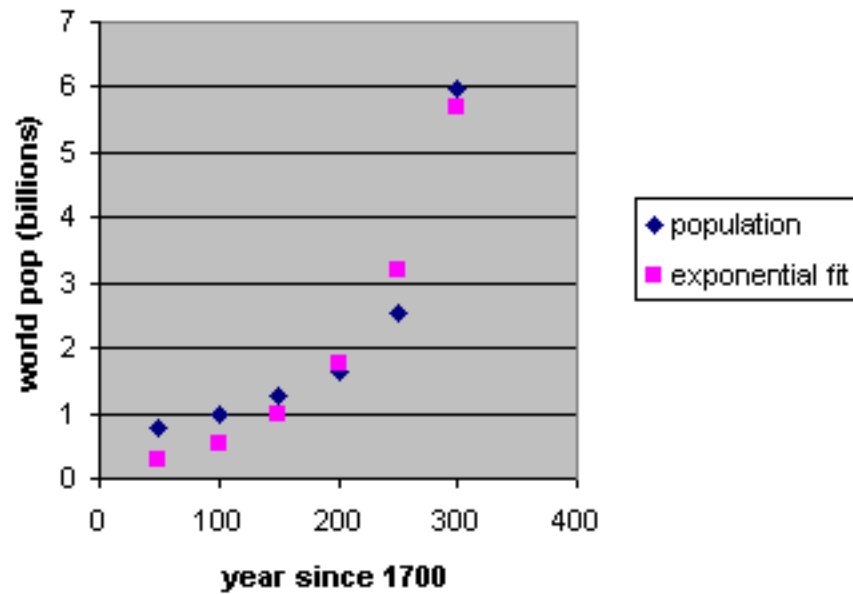
r_0	r_1	y_0	S
.01	.01	0.5	$= \text{sum}(d2:d7)$

least sum of squares S ?

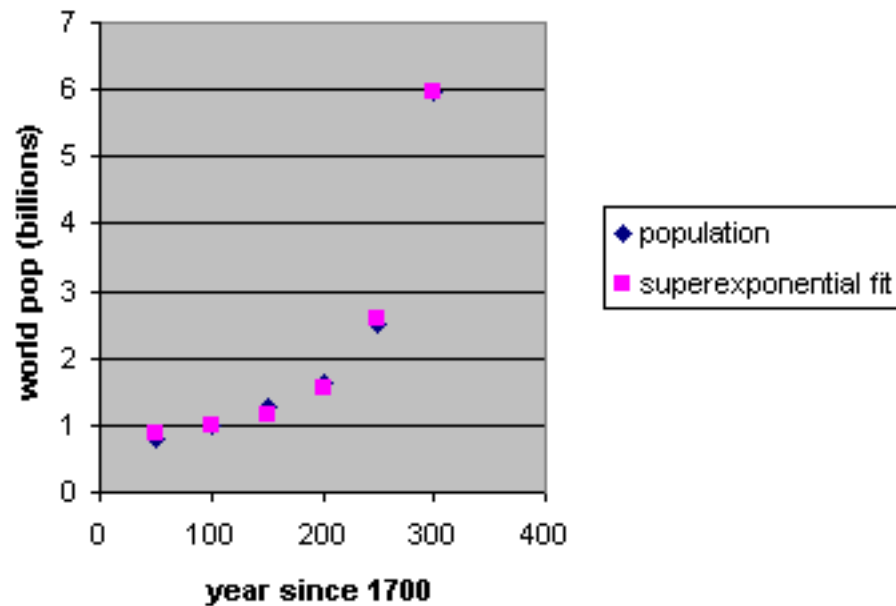
-
2. If Solver isn't there under Tools, you need to go to Tools -> Add-Ins, scroll down, and click that Solver should be added. Then you can use Tools -> Solver. If Solver isn't in the Add-Ins, your installation of Excel will not work for this project, and you will either need to install Solver or find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

Assessment

World population fit by exponential model



World population fit by superexponential model



The superexponential model fits world population data better than the exponential model on three counts: (1) The sum of squares is 30 times larger with the exponential model, meaning the best-fitting exponential model is much farther away from the data points; (2) visually, the points are right on top of each other with the superexponential model, but not for the exponential model; (3) there is an obvious “trend” in the exponential fit that suggests that the errors in fitting are systematic rather than random.

We currently have 6 billion people in the world. If we project a little bit into the future, with the exponential model, we get a world population in the years 2050 and 2100 of 10 billion and 19 billion, respectively. If we project using the superexponential model, we get a world population in 2050 and 2100 of 26 and 309 billion, respectively.

Table 1: World population and model predictions

model	2000	2050	2100
exponential	6 billion	10 billion	19 billion
superexponential	6 billion	26 billion	309 billion

Draw your own conclusion.