

Phenomenological Models and Sustainability using Gapminder

A struggle for existence inevitably follows from the high rate at which all organic beings tend to increase. Every being, which during its natural lifetime produces several eggs or seeds, must suffer destruction during some period of life, and during some season or occasional year, otherwise, on the principle of geometric increase, its numbers would quickly become so inordinately great that no country could support the product.

Charles Darwin, The Origin of Species, 1859

Learning Objectives

After completion of this module, the student will be able to

1. use Gapminder to explore “social, economic and environmental development at local, national and global levels”
2. perform and interpret logarithmic transformations for graphical display
3. define population growth rate, per capita growth rate, and exponential growth
4. build a data-driven phenomenological model with a minimal number of parameters using the framework of differential equations
5. make predictions based on a mathematical model

Knowledge and Skills

1. logarithmic transformation
2. continuous time population models
3. fitting a straight line to data
4. doubling time of an exponentially growing population
5. population growth, per capita growth rate

Prerequisites

1. calculating percent changes
2. straight lines
3. natural logarithm, exponential function
4. graphing in EXCEL
5. fitting a straight line to data points in EXCEL and displaying the equation

Gapminder

Gapminder (<http://www.gapminder.org/>) “is a non-profit venture promoting sustainable global development and achievement of the United Nations Millenium Development Goals by increased use and understanding of statistics and other information about social, economic and environmental development at local, national and global levels.” (<http://www.gapminder.org/about-gapminder/>)

In-class Activity 1

To get a sense for how various variables depend on population size, explore **Gapminder World**: Display *Population > Population, total* on the horizontal axis and the following secondary variables on the vertical axis.

- *Income per capita (GDP/ capita, inflation-adjusted \$)*
- *Births and deaths > Population growth (annual %)*
- *Children per woman (total fertility)*
- *Geography and Infrastructure > Agricultural land (% of land area)*
- *Energy > Oil > Oil consumption (thousands of barrels daily)*
- *Energy > Oil > Oil consumption per person (barrels yearly)*
- *Environment > Emission > Total CO₂ emissions (thousand tonnes)*
- *CO₂ emissions (tonnes per person)*

Circles of different colors and sizes are shown in the graph. Explain what the size and the color of a circle means. Use the **Play** button to see how each of the variables changes over time. Hold your cursor over geographic regions. Explore what happens when you change the axes from **lin** to **log**. Pick six countries, including the U.S., China, and India. Build a data set for each of the variables for the year 2005 in the spreadsheet under the *Gapminder* tab.

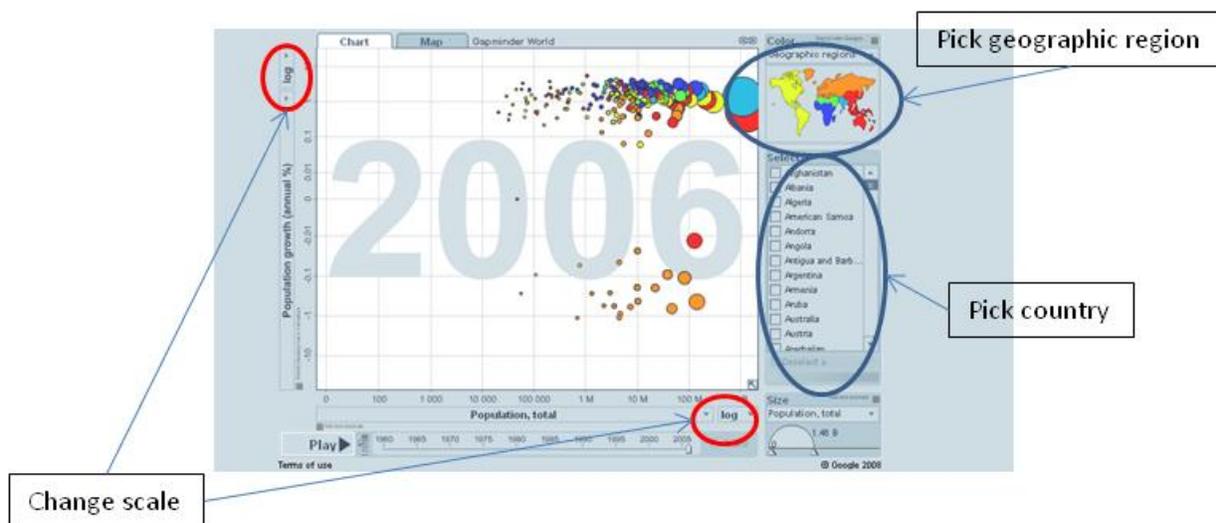


Figure 1: Screenshot of Gapminder

In-class Activity 2

Explore whether there are any relationships among the indicators you investigated in In-class Activity 1. Feel free to add other indicators. Record your observations in the spreadsheet under the *Gapminder* tab.

Logarithmic Transformations

Campbell (1996), page 1109, shows a graph that illustrates the decrease in fecundity at high population densities in a small herb, plantain (*Plantago major*). The data can be found in the spreadsheet under the Plantain tab. On the horizontal axis is the number of seeds planted per m², on the vertical axis is the average number of seeds per reproducing individual (Figure 2). We make two observations: (1) The axes are on a scale where multiples of 10 are equidistant. (2) The points seem to follow a straight line.

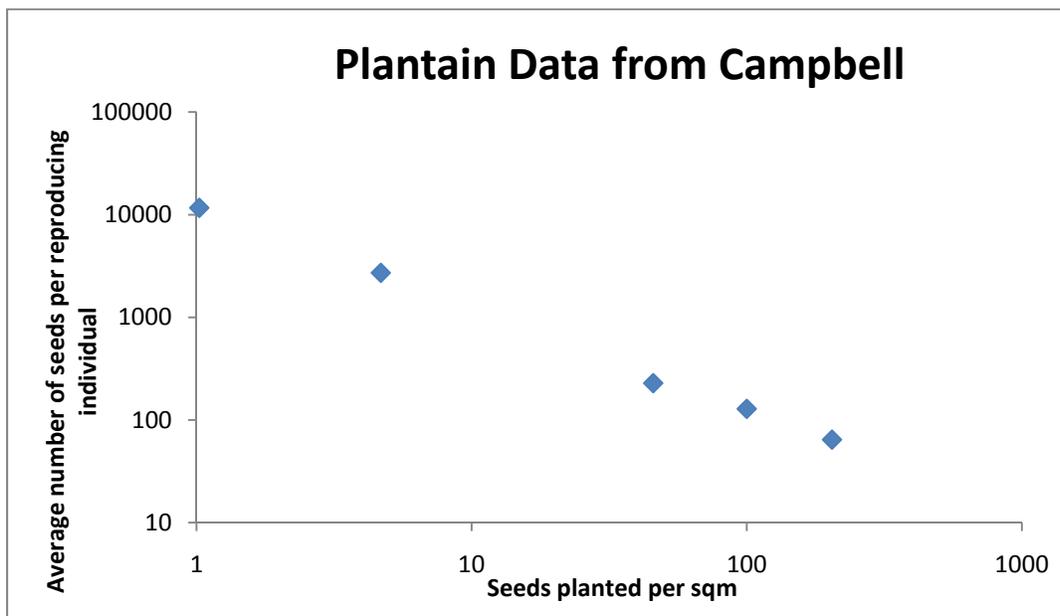


Figure 2: Data points on a graph where both axes are logarithmically transformed. (Drawn after Campbell, 1996¹.)

On either axis, the numbers span several orders of magnitude. Graphing the points in this way makes it easier to see the relationship since the data almost follow a straight line on this graph.

Data from the Christmas Bird Count was used in Van Bael and Pruett-Jones² as a proxy for population size. The data can be found in the spreadsheet under the Parakeet tab. If we use a logarithmic scale on the vertical axis, then a straight line fits the data points. The graph is shown in Figure 3.

¹ Campbell, N.A. (1996) *Biology*. Fourth Edition. The Benjamin/Cummings Publishing Company, Inc.

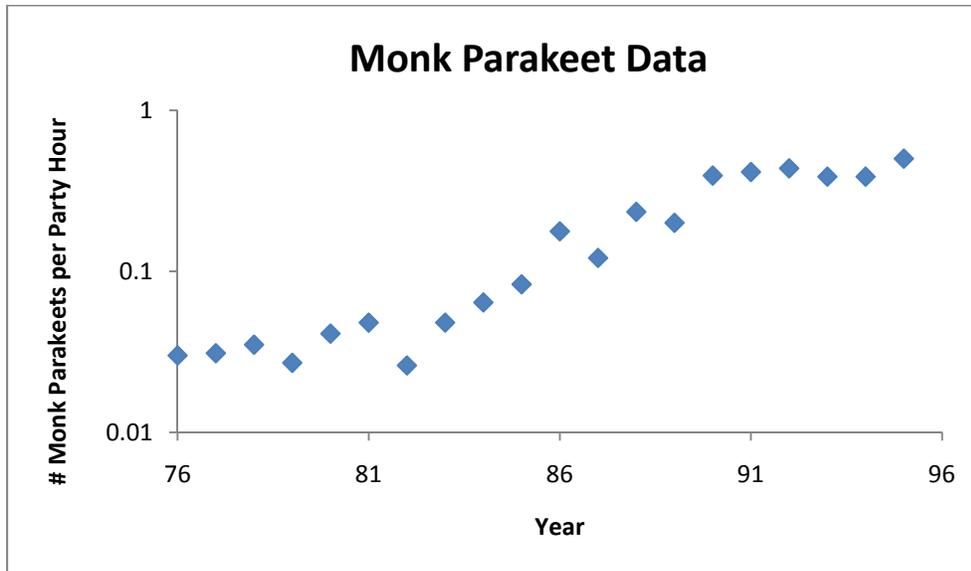
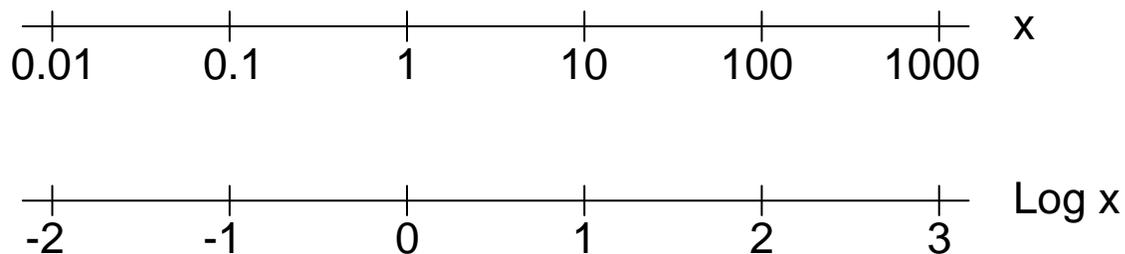


Figure 3: Data points on a graph where the vertical axis is logarithmically transformed.

We will learn what functions are used to describe linear relationships after the data is transformed logarithmically. The first step in understanding these relationships is to gain familiarity with logarithms to base 10.

The Logarithmic Scale

A scale where multiples of 10 are equidistant as in the graph above is called a *logarithmic scale*. It is called logarithmic since the logarithms of the labels on the axis below are equidistant (here: $\log x$ or $\text{Log } x$ means the logarithm to base 10):



² Van Bael, S. and S. Pruett-Jones. 1996. Exponential population growth on Monk Parakeets in the United States. *Wilson Bulletin* 108(3):584-588.

In-class Activity 3

(a) On the two axes above find the following numbers: $x=0.05, 0.2, 8, 15, 750$. (b) Why do you think we choose logarithms to base 10, instead of some other base? (c) Can you plot negative numbers on a logarithmic scale? (d) As x approaches 0, where would you find x on a logarithmic scale?

The two most frequent transformations of a relationship $y = f(x)$ are (1) both axes are logarithmically transformed or (2) the y -axis is logarithmically transformed and the x -axis is on an arithmetic (linear) scale. In either case, when such a transformation results in a straight line, we can find the analytical form of the relationship.

Case 1: Both axes are logarithmically transformed

If the relationship between $\log y$ and $\log x$ is linear, we can write

$$\log y = B + a \log x$$

where B is the intercept on the vertical axis and a is the slope.

$$\log y = B + a \log x$$

$$y = 10^{[B+a \log x]}$$

$$y = 10^B \underbrace{10^{a \log x}}_{10^{\log x^a} = x^a}$$

With $b = 10^B$, we can now write

$$y = bx^a$$

This is a power function. We can summarize this result.

If both axes are logarithmically transformed and a straight line results, then the relationship between x and y is a **power function**: $y = bx^a$

Case 2: The x -axis is on an arithmetic scale and the y -axis is logarithmically transformed

If the relationship between $\log y$ and x is linear, we can write

$$\log y = C + mx$$

where C is the intercept on the vertical axis and m is the slope.

$$\log y = C + mx$$

$$y = 10^{[C+mx]}$$

$$y = 10^C \underbrace{10^{mx}}_{(10^m)^x}$$

With $c = 10^C$ and $a = 10^m$, we can now write

$$y = ca^x$$

This is an exponential function. We can summarize this result.

If the x -axis is on an arithmetic scale and the y -axis is logarithmically transformed and a straight line results, then the relationship between x and y is an **exponential function**:

$$y = ca^x$$

In-class Activity 4a

Graph each of the following functions in Excel and transform the axes so that a straight line results. Use the Trendline option to fit a straight line. Display the equation on the chart.

(a) $y = 2x^3$

(b) $y = 2.5 \times 3^x$

In-class Activity 4b

Fit functions to both the Plantain and Parakeet data in the spreadsheet.

Data on Population Growth

The U.S. Census Bureau keeps track of the world population, calculates annual growth rates, and uses models to predict future population sizes. For instance, the total midyear population for the world between 1950 and 2050 can be found at <http://www.census.gov/ipc/www/idb/>. This data is also

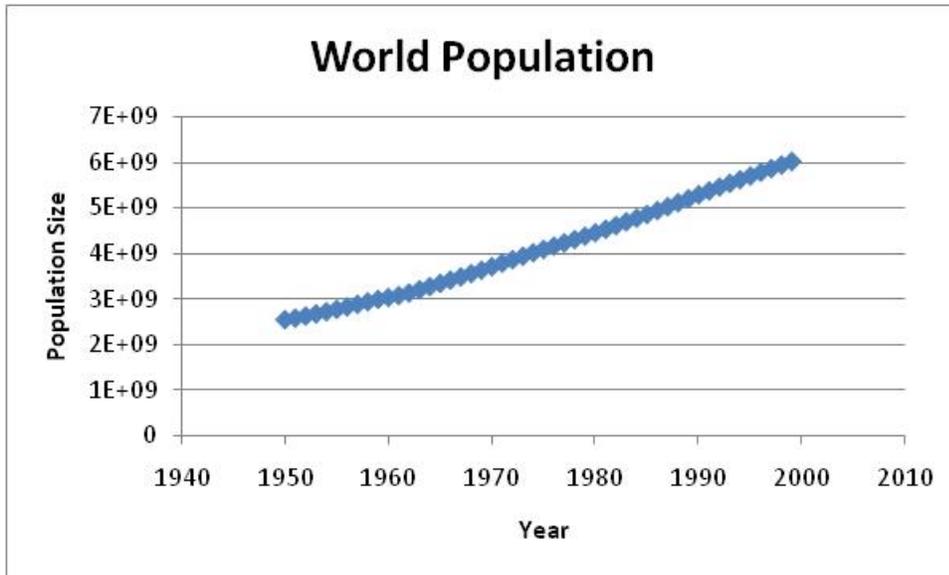


Figure 4: World population between 1950 and 1999. Data from U.S. Census Bureau.

available under the World Population tab in the accompanying spreadsheet (labeled World Population—the sheet is protected to avoid accidental overwriting of data). The population size between 1950 and 1999 is shown in Figure 4. We see a steady increase.

In-class Activity 5

How does the population change from year to year? Let's build a very simple model where the population size at time $t + 1$ is a multiple of the population size at time t

$$(1) \quad N(t + 1) = R(t)N(t)$$

In your spreadsheet under the *WorldPopulationWorksheet* tab, calculate the function $R(t)$ for the years 1950 to 2005. Set the year 1950 equal to 0 and plot $R(t)$. Transform the vertical axis logarithmically. Compare the two plots.

The **average annual per capita growth rates** in Column C of the spreadsheet are calculated using the formula

$$(2) \quad r(t) = \ln \frac{N(t+1)}{N(t)}$$

where t denotes the year, $r(t)$ the growth rate from midyear t to midyear $t+1$, $N(t)$ the population at midyear t , and \ln the natural logarithm. The relationship between $R(t)$ in Equation (1) and $r(t)$ in Equation (2) is as follows

$$r(t) = \ln R(t)$$

In-class Activity 6

1. Use Equation (2) to calculate the annual growth rate in percent from midyear t to midyear $t+1$ for the world population for $t = 1990, 1991, 1992, 1993$, and 1994:

| Year | Midyear Population | Annual Growth Rate (%) |
|------|--------------------|------------------------|
| 1990 | 5,282,371,928 | |
| 1991 | 5,365,708,797 | |
| 1992 | 5,448,725,522 | |
| 1993 | 5,529,987,425 | |
| 1994 | 5,610,065,463 | |
| 1995 | 5,690,982,026 | |

Compare your results with the annual growth rate that the U.S. Census Bureau calculated.

2. The world population reached about 3 billion in 1959. How many years did it take the human world population to reach 6 billion, 9 billion? What is the percent increase during each of these time intervals?

3. Suppose a population doubles in size within a single year. What is the percent increase during that year? What is the average annual growth rate in % during that year based on Equation (2)?

In-class Activity 7

Assume that a population size follows the following equation

$$(3) \quad N(t) = N(0)e^{at}$$

where the parameter a is the per capita growth rate. Assume that the population has a constant per capita growth rate of $a = 1.4\%$.

1. Use Equation (3) to calculate how many years it will take a population to double in size?

2. To calculate the doubling time of a growing population with constant per capita growth rate, you divide the percent growth rate into 70. Apply this rule to check your answer in In-class Activity 7, Problem (1). Use Equation (3) to explain where this "rule of 70" comes from.

Back to Data

When looking at the growth rate of the world population between 1950 and 1999 (Figure 5), we see that the annual growth rate varies from year to year. We can make two observations for the time period

between 1950 and 1999: (1) The world population increased every year during this time interval; and (2) there are periods during this time interval when the rate of growth is increasing and when it is decreasing.

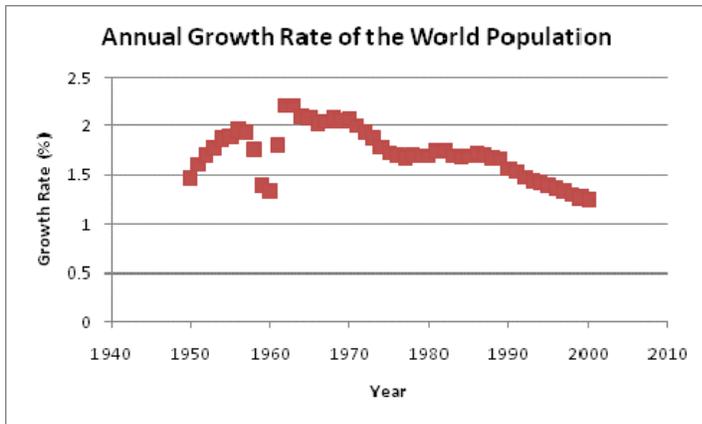


Figure 5: Annual growth rate between 1950 and 1999 of the World Population. (Source: U.S. Bureau of Census)

In-class Activity 8

1. As long as the growth rate is positive, the population size increases, regardless of whether the growth rate is increasing or decreasing. How does the shape of the graph of a population size versus time change when the population is increasing at an increasing rate and increasing at a decreasing rate?
2. How can you extend this to populations that are decreasing in size?
3. Can you come up with other examples where a quantity is increasing/decreasing at an increasing/decreasing rate?

| | Increasing rate | Decreasing rate |
|---------------------|-----------------|-----------------|
| Increasing quantity | | |
| Decreasing quantity | | |

The dynamics of the human population are quite complex. Economic and political factors play a significant role and are difficult to model. For instance, the dip in the annual growth rate from 1959-1960 was due to the Great Leap Forward in China where a combination of natural disasters and decreased agricultural output during social upheavals caused a sharp rise in death rates and a decline in its birth rate.

In-class Activity 9

We will use the linear regression tool in Excel and the annual average growth rate between 1950 and 2005 to predict future growth rates between 2006 and 2050. The first step is to pick a time interval that you wish to use for predicting future growth. How did you choose?

The next step is to plot the growth rate versus year as a scatter plot. Set your starting year equal to 0. Use the linear regression tool to fit a straight line and display the equation of the straight line. The equation will be of the form

$$(4) \quad r(t) = mt + b$$

Use Equation (4) to predict growth rates for each of the years between 2006 and 2049.

Since

$$(5) \quad r(t) = \ln \frac{N(t+1)}{N(t)}$$

It follows that

$$(6) \quad N(t + 1) = N(t)e^{r(t)}$$

Which can be used to iteratively predict the population sizes in the future. What is your prediction for 2050? How does it compare to the predictions of the U.S. Census Bureau?

In-class Activity 10

Return to the six countries you picked in In-class Activity 1. Assuming that the population in 2050 is predicted to be about 9.5 billion, what impact on sustainability would this have if the entire world adopted the standard of living of each of the six countries?

Appendix: Where Does the U.S. Census Bureau Formula Come From?

In populations with overlapping generations, we model the population size in continuous time. The modeling framework is **differential equations**. If $N(t)$ denotes the population size at time t , then the change in population size during a time interval of length Δt , say between time t and time $t + \Delta t$ is given by $\Delta N = N(t + \Delta t) - N(t)$. The **average growth rate** during this time interval is then calculated as

$$(7) \quad \text{average growth rate} = \frac{\Delta N}{\Delta t} = \frac{N(t + \Delta t) - N(t)}{\Delta t}$$

A population with overlapping generations is continually changing. To describe this change, we will calculate the **instantaneous rate of change**, that is, the rate of change during an **infinitesimal time interval**. To calculate this growth rate we take the limit as $\Delta t \rightarrow 0$:

$$(8) \quad \text{instantaneous growth rate} = \frac{dN}{dt} = \lim_{\Delta t \rightarrow 0} \frac{N(t + \Delta t) - N(t)}{\Delta t}$$

You recognize this as the derivative of the population size $N(t)$ at time t . The derivative $\frac{dN}{dt}$ is the **population growth rate**. The **per capita growth rate** at time t is obtained by dividing the population growth rate by the population size

$$\text{per capita growth rate} = \frac{1}{N} \frac{dN}{dt}$$

Our first model says that the per capita growth rate is equal to a function, which we denote by $r(t)$. We thus have as our first model

$$(9) \quad \frac{1}{N(t)} \frac{dN}{dt} = r(t)$$

Let's assume for the moment that the per capita growth rate is constant and let's denote it by a , that is, $r(t) = a$ for all t , then one can show using integral calculus that the solution of Equation (9) is

$$(10) \quad N(t) = N(0)e^{at}$$

where $N(0)$ denotes the population size at time 0. We can confirm this by differentiating Equation (10). We find

$$\frac{dN}{dt} = a \underbrace{N(0)e^{at}}_{N(t)} = aN(t) \quad \text{and} \quad N(0) = N(0) \underbrace{e^{(a)(0)}}_{=1}$$

We can now check why Equation (2) holds when the per capita growth rate is constant. Namely, looking at Equation (10), we see that

$$N(t+1) = N(0)e^{a(t+1)} \quad \text{and} \quad N(t) = N(0)e^{at}$$

We find for the ratio $\frac{N(t+1)}{N(t)}$

$$\frac{N(t+1)}{N(t)} = \frac{N(0)e^{a(t+1)}}{N(0)e^{at}} = e^{a(t+1)-at} = e^{at+a-at} = e^a$$

Taking natural logarithms on both sides, we obtain

$$\ln \frac{N(t+1)}{N(t)} = a$$

This is Equation (2) in the case when the per capita growth rate is constant. To understand why Equation (2) can be used to estimate the per capita growth rate regardless of whether the per capita growth rate is constant, we need to be able to solve a differential equation by integration. [If you have not covered this in calculus, you can skip over the following calculation for now and go directly to Equation (11).] If the reproductive value or the per capita growth rate varies over time, the per capita growth rate can be approximated by solving the differential equation in Equation (9) from one time step to the next. Separating variables in Equation (9) yields

$$\frac{dN}{N} = r(t)dt$$

We will integrate this over the time interval $[t, t+1]$. Since the antiderivative of $1/x$ is $\ln x + C$, we find

$$\ln N(t+1) - \ln N(t) = \int_t^{t+1} r(s)ds$$

The left-hand side can be simplified using the rule $\ln x - \ln y = \ln \frac{x}{y}$. The right-hand side cannot be simplified without knowing more about the function $r(t)$. However, the right-hand side can be interpreted as the average value of the function $r(t)$ over the time interval $[t, t+1]$ using the formula of calculating the average value of a continuous function $f(x)$ over the interval $[a, b]$

$$f_{avg} = \frac{1}{b-a} \int_a^b f(x)dx$$

We therefore find that

$$(11) \quad \text{average per capita growth rate in } [t, t+1] = \ln \frac{N(t+1)}{N(t)}$$

The formula in Equation (11) is the one used by the U.S. Census Bureau. It calculates the average per capita growth rate of the population for the time interval $[t, t+1]$. This is not the same as the instantaneous growth rate $r(t)$. In other words, when we use Equation (2), we find approximations for the instantaneous per capita growth rate $r(t)$ at time t by approximating $r(t)$ by the average per capita growth rate over the time interval $[t, t+1]$.

Additional Resources

Design of Scientific Posters:

<http://www.hhmi.org/coolscience/resources/SPT--FullRecord.php?ResourceId=29>

Statistics for Journalists:

<http://www.robertniles.com/stats/>

Future Explorations

Data used in Gapminder can be downloaded from Google (see List of Indicators and download the “indicator list” to Excel). The following spreadsheets are included in the Resources:

OilConsumptionGapminder.xls

OilConsumptionPersonGapminder.xls

PopulationGrowthGapminder.xls

TotalCO2EmissionGapminder.xls

TotalIncomeGapminder.xls