6.7 Population Models

Population models are used to predict the future size of a population and its long term behavior. Population models come in two flavors:

- models on the overall population size
- models that distinguish by age group.

The first type of model is of interest in questions such as how much food is necessary to feed the growing world population or for predictions on the needs for energy and natural resources, at all levels of govenance (city, state, country). The second type of model is used, for example, to predict how much money is needed to pay Social Security (a hot topic in 1997/98), or to predict the number of students that are expected to enroll in the California State University system (Tidal Wave II). In the latter type of models, the population is divided into age groups, often spanning either five or ten years: 0-5, 6-10, 11-15,... or 0-10, 11 - 20, 21-30,....

We will start by deriving a model for the overall size of a population. First, we define variables:

$$\begin{aligned} n &= \text{time} \\ P(n) &= \text{population at time } n \end{aligned}$$

To derive the recursive model equation, we follow the paradigm

$$\begin{aligned} \text{new} &= \text{old} + \text{change} \\ P(n+1) &= P(n) + \text{change in population} \end{aligned}$$

Now we have to make assumptions about the change in population. To do this, we need to think about which factors influence the size of the population over time. Some examples are

- births/deaths (influenced by gender distribution, age structure, ethnicity, environment (3^{rd} world versus industrialized nation))
- immigration/emigration
- wars/man-made disasters
- natural disasters (earthquakes, hurricanes).

For a first model, we will take into account only births and deaths. At this point we need to make an assumption about the number of births and deaths per time unit. The simplest scenario is a fixed number of births and deaths per time unit, which leads to a linear function for the population size (since linear functions have constant first unit differences). However, this does not seem very realistic. A more likely assumption would be that a larger population results in more births and deaths than a smaller population. This translates into a proportional relation between the number of births and deaths and the population size:

$$\begin{aligned} \text{births in year } n &= b \cdot P(n) \qquad & b &= \textit{birth rate} \\ \text{deaths in year } n &= d \cdot P(n) \qquad & d &= \textit{death rate} \end{aligned}$$

Altogether, the change in population is given as the number of births minus the number of deaths:

$$\begin{aligned} \text{change in population} = \underbrace{b \cdot P(n)}_{\text{increase}} + (\underbrace{-d \cdot P(n)}_{\text{decrease}}) &= (b-d) \cdot P(n) \\ &= c \cdot P(n) \end{aligned}$$

where $c = b - d$ is the (net) *growth rate*. Notice that we have assumed that the birth and death rates are constant over time (which may be true in the short run, but most likely is not true in the long run due to societal changes, birth control methods, etc.). However, this is just a first model, and time-dependent birth and death rates can be added as a refinement later if needed. Thus, we have:

$$P(n+1) = P(n) + c \cdot P(n)$$

$$\boxed{P(n+1) = (1+c) \cdot P(n)}$$

This is a model of the form discussed in Section 6.1, for which we have derived the explicit solution. In the population example, the constant of proportionality is given by $1+c$. This is the factor that is taken to the n$^{\text{th}}$ power in the explicit solution and multiplied by the initial value for the population:

$$P(n) = (1+c)^n \cdot P(0).$$

The factor $(1+c)$ is called the *growth factor*. It indicates the factor by which the population increases per time unit. We can immediately derive the following behavior:

$$\begin{aligned} &\text{if } 1+c > 1 \Rightarrow \text{population increases} \\ &\text{if } 1+c = 1 \Rightarrow \text{population stays the same} \\ &\text{if } 1+c < 1 \Rightarrow \text{population decreases} \end{aligned}$$

Does this make sense?

$$\begin{aligned} 1+c > 1 &\Rightarrow c > 0 \Rightarrow b-d > 0 \quad \text{or} \quad \text{more births than deaths} \\ 1+\mathrm{c} = 1 &\Rightarrow c = 0 \Rightarrow b-d = 0 \quad \text{or} \quad \text{births and deaths balance exactly} \\ 1+c < 1 &\Rightarrow c < 0 \Rightarrow b-d < 0 \quad \text{or} \quad \text{more deaths than births} \end{aligned}$$

In order to make predictions with this model, we need to know the **initial population size** and the **birth** and **death rates**. Let’s start with a simple example of cell growth.

At 8 AM, a biologist has started a culture in a Petri dish with cells occupying an area of 1.5 cm^2. The cells divide regularly, and the culture contains plenty of nutrients, which ensures that no cells will die. The Petri dish is photographed every 10 minutes, and the area occupied by the cells is measured on the Polaroid photographs. As all the cells have the same size, the number of cells can be computed from the total area occupied. The biologist would like to develop a model for the area occupied by the cells; the experiment will provide the data against which the model can be tested to validate his assumptions about the cell growth.

Given: Cells initially occupy 1.5 cm^2
Regular cell division, no cells die
Data is collected every 10 minutes

Now we translate this information into mathematical expressions, starting with the model variables.

Variables: $n =$ time (in minutes)
$c(n) =$ area (in cm^2) occupied by the cells at time n

Wait a minute - is this correct? Recall what we said n to be: Time in minutes. Therefore, $c(0) = 1.5$ translates to: At time 0 minutes, the area occupied by the cells is 1.5 cm^2. However, the area occupied was equal to 1.5 cm^2 at 8 AM, not at time 0 minutes. To correct this problem, we modify the meaning of the input variable. Instead of measuring time in absolute terms, we measure time relative to 8 AM:

Variables: $n =$ time (in minutes) after 8 AM
$c(n) =$ area (in cm^2) occupied by the cells at time n

With this definition, $c(0) = 1.5$ translates properly. Defining the input variable as measuring time in such a way that $n = 0$ corresponds to the initial time is very common and insures that we know one quantity of the explicit solution immediately.

Given: $c(0) = 1.5$
birthrate b, needs to be determined from data
death rate $d = 0$ (as no cells die)

Using the explicit solution, we get

$$\begin{aligned} c(n) &= (1+(b-d))^n c(0) \\ &= (1+b)^n\, 1.5 = 1.5 \cdot (1+b)^n \end{aligned}$$

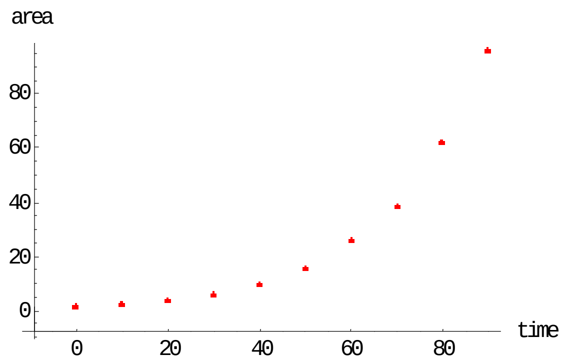
Before we can make any predictions, we need to determine the value of b from the experimental data. Here is what the biologist measured:

Time (AM)	8:00	8:10	8:20	8:30	8:40	8:50	9:00	9:10	9:20	9:30
Area (cm^2)	1.5	2.37	3.82	5.98	9.45	15.26	25.89	37.94	62.19	95.84

Translating the table to reflect the meaning of the input variable results in:

Time (in minutes after 8 AM)	0	10	20	30	40	50	60	70	80	90
Area (cm^2)	1.5	2.37	3.82	5.98	9.45	15.26	25.89	37.94	62.19	95.84

The shape of the data graph indicates that an exponential function is likely, which is in agreement with the model we have developed.



We still need to find the growth factor. We start with the explicit solution and solve for the growth factor 1 + b:

$$c(n) = (1+b)^n c(0)$$

$$\frac{c(n)}{c(0)} = (1+b)^n \qquad \text{divide by } c(0)$$

$$\left[\frac{c(n)}{c(0)}\right]^{1/n} = \left((1+b)^n\right)^{1/n} \qquad \text{take both sides to power } 1/n$$

$$\left[\frac{c(n)}{c(0)}\right]^{1/n} = 1+b \qquad \text{simplify}$$

In particular, for n = 10, we get

$$\left(\frac{c(10)}{c(0)}\right)^{1/10} = 1+b\,.$$

This is exactly the unit ratio! As the model solution is an exponential function, we know that the unit ratios are constant. Thus, the **growth factor equals any of the unit ratios**. However, if we start from data, then the unit ratios are only approximately constant, and we have to figure out which of these possible values for the growth factor should be used. Let's compute the unit ratios for the cell data (using the palette function **UnitRatios** from the DataFit palette):

{1.0468, 1.04889, 1.04584, 1.04682, 1.04909, 1.05428, 1.03895, 1.05066, 1.0442}

They look very close - which one should we use: The smallest? The largest? The first one? The average of all the values? The goal is to find a method that will work on all kinds of data, even if the values are not as close as in this example.

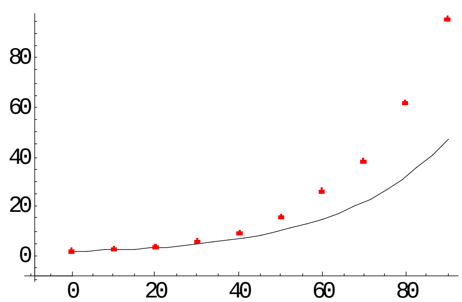
Activity 6.7.1

Below are the graphs of four different models of cell growth, each one displayed together with the data. These four models result from the four different ways of selecting the growth factor suggested above, namely using

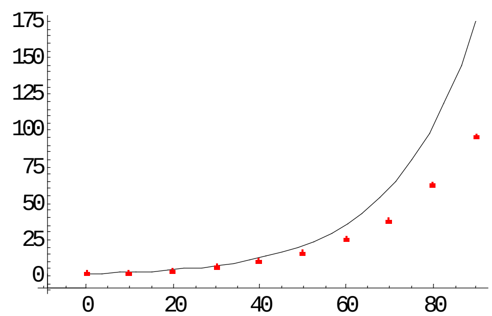
a) the smallest b) the largest c) the first d) the average unit ratio

as the value of the growth factor 1 + *b*.

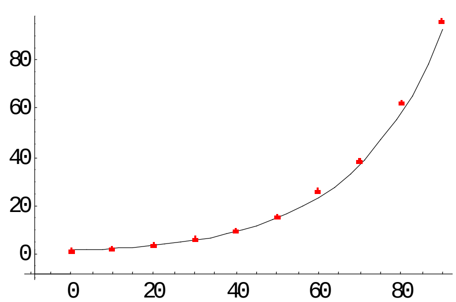
a) (1+*b*)=1.03895



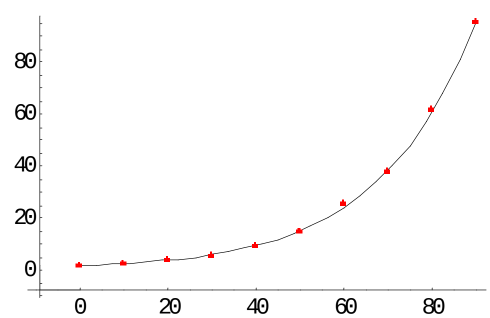
b) (1+*b*) = 1.05428



c) (1+*b*) = 1.0468



d) (1+*b*) = 1.04728



Describe for each case how the model compares to the data. Which model would you select as the one that best fits the data? (Give reasons for your answer.) Do you think that the particular method (smallest, largest, first, average unit ratio) by which the growth factor was selected for your best fitting model would work equally well with other data? Why or why not?

Generally, taking the average unit ratio makes the most sense, as this method takes into account all values (unlike choosing the first unit ratio) and also balances out the effect of very large or very small values. With this choice of growth factor, the (analytical) model is given by

$$c(n) = 1.5 \cdot 1.04728^n .$$

How does this analytical model, which used the data to compute the growth factor, compare to the model obtained when using a least squares fit for an exponential function? Using the palette function **ExpoFitFunc**, we get the following result (assuming the data is in a list called A):

```
ExpoFitFunc[A]
```

$1.49917\ 1.04754^x$

Identifying the relevant parameters in the fitted function, we get $c(0) = 1.49917$, and a growth factor of 1.04754 (this is the constant that is raised to the power of x). Listing the parameters for the two types of model in the table below, we see that they are quite close.

	Analytic Model (from assumptions)	Least Squares Fit Model
$c(0)$	**1.5**	**1.49917**
growth factor $1+b$	**1.04728**	**1.04754**

So what is the advantage of an analytical model over a least squares fit model? The analytical model gives a meaning to the constants (initial population, growth factor) and also gives insight into the “Why?” explaining the underlying process that results in this particular type of model. From the derivation we can conclude that an exponential model works when the change in population is proportional to the population size. Therefore, we can identify the meaning of the parameters in the model solution that results from an exponential least squares fit.