Lab Manual for Calculus: A Modeling Approach for the Life Sciences

Joseph M. Mahaffy Alexandra Chávez-Ross

August 2006

Contents

Chapter 0

Preface

This lab manual has been developed to accompany the text *Calculus: A Modeling Approach for the Life Sciences*, which the authors created to give a modern and biologically relevant view of Calculus. This lab manual is designed to provided biologically interesting examples and teach students how to use several software packages. In particular, the primary computer tools are the spreadsheet software Excel, which has excellent graphing capabilities and several valuable data analysis tools, and the mathematical symbolic software Maple, which has excellent algebraic and Calculus related abilities that the students need to learn about.

The course was originally developed at San Diego State University when the Mathematics and Computer Sciences departments were together. The Calculus for the biology majors was being taught from a Business Calculus text, and the students were not appreciating the importance of mathematics to modern biology. Furthermore, the biology faculty at San Diego State University were frustrated by the lack of mathematical skills emerging from students who had taken the two semester sequence. Professor Mahaffy began by changing texts and introducing the students to computer projects that expanded on the material in class with the able assistance of Professor Roger Whitney from the Computer Science Department. The class began with surplus computers leftover from Professor Whitney's NSF equipment grant that were seriously underpowered (Mac Pluses). Still the initial projects proved to create interest by the students that had not been seen before in this class. Thus, SDSU's College of Sciences helped us develop a better lab with improved equipment, which got other faculty in Mathematics interested in using the computer labs to aid in the instruction of their classes. The result is that we now have excellent support for our computer lab, and an excellent environment for student learning.

The available texts for this course continued to be weak and dated, so Professor Mahaffy developed a website with extensive lecture notes and continued to improve the computer labs to go along with the material on the web. Recently, the lecture material was developed into preliminary editions by Prentice Special publications. However, the course was developed as an integrated course that taught both skills in Calculus and techniques of the computer, which students found particularly useful in their upper division biology courses. This three unit course has evolved to where the students have two hours of lecture each week on the material in the text and two hours of computer lab (a one unit activity). The result has been that students have developed a much better appreciation of the role of mathematics in biology and have learned valuable computer skills that help them solve quantitative problems in biology. They no longer feel that this course is useless to them as when they were learning from the Business Calculus text.

This lab manual will parallel the material in the text *Calculus: A Modeling Approach for the Life Sciences*, but material in this lab manual would be valuable to any instructor who wants to teach mathematical modeling to biology students and should provide valuable examples for other faculty teaching mathematical modeling. The text *Calculus: A Modeling Approach for the Life Sciences* is based on a dynamical systems approach to learning Calculus, but it also stresses the importance of using real data with the help of computers to appreciate the relevance.

Joseph M. Mahaffy
Alexandra Chávez-Ross
August 2006

Chapter 1

Introduction

The biological sciences are becoming more quantitative as research is expanding. Mathematics is playing a significant role in understanding the complex relationships that are being discovered in biology. The advent of computers has made many of the complex problems tractable and has greatly facilitated the analysis of large sets of data. Modern students of biology need to know more about mathematical models and computers to better understand what are the abilities and limitations of the theoretical models for biology.

We have developed an integrated course of Calculus for the life sciences that emphasizes mathematical models from biology and uses computers to better understand the concepts taught in the course. This Computer Laboratory Manual has been developed to accompany the text *Calculus: A Modeling Approach for the Life Sciences* by the authors of this manual[1]. However, the material for the computer labs developed here can be used as supplemental material for a variety of courses in mathematical modeling for biology. This manual is designed to provide numerous biological problems (along with a few purely mathematical problems from Calculus) that should help the student learn more about both computer techniques and mathematical models.

1.1 Background for the Computer Laboratory

Most of the problems are designed to illustrate some biological idea using real biological data. The computer lab problems were originally designed for students to work in small groups (usually two), but can be adapted to other schemes. Based on consultation with faculty in Biology, we developed the computer labs to have mathematical, biological, and writing components. Thus, these labs have been written with the idea that a biological problem that extends from the lecture material in the text would be analyzed using the spreadsheet software, Excel, or the symbolic mathematical software, Maple. Then the students have been required to write a formal lab report, using complete sentences with the software Word, which nicely imports the graphic material from either software package (and are parts of almost every exercise) and can create good tables that summarize information acquired from the analysis. Many of the exercises ask the students to think a little deeper about the biological implications of the study, then write a cogent paragraph about their results.

1.2 Graphing in Excel - A Sample Problem

Below is an example of a quadratic function. Details are provided to guide you through key steps for creating good Excel graphs for your lab report, showing many of the features that are required in the graphs. In the process, we will demonstrate a couple of valuable techniques in Word to make the lab write-up look better.

Example 1: Consider the quadratic equation given by the function

$$f(x) = 2 + 2x - x^2.$$

Find the x and y-intercepts and the vertex of the parabola. Graph $f(x)$ for $-3 \le x \le 3$. Label the function and the vertex on the graph.

Creating a Graph with Excel

Excel is a spreadsheet software package, so it is not ideally set up to create graphs (which will be contrasted later when we show graphing methods in Maple). However, it is excellent for adding labels and other design features, which makes it a valuable tool for creating good presentations of data and functions. To create a good graph in Excel, the user needs to create a data base of the x and y values that will be graphed. The rule of thumb for a good graph in Excel is to have 40-60 points on the graph. (Below we will provide information to create a general *Graphing Template* that always uses 50 points and can be used for most graphing situations, where one is asked to graph a function in Excel.)

In the example above, the user is asked to graph $f(x)$ for $-3 \leq x \leq 3$, so we begin in Excel by creating a table of numbers from $x = -3$ to $x = 3$. Since the difference between the x values is 6, then if we choose an increment size of 0.1, there will be 60 points for our function evaluations. For this example, the list of x values are placed in the first column of the Excel spreadsheet. We use the first cell, **A1**, as a location for the label x. In the cell below, **A2**, we put the x value of the left endpoint of the interval. In the next cell below, **A3**, we type **= A2 + 0.1**, which adds 0.1 to -3, giving -2.9. Note that you can type =, then click on **A2**, then continue typing **+ 0.1** to get the same result. From **A3**, we fill down until we reach $x = 3$ at **A62**. To fill down, you can highlight the elements in the spreadsheet that you want to fill down. You highlight using the cursor (thick white cross) and by holding onto the left button of the mouse from cell **A2**, pull the cursor down along column **A**, highlighting until the cell **A62**. Follow this by either typing **Ctrl-D** or going to the **Edit** item on the menu, finding fill, and selecting down. Alternatively (and simpler), you highlight the cell that you want to fill down from. You move the cursor to the lower right corner until it changes to a simple +. At this point, you move the mouse down to the cell that you want to fill to (**A62**).

To make your formulae easier to understand, we want to take advantage of the variable naming feature of Excel. Begin by highlighting all the cells from **A1** to **A62**, which is done by holding the left button of the mouse, while you move through all the desired cells. Next we click on the **Insert** label in the toolbar at the top of the spreadsheet. Choose the option **Name** followed by the option **Create**, and accept the default setting of **Create names in Top Row**. This gives all the cells in the first column the label of x, which will make reading the formula for our function easier to understand.

The next step in the process is to create the y values for our graph from the function $f(x)$. We use the first cell in the second column, **B1**, as a location for the label y. In the cell below, **B2**, we put the y value in by evaluating our function $f(x)$. This is done by typing **= 2 + 2*x - x^ 2** in that cell. Note that we need to tell Excel to multiply the 2 and the x value using the symbol *, and Excel uses **x^ 2** to represent x^2. Next you fill down Column B from **B2** to **B62**, following the same fill down procedure as noted above for Column A.

Now we want to graph **Columns A** and **B**. First you highlight the data to be graphed (you can do this very quickly by just clicking on the letter **A** of **Column A** and holding the Ctrl button click on the letter **B** of **Column B**). Then you click on the icon for **Chart Wizard** on the toolbar (if the icon is not shown, look for it in the **Insert** menu on the toolbar.) Select the **XY (Scatter)** chart type of graph, then select the chart sub-type in the lower right corner (straight lines). (For this text, we exclusively use the **XY (Scatter)** chart type, and we almost always select either the default (for data) of points only or the lower right corner (for lines/functions) and ignore all other sub-types. Proceed with **Next** twice to get to the page (Step 3) where you begin by entering a **Title** and **Label the axes**. For this example, the title given is **Quadratic Function** with simple **x** and **y** labels for the axes. Now select **Gridlines** and check **Major** for the x-axis, then select **Legend** and unselect **Show legend**, as there is only the one graph. At this point, you can select **Finish**. The result appears as follows.

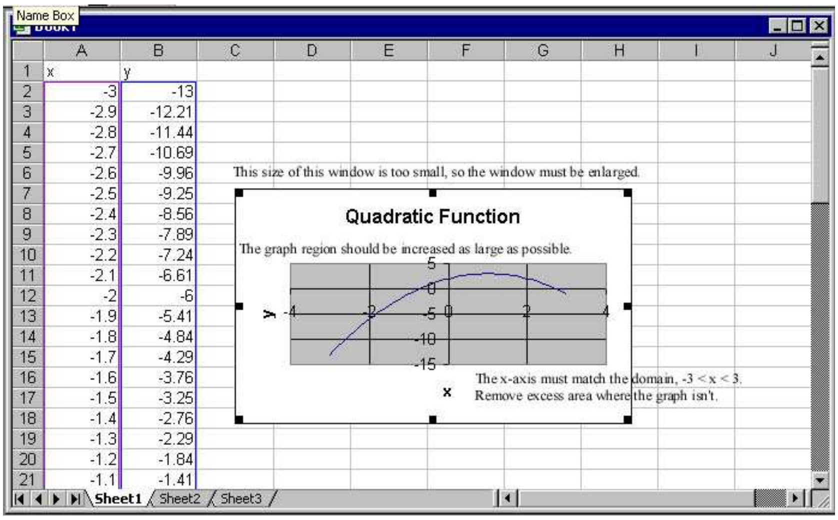


Figure 1.1: Initial Excel plot for graphing a quadratic.

This is clearly a poor looking graph. There are three comments that have been entered showing where we need to improve the graph. First the window is increased by moving the mouse to black squares and using the left button of the mouse to increase the viewing frame (about 8 columns and 19 rows is a fairly good size). To adjust the x-axis to the correct domain, you double click (or use the right button of the mouse) when you point to the x-axis and see the label **Value (X) axis**. In the window that pops up, you select Scale and change the **Minimum** to -3, the **Maximum** to 3, and the **Major unit** to 1. Finally, we need to make the actual area of the graph (grey region) larger. The fonts for the titles and axis labels are oversized, so we change the fonts on all of these. To change these we simply point with the mouse, then either double click or use the right button of the mouse. The title works well with a font of 12 or 14, usually bold, while the axis labels and axis numbers are best using either regular or italic fonts with size of 10 or 12.

x	y
-3	-13
-2.9	-12.21
-2.8	-11.44
-2.7	-10.69
-2.6	-9.96
-2.5	-9.25
-2.4	-8.56
-2.3	-7.89
-2.2	-7.24
-2.1	-6.61
-2	-6
-1.9	-5.41
-1.8	-4.84
-1.7	-4.29
-1.6	-3.76
-1.5	-3.25
-1.4	-2.76
-1.3	-2.29
-1.2	-1.84
-1.1	-1.41

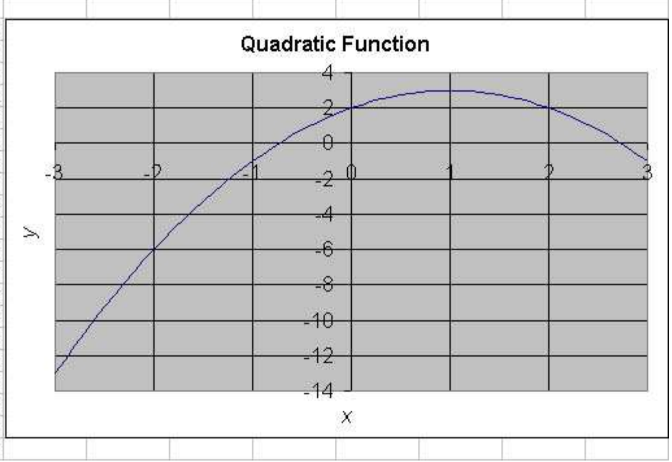


Figure 1.2: Adding labels to the graph of the quadratic.

Before finishing the graph by adding the required labels, we need to complete the algebra to obtain the intercepts and the vertex. The y-intercept is very easily found by simply evaluating $f(0) = 2$, so the y-intercept is $(0, 2)$. The x-intercepts are found by solving $f(x) = 0$, which is equivalent to solving

$$x^2 - 2x - 2 = 0.$$

This is solved using the quadratic equation (or completing the square if you prefer this technique.) You may refer to the *Function Review and Quadratics* chapter of the textbook for some examples in how to solve quadratic equations [1]. Later we will show how the symbolic mathematics software Maple easily does this computation for you. The quadratic formula gives

$$x = 1 \pm \sqrt{3}.$$

The x-coordinate of the vertex is halfway between the x-intercepts, so the x-coordinate of the vertex is 1. Since $f(1) = 3$, it follows that the vertex occurs at $(1, 3)$. (Section 3 of Chapter 4 of the text discusses some of the techniques for finding the vertex of a parabola.)

We now return to the graph in Excel to add the labels and make other changes to improve the look of our graph. First, to make the graph easier to see, we double click (it will say **Plot Area**) on the background and change its color (usually to a lighter one, though you might want one of the patterns or even import a relevant picture into the background). Next we make the function easier to see by double clicking on the function or click the right button of the mouse so you will get the **Format Data Series** menu. On the **Patterns** option choose the next thicker type from the **Weight** window, and we choose usually a dark color. Because the default graph leaves little room to label the vertex, we change the scale on the y-axis by double clicking on the y-axis and changing the scale in this case from $y = -15$ to $y = 5$ in increments of 5.

Adding Labels to the Graph

First, make sure that the graph is activated by seeing the black squares at the corners and in the middle of all edges. (If this is not the case, then clicking on the white area of the graph activates the graph.) We begin by adding the label for the vertex of this parabola. Click in the white area labeled f_x that appears just above the column labels. (Not all Excel versions have the f_x label, but instead you should have the = sign on the left hand side of the window). In this region, we type **Vertex: (1, 3)**, then Enter the result. The label will appear in the middle of the graph. For the label shown on the graph in Figure 1.3, the font was changed to Times New Roman (the most common used in Math) and increased its size to 14 points. The box had to be resized to fit this font and larger size by dragging one of the corners with the mouse. Use the mouse to position the box near the vertex, then highlight the text you typed and right click on the the box and select **Format Text Box**. The next menu item you choose is **Color and Lines** from which you go to the the **Color** entry and select the same color as you selected for the graph background. This allows your label to block out the gridlines and be more visible.

For labeling the graph, you need to type in the equation in the white area labeled f_x (or on top of the column labels). For this function, we typed **y = 2 + 2x - x2**. As in the label discussed above, the font was changed to 14 point Times New Roman and the background was made to match the graph background. To finish the equation, we highlighted each of the variables (individually) and made them italic by clicking on the *I* on the menu bar. Finally, to get the x^2, we highlighted the final 2 in the equation, then went to the main menu bar under **Format**, then accepted **Selected Object** and on the submenu checked the box for **Superscript** under the category **Effects**. Alternately, we highlight the final 2 in the text box, right click on it and select **Format Text Box again**, then check on the **Superscript** white square. In Figure 1.3 you may see the final result of the graph created, which is how it should appear in your report.

It is important to note that you should be saving your work at regular intervals so that you don't need to start your work from scratch. It also protects you from possible system problems so that you don't lose everything you've worked hard to obtain. Now we are ready to create the actual Lab Report for **Example 1**.

Creating the Lab Report with Word

One objective of the Computer Laboratory is to teach students to learn about technical writing. Thus, the answers to the lab questions should be typed in a Word document in clear, concise, and complete sentences. Above we have done all the calculations needed to answer the questions and have created a good graph to include in our document. We just need to add that once you have copied and pasted your final version of the graph created in Excel into your Word document, you should change the size of the graph to be about 3 inches in height height and 4 inches in width (the actual graph, not the window). This means **DO NOT** to use the default size in Excel. The graph should be clear and readable in your report document to demonstrate the results being exhibited in the problem.

There remains one additional task to learn to be able to create a good document. The x-intercepts are $1 \pm \sqrt{3}$, which is not a standard form on the typical computer keyboard.

Adding Equations to a Document

In Word (assuming you don't have an abbreviated installation), there exists a special program to create mathematical equations. When you are in Word and are where you want to enter a mathematical expression, you click on **Insert** on the top menu, followed by **Object**, and then choose **Microsoft Equation 3.0** (or alternatively it might say **Mathtype** or **Equation Editor**). A second window will open up with symbols which you can insert. To obtain the right x-intercept, for example, you type **(1+** followed by the **square root icon** and a **3**, then the **right arrow key** to indicate that the next part is no longer inside the square root. Follow this with **,0)** and press

Escape. The result will be $(1+\sqrt{3}, 0)$.

Below is an example of how this question might appear in a Word document, using the equation capabilities of Word and copying and pasting the Excel graph into the document.

Example 1: Consider the quadratic equation given by the function

$$f(x) = 2 + 2x - x^2.$$

Find the x and y-intercepts and the vertex of the parabola. Graph $f(x)$ for $-3 \leq x \leq 3$. Label the function and the vertex on the graph.

Solution: The x-intercepts occur at $(1-\sqrt{3}, 0)$ and $(1+\sqrt{3}, 0)$. The y-intercept is at $(0, 2)$, while the vertex is located at $(1, 3)$. Below is the graph of $f(x)$.

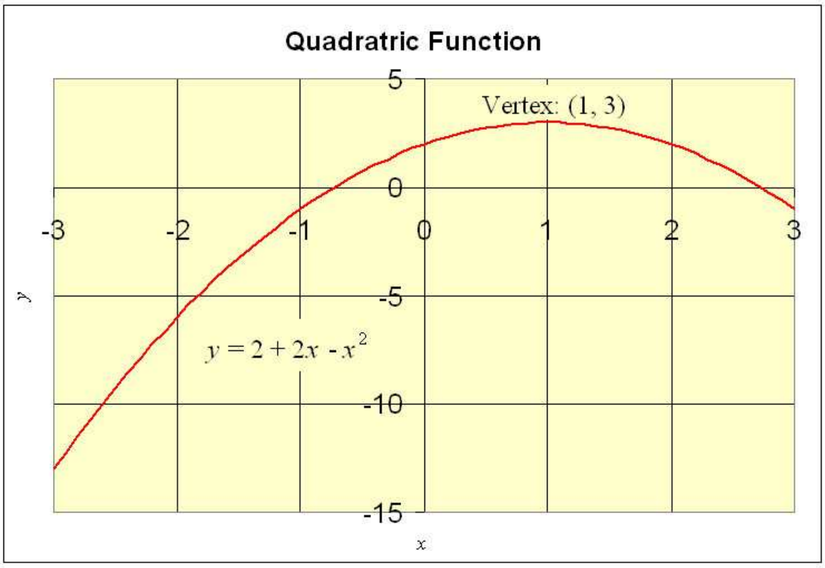


Figure 1.3: Graph of $f(x) = 2 + 2x - x^2$.

The above answer is an example of a complete solution to this example. Be sure to **Save** your file at this point!

1.3 Creating a Generic Graphing Template

Any calculus course requires the creation of many graphs. In fact, Calculus itself will give you new techniques to sketch the graph of a function. Refer to the *Applications of the Derivative-Graphing* chapter of the text book to learn how [1]. This computer lab manual has many problems that require graphing, so below we present an algorithm for creating a graphing template for a "nice" function, which requires that the user only has to input the domain and the function, then it is ready to graph in Excel. You are not supposed to follow all the procedure described in Section 1.2 every time you need to make the graph of a function. Moreover, by a "nice" function, we mean a smooth graph for which you will need enough points to be plotted. However, the student will still need to do many of the procedures outlined above to make the graph look good.

Consider a function $f(x)$ that we want to graph on the interval $x \in [a, b]$. Below is a detailed procedure to create a graphing template that can be used for any graph by only changing the inputs for a, b, and $f(x)$. It can also be used for graphing multiple functions in Excel. For our starting example, we will choose $a = 0$, $b = 5$, and $f(x) = x^2 - 2x$, but this is only for set up purposes.

1. Open an Excel file and give it a generic name, like **grphtmp.xls**.
2. Put the labels x in **A1**, $f(x)$ in **B1**, a in **G1**, and b in **G2**.
3. Insert the values **0** in **H1** and **5** in **H2**.
4. Highlight the endpoint cells and their values **G1**, **H1**, **G2**, and **H2**, then go to the menu bar and select **Insert**, followed by **Name**, followed by **Create**, and accept the default **Create Names in Left Column**. This will allow you to change your endpoints by simply changing the cells **H1** and **H2**.
5. Go to the cell **A2** and type $= a$ or alternately, type = and click on cell **H1**. You should see the value **0** appear in **A2**.
6. Go to the cell **A3** and type = **A2** + (b - a)/**50**. Again you can simply click on the locations for **A2**, b, and a. The result will be **0.1** appearing in **A3**.
7. Fill down **Column A** until you reach **A52**, which have the value $b = 5$.

8. Highlight **Column A** (**A1** through **A52**), then select **Insert**, followed by **Name**, followed by **Create**, and accept the default **Create Names in Top Row**. This creates your independent variable x and makes 51 evenly spaced points along the domain.

9. In **B2**, type your function, which for this example you would enter = x^ **2 - 2***x.

10. Fill down **Column B** until you reach **B52**, which gives you the corresponding $f(x)$ values.

11. To create a graph, simply highlight **Columns A** and **B**, then click on the **Chart Wizard** icon in the menu bar (the symbol that looks like a yellow, blue, and red bar chart). From there, you follow the usual directions noted in the Example above of selecting the **XY-Scatter Plot**, etc.

Having created this graphing template and saved it, then when you want to create a new graph, you simply make a new copy of this file (renaming it). You open the new file and simply change the entries in **H1** and **H2** to match your new domain endpoints, and change the entry in **B2** to match your new function. You have to **Fill Down** in **Column B** to complete the table, then the new function can be easily graphed. If you have more functions to put on the same graph, then you can enter them in **Column C**, **Column D**, etc. To graph multiple functions, you simply highlight all of the relevant columns, then follow the same procedure as above with **Chart Wizard** to produce the desired graph.

1.4 Graphing a Function and Data Points

In the mathematical modeling of biological problems, biological data is collected, then a mathematical model is often proposed to describe that data. Often one searches for the least squares best fit to the data, then sees if the model describes the data well. Graphically, one wants to display the data and the model, then observe how well they match. Courses in statistics can make provide more quantitative information on the goodness of fit for a mathematical model. In this part of the lab manual, we want to describe how to display both data and a function using Excel. More than just sketching the graphs of functions, we could say this is one of the most relevant features that Excel can provide. Other software, like Maple, helps plot graphs of functions in a more straight forward way.

Below we show data that compares the doublings/hr, denoted μ, to the rate of mRNA synthesis/cell, denoted r_m (see Section 1 of the *Function Review and Quadratics* chapter of the text book for a proper biological introduction of this example [1]). It can be shown that a good mathematical

μ	0.6	1.0	1.5	2.0	2.5
r_m	4.3	9.1	13	19	23

Table 1.1: Data indicating growth of a bacterial cell in terms of μ (in doublings/hr) and the rate of mRNA synthesis, $r_m \times 10^5$ (in nucleotides/min/cell) [2].

model for these data is given by the function

$$r_m = 9.14\mu.$$

This model shows a linear relationship of the doublings/hr with the rate of mRNA synthesis. Note that when there is no growth ($\mu = 0$), then there is no synthesis of mRNA as one would expect.

We begin this graphing exercise by simply plotting the data in the table. In **Column A**, we list the values of μ, and **Column B** contains the data for r_m. We highlight these two columns and use the **XY Scatter Plot** in **Chart Wizard** to graph these data. In this case, we use the default graph type, which produces data points. You may increase the size of the data points so they stand, out by right clicking on either of the points selecting **Format Data Series** and on the **Marker** column increase the default size to 7 points. After going through the menu, adding vertical gridlines, removing the legend, and giving the graph a title and labeling the axes, we see the following graph in Excel.

Next we want to add the theoretical model to the graph. Since it is a straight line, we only need to use two points. However, in Excel, it is usually best to use at least three points when entering a line to avoid having Excel confuse which number pairs with which. (Excel often defaults to using the rows as x and y values, which reverses our use of columns, the more standard way to enter data and produces the wrong graph.) For this problem, we will use **Column D** for the theoretical values of μ, taking the values $\mu = 0, 1, 2, 3$. The corresponding values of r_m are computed in **Column E** ($r_m = 0, 9.14, 18.28, 27.42$). To add these values to the graph,

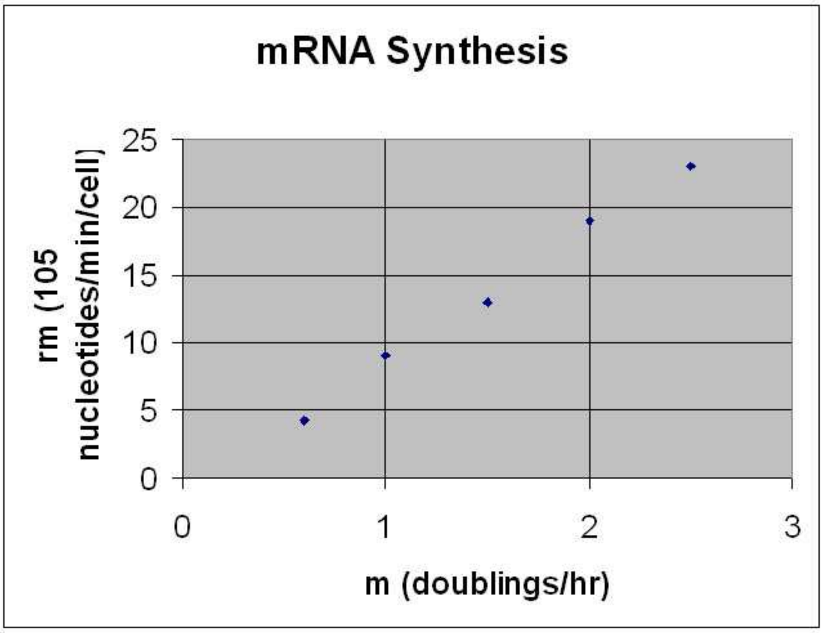


Figure 1.4: Data for mRNA synthesis as a function of the doublings/hr.

we activate the graph by clicking in the white border region of the graph. Then we go to the top menu and click on **Chart** and select the entry **Add Data**. When the box pops up, you simply take the mouse and highlight the desired data in **Columns D** and **E**. Another box will pop up that says **Paste Special**. In this box, it will already have **New Series** and **Values of** Y **in Columns** checked, but you will have to check in addition the box that says **Categories (**X **Values) in First Column**. The result is the following graph in Excel.

At this point the mathematical model consists of data points. This model must be converted to a line. The easiest way to change this into a line is to double click (or right click) on one of the data points on the line. A box entitled **Format Data Series** pops up. In this box you select a color and weight for the line of the model and check **None** under the category of **marker**. This changes the data points into a line as desired. We complete our work to make the graph look good by adjusting the **Scale** of the μ-axis. We change the background color, adjust the font size and style and transform the labels to μ (using the symbol font), r_m (using the subscript option under font from the **Format** menu), and 10^5 (using the superscript option), and make the data points round and larger (by double clicking on

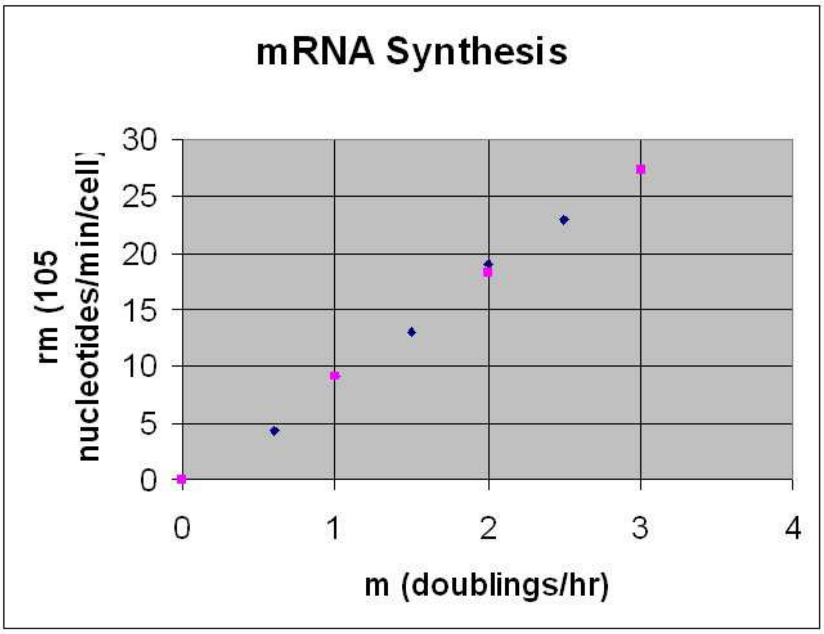


Figure 1.5: Graph of data and straight line for model (as points).

the data points and following the menu). The result is the graph in Excel given in Figure 1.6.

1.5 Computer Laboratory Exercises

Numerous examples and exercises in this lab manual use connections to the web. At some point it is hoped there will be a CD to accompany this lab manual, but for now the reader can consult the website of J. M. Mahaffy. When possible, the specific web address will be given. To simplify listing of these web addresses, we note that the beginning of the web address begins with the following:

`www-rohan.sdsu.edu/~jmahaffy/courses`

which will be given as **...** in the text below. Most of these exercises were produced for a course run from the web. Next to the problem, the reader will find a letter and a number, such as **C3**. The reader can find the original problem by visiting the website:

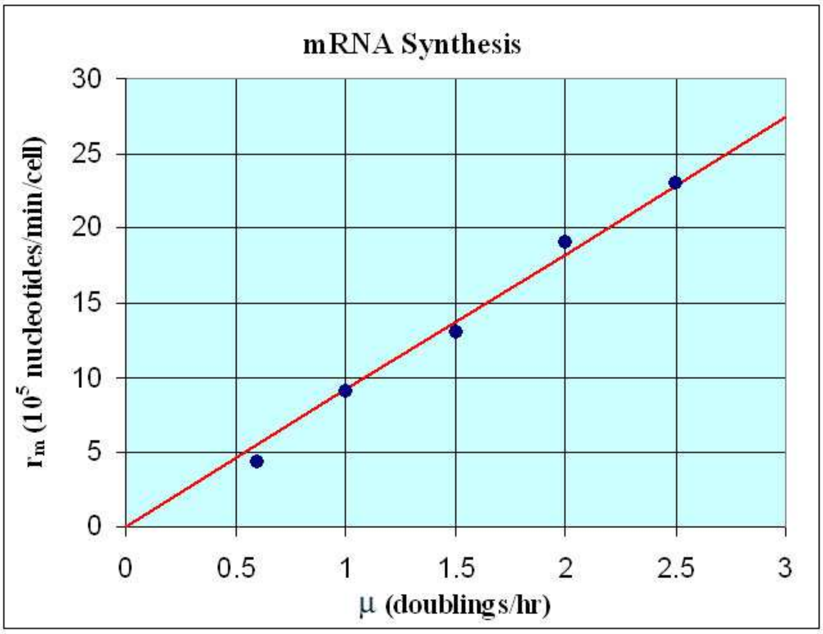


Figure 1.6: Final graph showing the model for mRNA synthesis as a function of the doublings/hr and data points from experiments.

`.../s00a/math121/labs/labc/q3v1.htm`

where the **labc** is from the letter **C**, denoting Lab C, and the **q3v1** is from the number **3**, denoting Question 3 Version 1. (There generally three versions of each problem.)

1. (A1) This question introduces the use of Excel and Word for producing good laboratory documents. This exercise can be started from scratch or one can download a starting graph at

`.../s00a/math121/labs/laba/q1v1.htm`

Consider the two lines given by the formulae

$$y = 2x - 1 \qquad \text{and} \qquad y = -\frac{x}{2} + 2.$$

Graph these two lines in Excel. Generate a title for the graph, label the x and y-axes appropriately, adjust the scale so that the domain (x) and range (y) both go from -4 to 4, and expand the graph to a square. Use Excel's labels to identify each line with its equation. Also, label the coordinates of the point of intersection on the graph. Copy this graph into a Word laboratory document. Write a short paragraph detailing how to find the point of intersection. Are these lines perpendicular? (More information is available in Section 3 of the *Linear Models* [1] for procedures to find the point of intersection and for definitions of parallel and perpendicular lines.) Explain your answer.

2. (A1) Consider the quadratic equation

$$y = \frac{x^2}{6} - \frac{x}{6} - \frac{1}{3}.$$

Use Microsoft Equation or Equation Editor to write this quadratic equation with full-sized fractions (as seen above). Also, write this equation in full factored form, then find the roots of this equation.

3. (A2) a. Consider the following functions, $f(x)$ and $g(x)$:

$$f(x) = 2x - 1 \qquad \text{and} \qquad g(x) = 4 - 2x - x^2.$$

Use Excel to create a single graph of both functions for $x \in [-6, 4]$. Label the graph clearly with the domain adjusted to the given interval.

b. Find the x and y-intercepts for both functions. Give the slope of the line and the vertex of the parabola. Find all points of intersection between the curves. Write this information in a clear paragraph. If the answer uses the quadratic formula and has a square root, then give both the decimal answer to four significant figures and the answer including the square root. (Use Microsoft Equation or Equation Editor to write the answer that includes a square root.)

4. (A3) The text associated with this lab manual gives details supporting the idea that crickets chirping could be used as a type of thermometer, albeit a crude one. That text presented the classic folk "cricket thermometer," formalized by Dolbear, which satisfied the linear relationship:

$$T = \frac{N}{4} + 40,$$

where T and N were the temperature and the number of chirps/minute, respectively. The Bessey brothers later made careful measurements and did a linear least squares best fit to their data and obtained the linear relationship

$$T = 0.21N + 40.4.$$

a. Recordings of four crickets chirping at different temperatures are available at the web address

`.../s00a/math121/labs/laba/q3v1.htm`

In this question, you time the number of chirps/minute of the four crickets. Make a table listing the number of chirps/minute for each of the crickets along with the predicted temperatures from each of the models above.

b. Create a graph of each of the models (one graph with both models), showing clearly the data points that you gathered in Part a.

c. Give the units of the coefficients (slope and intercept) in each of the equations above.

d. Suppose that the error in counting chirps/min is less than 10 chirps/min. Find the range of temperatures for each of the crickets from your data for each of the models, taking into account this source of error. (Thus, if you found $N = 92$ for one cricket, then give the possible temperatures for N ranging between 82 and 102 with each of the models.) Use Word to create a table that gives the range of temperatures that each model gives for each cricket. Write a brief paragraph discussing the accuracies of the models from your lab experience, what are the major sources of error (list at least two), and how much agreement the different models have in predicting the temperature.

5. The thyroid gland in children is susceptible to carcinogenic action from ionizing radiation. The pooled data of many studies suggests that the risk of thyroid cancer is roughly linear with respect to the amount of radiation exposure before the age of 15. This is a typical dose-response that is assumed in many environmental studies of various toxins. More recently, a number of different studies show that our bodies respond to toxins in quite different and very nonlinear ways, but radiation exposure is often best described by this linear dose-response profile.

a. A pooled study [3] of children exposed to radiation from atomic bomb survivors, children treated for tinea capitis, children with irradiated tonsils, and infants with their thymus irradiated were compared against a control population. (Learn more about analysis of these studies in Statistics.) The resulting analysis showed that there was a linear dose-response of the excess relative risk (ERR) to the amount of radiation received (in Gy). (Note: A

Gy is a gray unit with 1 Gy = 1 Joule/kg. Typically, a whole body radiation dose of 10-20 Gy is fatal, which about the same energy as in $\frac{1}{4}$ g of sugar.) The best fitting line to the data is given by

$$ERR = 1 + 7.7\,d,$$

where ERR is the excess relative risk (factor of more cancers over the controls) and d is the dose of radiation. Make an Excel graph of this line over the domain $d \in [0, 5]$. Be sure to label your graph, including the axes. Place the equation of the line on the graph.

b. Below is an estimated table of the pooled data. Plot these data points and add them to the graph created in Part a.

d	0.1	0.4	0.7	1.3	1.7	3.5	4.2
ERR	2.1	4.5	5.1	11.0	8.0	31	29

Table 1.2: Data on the excess relative risk for children exposed to radiation based on the dose of radiation that they received as a child.

1.6 References:

[1] J. M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[2] H. Bremer and P. P. Dennis, "Modulation of chemical composition and other parameters of the cell by growth rate", *Escherichia coli and Salmonella typhimurium*, edḞ. C. Neidhardt, American Society of Microbiology, Washington, D. C. (1987).

[3] E. Ron, J. H. Lubin, R. E. Shore, K. Mabuchi, B. Modan, L. M. Pottern, A. B. Schneider, M. A. Tucker, and J. D. Boice Jr, (1995), "Thyroid cancer after exposure to external radiation: A pooled analysis of seven studies", *Radiation Research*, **141**, 259-277.

Chapter 2

Linear Models and Least Squares Best Fit

One of the most common models applied to biological data is the linear model. Due to the complexity of most biological problems, this type of model is rarely valid over a wide range of the variables. However, it does provide a useful type of model to begin our studies. One advantage of a linear relationship for a set of data is the easy algorithm to find the best possible straight line through the data, known as *linear regression* or the *linear least squares best fit* to the data. (Refer to the *Least Squares Analysis* chapter of the text book for more information [2].)

Assume that the linear model for a set of data points (x_i, y_i) is given by the equation

$$y = ax + b,$$

where the parameters a and b are chosen to "best fit" the data. The least squares best fit minimizes the square of the error in the distance between the y_i values of the data points and the y value of the line, which depends on the selection of the slope, a, and the intercept, b. The *absolute error* between each of the data points and the line is

$$e_i = |y_i - y(x_i)| = |y_i - (ax_i + b)|, \quad i = 1, ...n.$$

The least squares best fit is found by finding the minimum value of the function

$$J(a, b) = e_1^2 + e_2^2 + ... + e_n^2 = \sum_{i=1}^{n} e_n^2.$$

Details for how to find the best values of a and b uses Calculus of two variables and can be found in most statistics books. The details of this analysis are omitted, since it does require a little more knowledge of Calculus. However, the results are summarized below. Define the mean of the x values of the data points as

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The value for the slope of the line that best fits the data is given by

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

With the slope computed, the intercept is found from the formula

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - a\bar{x} = \bar{y} - a\bar{x}.$$

There are many computer programs that automatically compute a and b from data sets. In this chapter, we show how to use Excel's Trendline feature.

2.1 Excel's Trendline for Linear Models

The algorithm listed above is very tedious to compute by hand, but lends itself very well for a computer. This algorithm for finding the best slope and intercept for a linear model is embedded in Excel's graphing program under the name of *Trendline*. Below we demonstrate how to use this routine through an example of juvenile height.

Example: Juvenile Height

It can be readily seen that the growth of children follows a fairly linear model for a range of ages. Below is a table of data on the average height of children in the U. S. (This example is also studied in Section 5 of Chapter 2 of the textbook [2].)

The height h, is graphed as a function of age a. Note that the graph below is created in the usual manner of creating a table of data in Excel, then using **Chart Wizard** with its **XY-Scatter** option to create the graph with the data displayed as points. It is easy to see that the data almost lie on a line, which suggests a *linear* model. After the usual cosmetic changes

Age	1	3	5	7	9	11	13
Height (cm)	75	92	108	121	130	142	155

Table 2.1: Average juvenile height as a function of age [1].

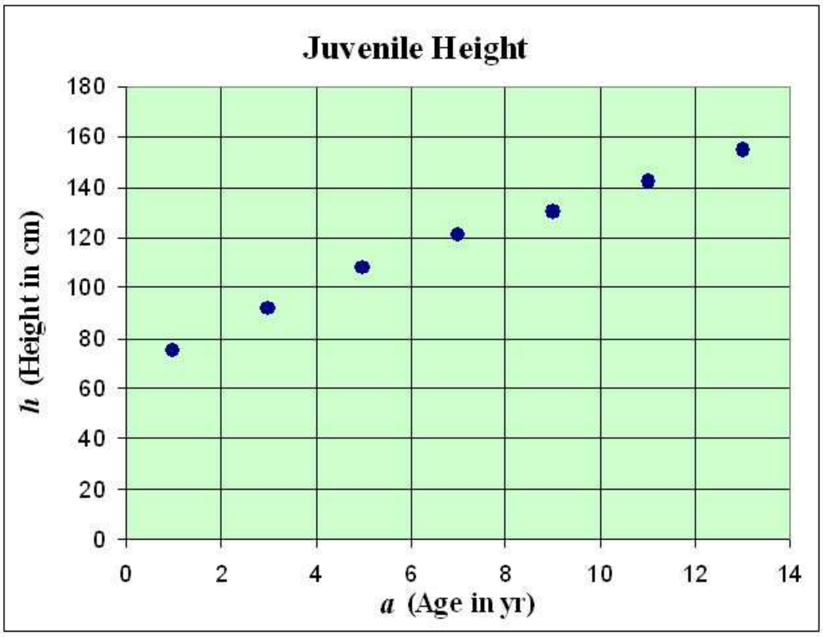


Figure 2.1: Graph of the juvenile height data.

to make the graph look good, we have the graph of Figure 2.1.

We want to use Excel's *Trendline* feature to find the best linear model through the data graphed above. There are two good ways to invoke Excel's Trendline. The easiest way is to right click on one of the data points, then select the option **Add Trendline**. Alternately, one can click (single) on one of the data points, then once the data points appear highlighted, return to the **Chart** on the **Main Menu** and select **Add Trendline**. With the menu that appears under Trendline, there are several options. Since the mathematical model in this case is a linear model, we stay with the default, **Linear**. Next select the **Options** folder in this window and check the box to **Display the equation on chart**. The result that appears in the middle of the graph is the equation

$$y = 6.4643x + 72.321.$$

We move this equation to a more visible part of the graph and make any other touch-ups to make the graph look good. For example, we changed the variables y to h (height) and x to a (age) to match the variables in our model, then moved the equation closer to the line. We also changed the font size and style (Times New Roman), then formatted the **Text Box** to have a **Pattern** that matches the background selected for the graph. The resulting graph is shown in Figure 2.2 and is ready for inclusion in a lab report.

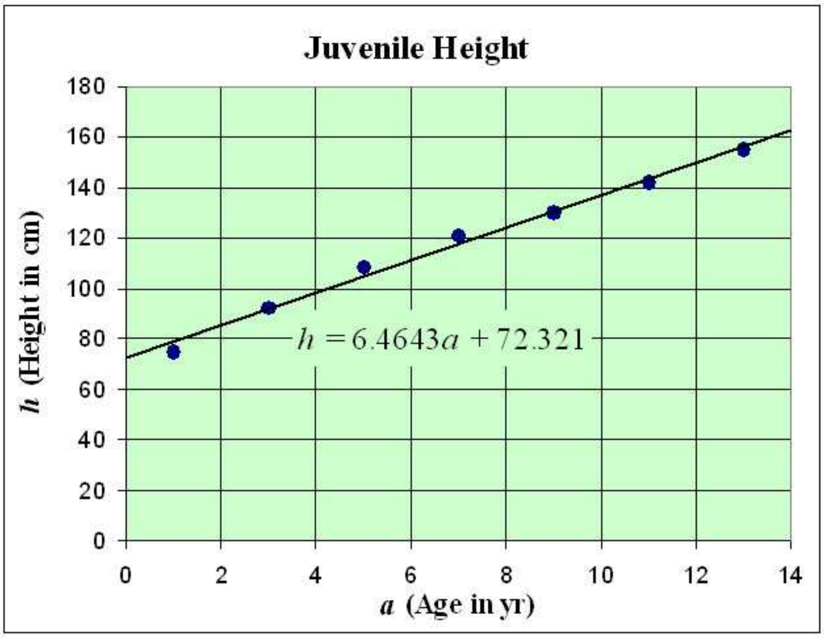


Figure 2.2: Juvenile height data with the best fitting straight line.

2.2 Computer Laboratory Exercises

To simplify listing of web addresses, the beginning of a web address below begins with the following:

`www-rohan.sdsu.edu/~jmahaffy/courses`

which will be given as **...** in the text below.

1. (B2) This problem examines some physiological data from the laboratory of Professor Carol Beuchat from San Diego State University. Animals have evolved different mechanisms for excreting waste nitrogen. The principle means of excreting nitrogen are uric acid , urea, and ammonia. Unfortunately, the latter two are toxic so require larger volumes of water for excretion. Uric acid uses less water, but it requires more energy (ATP) to produce. Thus, animals must weigh their needs of water versus energy when selecting a means of excretion.

a. Here we only examine the amount of urea excreted. First, a standard is run to determine the absorbance at 570 nm as a function of the concentration of urea. (This is a standard technique using spectrophotometry.) The data are listed below:

Urea conc. (mg/dl)	Absorbance
3	0.138
15	0.676
40	1.315
60	1.76
80	2.453
100	3.053
120	3.939
150	4.66
200	6.093

Use the EXCEL's trendline feature on a scatterplot to find the best straight line through the data, where

$$A = mU + b,$$

is the straight line describing absorbance (A) as a function of the concentration of urea (U) with EXCEL determining the slope (m) and intercept (b). Write the equation for the best linear model, and show the graph with both the data and this linear model.

b. Find the expected absorbance for a sample containing 53 mg/dl of urea and a sample containing 176 mg/dl of urea.

c. In practice, one will use the spectrograph to measure the absorbance, and use the relationship between the two to calculate the urea levels. In order to do this you must now solve for U as a function of A. That is find a function $f(A)$ such that

$$U = f(A).$$

This *inverse function* is also a straight line, so give its slope and intercept.

d. In Professor Beuchat's laboratory they found that the urine from a hummingbird kept at 10°C had an absorbance of 0.142. When the hummingbird was kept at 20°C, the absorbance for a urine sample was 0.201, while at 40°C the absorbance was 0.29. Find the corresponding values of concentration of urea for each temperature from the results in Part c. (Don't forget to include the correct units on your answer.) Can you explain your results with regard to either energy or water conservation by these hummingbirds?

e. We would like to see if the analysis of urine samples tells us about other species. The table below lists different animals and the corresponding absorbances measured.

Animal	Absorbance
Chicken	3.124
Duck (Fresh Water)	0.468
Duck (Salt Water)	0.782
Frog	0.267
Turtle	1.115
Tortoise	6.877

Find the corresponding values for the concentration of urea in each of these animals. A hummingbird gets its energy from nectar, which is high in water content. What animal has the most similar excretion pattern to a hummingbird? Do you see a pattern between the different animals here, and can you offer some explanations?

2. (B3) A few years ago some Exercise Physiologists at UCLA published a paper in *Nature* wherein they predicted that by the year 2004, the women's world record in the marathon would be faster than the men's record [3]. The mechanism for the improvement in performance is thought to be the improvement of training methods and the expansion of the talent pool. But the data was examined only to describe the trend, not to explain it. This problem examines the winning Olympic times for the 100 m races for both Men and Women. As the years have gone by, the times have improved for both Men and Women. Below we present a table with the data for the

winning times (in seconds)

Men's 100	time	Women's 100	time
Burke	12		
Jarvis	11		
Hahn	11		
Hahn	11.2		
Walker	10.8		
Craig	10.8		
Paddock	10.8		
Abrahams	10.6		
Williams	10.8	Robinson	12.2
Tolan	10.3	Walasiewicz	11.9
Owens	10.3	Stephens	11.5
Dillard	10.3	Blankers-Koen	11.9
Remigino	10.4	Jackson	11.5
Morrow	10.5	Cuthbert	11.5
Har	10.2	Rudolph	11
Hayes	10	Tyus	11.4
Hines	9.95	Tyus	11
Borsov	10.14	Stecher	11.07
Crawford	10.06	Richter	11.08
Wells	10.25	Kondratyeva	11.06
Lewis	9.99	Ashford	10.97
Lewis	9.92	Joyner	10.54
Christie	9.96	Devers	10.82
Bailey	9.84	Devers	10.94
Greene	9.87	Jones	10.75

a. Use EXCEL's trendline feature to find the best straight lines (one for Men and one for Women) through the data, where

$$T = mY + b$$

is the straight line for the best time (T) as a function of the Olympic year (Y) with EXCEL determining the slope (m) and intercept (b). Write the equations for the best linear models and show (on a single graph) the graphs of the data and linear model for both Men and Women. Be sure to label which lines correspond to the data for the Men and Women.

b. Use the model to determine the predicted year when the best time is 10.0 sec for Men and 11.0 sec for Women, then compare your prediction to the actual data.

c. Use the model to predict the time for the 2000 and 2004 Olympics for both Men and Women in this event. Give the percent error between the actual and predicted value in 2000.

d. According to the model, which Olympics will first see Women outrunning the Men? Give a short discussion on the validity of this prediction and why you think it is true or false. What fundamental premise do you consider to be critical? Can you formulate another model that might be more valid?

3. (B1) A hyperlink at

`.../s00a/math121/labs/labb/q1v1.htm`

is provided for finding an applet for the least squares best fit to a quadratic equation given by

$$h = ax^2 + bx + c.$$

Since there are only three data points, the best quadratic always fits the data exactly.

a. Enter the data set number 1 in the upper right corner of the applet to get your first data set. Adjust the parameters for the coefficients a, b, and c until you find a sum of squares value of 0.0. (Hint: The coefficients are integer valued.) Write the equation for this quadratic, then determine the values of the x and h-intercepts and the coordinates of the vertex.

b. Repeat the process in Part a. for the data set number 2 in the upper right corner of the applet.

c. Write a brief discussion on what you observe as you increase c. Similarly, write a brief description of what happens to the graph of the quadratic as you change b and a. Be sure to describe what happens as a changes sign. You may find that the changes caused by varying b are different when a is less than zero. In your discussion, you should clearly write about each case of the parameters, a, b, and c, as they vary, noting differences when the parabola changes from pointing up to pointing down. For each case, describe what happens to the vertex (direction of shifts), the distance across the parabola, and/or the direction of opening of the parabola.

2.3 References

[1] David N. Holvey, editor, *The Merck Manual of Diagnosis and Therapy* 15^{th} ed., Merck Sharp & Dohme Research Laboratories, Rahway, NJ, (1987).

[2] J. M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[3] B. J. Whipp and S. Ward (1992). Will women soon outrun men?, *Nature*, **355**: 25.

Chapter 3

Functions and an Introduction to Maple

This chapter introduces the powerful symbolic algebra program Maple. Maple handles algebraic manipulations very easily and accurately, produces graphs easily, and performs many other mathematical calculations. In this lab manual, we only introduce a very small sample of Maple commands, but using the Maple tutorials and its help page, you can easily extend your working knowledge with this program. The reader should be aware that Mathematica is an equivalent software package that is used on other campuses. (It is entered slightly differently, so this manual will not provide the novice with information to use that software, but it handles the same problems shown below with Maple.)

We begin by showing how a typical Maple window should appear after invoking the program. The window shown in this figure shows Maple solving the quadratic equation and graphing a special case.

3.1 Maple - A Sample Problem

This example is designed to introduce the reader to basic Maple commands. Two functions are used to show how to enter functions, graph these functions, and find points of intersection.

Example 1: This example is meant to introduce Maple. Consider the

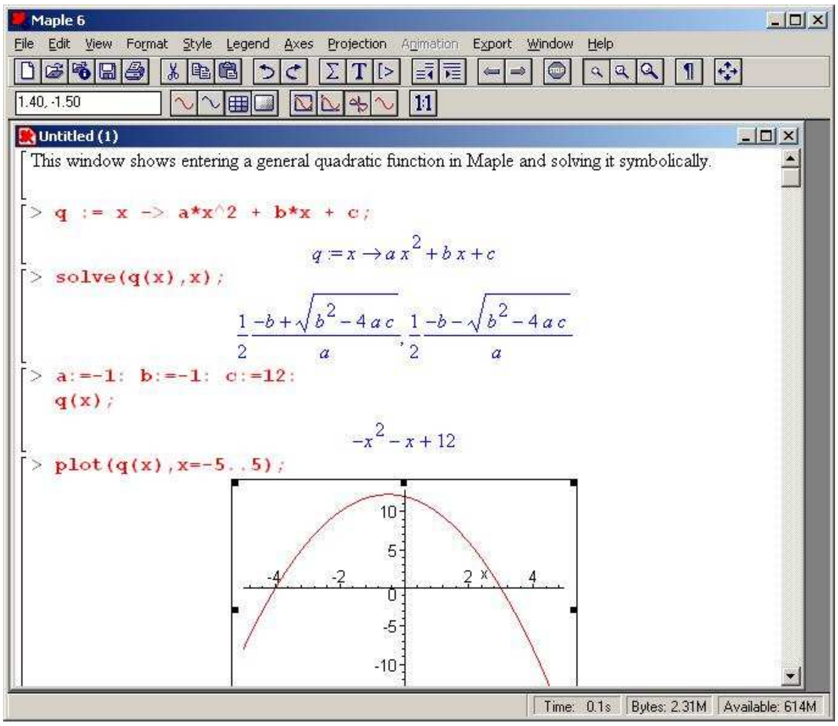


Figure 3.1: Maple program window with the solution of the quadratic equation.

following two equations:

$$f(x) = x + 2 \qquad \text{and} \qquad g(x) = x^2 - 2x - 2.$$

We want to find the x and y-intercepts of both functions, determine where these functions intersect, then sketch a graph of the functions.

Solution The first Maple command shows how functions are entered in Maple. (There are several ways to enter functions in Maple, but we are using the form that creates functions $f(x)$ as most frequently used in standard textbooks.) To enter the functions in Maple, we type

```
> f := x -> x + 2;
> g := x -> x^2 - 2*x - 2;
```

The output of this command by Maple will produce functions in a math font that looks like the functions you see in textbooks.

$$f := x \rightarrow x + 2;$$

and

$$g := x \to x^2 - 2x - 2;$$

The part given by

```
f := x ->
```

is used to define the function $f(x)$ in Maple. The semicolon (;) is crucial for ending all Maple commands. The semicolon completes each Maple command that produces an output, while a colon (:) ends a Maple command, but doesn't result in any output being observed. To execute a Maple command, you need to hit the **Enter** key. You can execute the command from any point in the command, not necessarily at the end of the command. (If you desire going to the next line in Maple not executing the command, then you must use **Shift** and **Enter**.)

Now it is easy to have Maple evaluate the function. For example, the function f evaluated at $x = 5$ is found by typing

```
> f(5);
```

which yields **7**. It follows that finding the y-intercepts of these functions is easily done by the following Maple command:

```
> f(0); g(0);
```

It follows that the y-intercepts are 2 and -2 for f and g, respectively.

To find the x-intercepts of these functions, there are two Maple commands. The **solve** command solves the equations algebraically (exactly), while the **fsolve** command solves the equations numerically. Since f and g are linear and quadratic functions respectively, it is always possible to find an analytical solution. The x-intercept for $f(x)$ is $x = -2$ and can be found in Maple by typing

```
> solve(f(x)=0, x);
```

The x-intercept for $g(x)$ can be found using the quadratic formula or in Maple, we type

```
> solve(g(x)=0, x);
```

Maple returns the exact solution

$$1 + \sqrt{3},\ 1 - \sqrt{3}$$

Alternately, we could use the Maple command **fsolve**, which gives the floating point or decimal solution to the equation. Thus, typing

```
> fsolve(g(x)=0, x);
```

gives the answers $x = -0.7320508076$ and $x = 2.732050808$. The x next to the equation indicates to Maple that you want to solve the equation for x and not for any other possible variable.

It is usually a good idea to have a graph of the functions with which you are working. The purpose of the graph is often just to visualize the functions in the problem. Graphing in Maple is very simple (though the output is not as elegant as it is in Excel). To graph the two functions $f(x)$ and $g(x)$, you simply type

```
> plot({f(x),g(x)},x = -5..5);
```

This plot command has numerous options, and we will show several of these later in the lab manual. Above this plot command simply tells Maple to plot the two functions $f(x)$ and $g(x)$ over the interval $x \in [-5, 5]$.

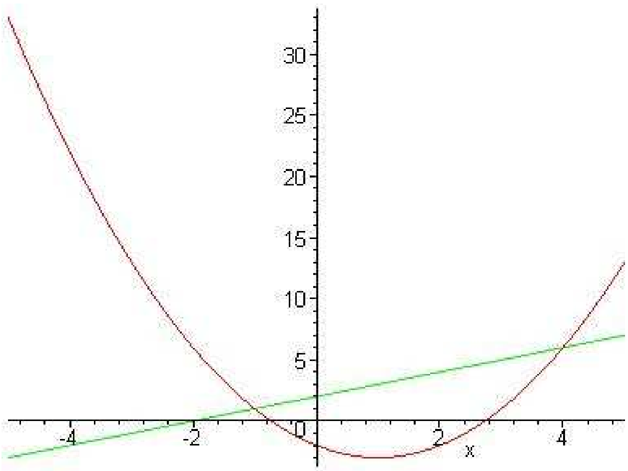


Figure 3.2: Graph of $f(x)$ and $g(x)$.

To find the points of intersection we need to set $f(x) = g(x)$ and solve for x. Again we can do this using either **solve** or **fsolve**. (I would recommend against using **solve** if you have any polynomial of degree higher than 2.)

```
> solve(f(x)=g(x),x); fsolve(f(x)=g(x),x);
```

Notice that you can put multiple Maple commands on one line, and Maple does the operations in the order you place them.

If we want both the x and y values of the points of intersection, then we need the following (assuming 2 points of intersection, it varies slightly if there is only one point):

```
> xs := fsolve(f(x)=g(x),x); f(xs[1]); f(xs[2]);
```

The **xs** stores the values of x created by the **fsolve** command. Since there are two values **xs[1]** gives the first x created by **fsolve** and **xs[2]** gives the second one. Writing **f(xs[1]);** gives the y value as it is the function evaluated at that x value. Note that if there was only one value, then it is **xs** and you get the y value by typing **f(xs);**. The Maple command above gives the points of intersection as $(-1, 1)$ and $(4, 6)$.

3.1.1 Maple - Algebraic Assistance

Maple is a powerful symbolic algebra package. This makes it a powerful tool to help with algebraic simplifications that are tedious. Here we introduce a few commands that are helpful with algebraic manipulations. Most of these commands are fairly obvious what they do by their names.

We begin with the commands **expand**, **factor**, and **simplify**. In several of the least squares problems, we need to find the vertex of the parabola to find the best fit of a straight line through the origin, but multiplying many terms becomes tedious. The **expand** operation can easily produce a polynomial as is shown below.

```
> expand((x-2.3)^2+(x-4.5)^2+(x-6.2)^2+(x-8.1)^2);
```

Maple gives the answer

$$4x^2 - 42.2x + 129.59,$$

which has a vertex at $x = 5.275$. We recall that the vertex of a quadratic function occurs at

$$x = -\frac{b}{2a}.$$

We can obtain that vertex with the following Maple command (assuming it directly follows the expand command above):

```
> -coeff(%,x,1)/(2*coeff(%,x,2));
```

Maple uses the % symbol to represent the previous result obtained by Maple (such as the answer to the expand command). **Be very careful** using the % symbol in Maple as it represents the last response from Maple that was entered and not necessarily the expression in the line above the %, since you can use the cursor to go to other parts of the worksheet. The **coeff(%,x,1)** command written above tells Maple to use the coefficient of the previous expression found (%) that corresponds to x^1. Similarly, **coeff(%,x,2)** tells Maple to use the coefficient of x^2 of the previous expression found.

The **factor** command can easily factor polynomial expressions if possible. Consider the example:

```
> factor(x^4-x^3-7*x^2+x+6);
```

Maple gives the answer

$$(x-1)(x+2)(x+1)(x-3).$$

Often there are expressions that are not easily simplified by hand, so Maple provides a valuable tool for simplifying these expressions. Consider the following example, which will occur when we study the derivative. We would like to simplify the expression

$$\frac{(x+h-5)^3-(x-5)^3}{h},$$

which is not something that one would like to perform by hand. Maple readily does this operation as follows:

```
> simplify(((x+h-5)^3-(x-5)^3)/h);
```

Maple gives the answer

$$3x^2+3xh-30x+h^2-15h+75.$$

3.2 Functions with Asymptotes

A rational function is a polynomial divided by another polynomial. This form of a function may have horizontal or vertical asymptotes . The vertical asymptotes often occur where the function is undefined. The horizontal asymptotes are found by looking at very large values of the function. All of these properties are easily done in Maple. Creating graphs with asymptotes, especially if you want to display the asymptote, creates special challenges in Excel. In this section, we show how to graph a function with asymptotes in both Maple and Excel and show how Maple can help find these asymptotes.

Example 2: Consider the following functions:

$$f(x) = x + 1 \qquad \text{and} \qquad g(x) = \frac{x}{x^2 - 4}.$$

Graph these functions. Find all vertical and horizontal asymptotes, then determine the points of intersection.

Solution: We begin with a command to graph these functions in Maple. Since $g(x)$ has vertical asymptotes, we want to limit the range of the graph. The **plot** command in Maple allows the graphing of $f(x)$ and $g(x)$ as follows:

```
> plot({f(x),g(x)}, x=-5..5, y=-10..10, discont=true);
```

As before, the domain is defined with $x \in [-5, 5]$. Two other options are included: the range limited to $y \in [-10, 10]$ and letting Maple know that we suspect vertical asymptotes, **discont=true**.

Maple can be used to find the vertical asymptotes by finding when the denominator is zero. This is accomplished by either typing

```
> fsolve(x^2-4=0,x);
```

or alternately,

```
> dn := denom(g(x));
> fsolve(dn=0,x);
```

where **dn** is a dummy variable for the denominator of $g(x)$. (In Maple, the characters (**:=**) represent "defined to be" and are used to assign an expression or value to the right of those characters to a user defined variable to the left of the characters. For example, **x := 5** assigns the value of 5 to x.) The vertical asymptotes occur at $x = -2$ and 2 for $g(x)$.

To find the horizontal asymptote you can use Maple's limit command.

```
> limit(g(x),x=infinity);
```

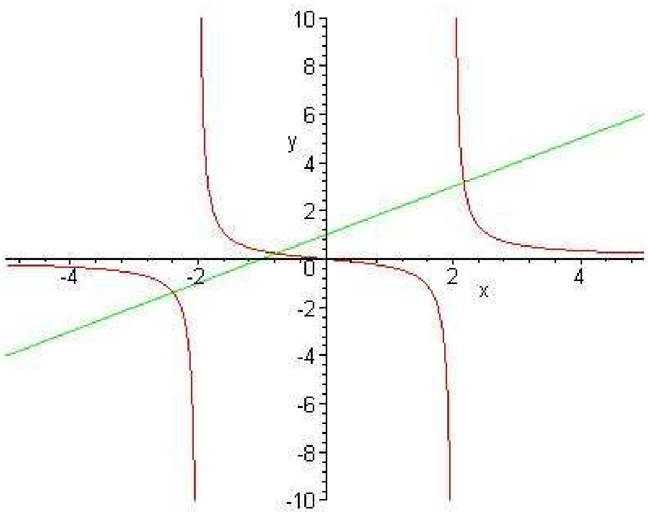


Figure 3.3: Graph of $f(x)$ and $g(x)$.

which if finite gives the horizontal asymptote to the right ($y = 0$), while

```
> limit(g(x),x=-infinity);
```

gives the horizontal asymptote to the left ($y = 0$). Thus, the function $g(x)$ has a horizontal asymptote at $y = 0$.

To find where $f(x)$ and $g(x)$ intersect, you use the fsolve command. We modify this slightly from Example 1 because this command would only find one of the 3 solutions. To find all solutions you need to limit where Maple looks for the solutions. The second expression in the **fsolve** command below gives the range where this Maple routine searches for the x value of the intersection, and this range is found from the graph.

```
> x1 := fsolve(f(x)=g(x),x=-5..-2); f(x1);
> x2 := fsolve(f(x)=g(x),x=-2..2); f(x2);
> x3 := fsolve(f(x)=g(x),x=2..5); f(x3);
```

The resulting points of intersection are $(-2.3914, -1.3914)$, $(-0.7729, 0.2271)$, and $(2.1642, 3.1642)$.

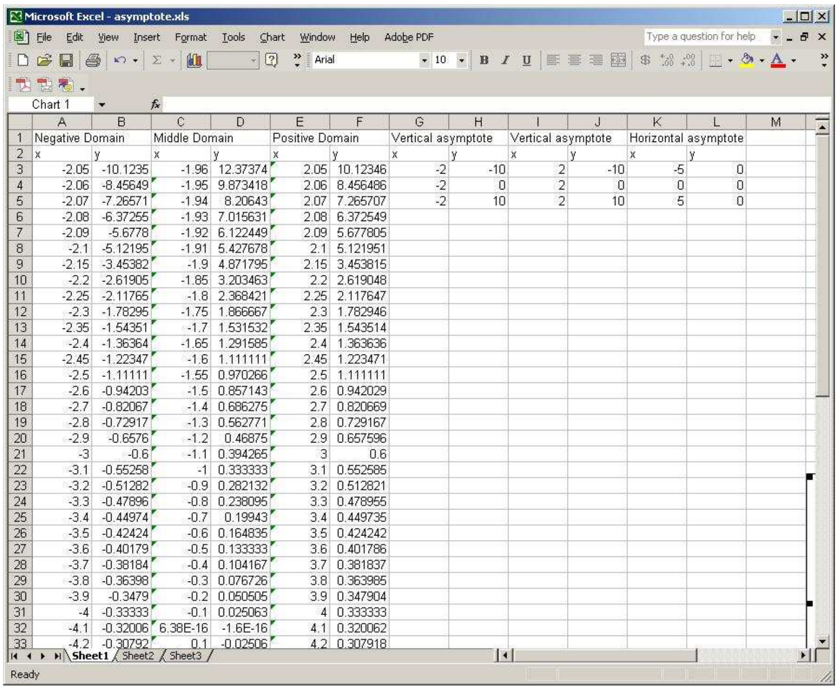
In the next example, we use the rational function above to show how to graph a function in Excel showing all the asymptotes.

Example 3: Consider the function:

$$g(x) = \frac{x}{x^2 - 4}.$$

Use Excel to create a graph of this function, showing all vertical and horizontal asymptotes.

Solution: We will take advantage of the information learned in the previous example about the vertical and horizontal asymptotes to create the Excel graph of this function. The vertical asymptotes at $x = \pm 2$ divide our domain into three distinct pieces. This means the domain is: $(-\infty, -2) \bigcup (-2, 2) \bigcup (2, \infty)$. To create a graph in Excel, we must separately evaluate the function in these three domains, then combine the graph into a single graph. Furthermore, because of the vertical asymptotes, we need to evaluate the function in shorter intervals near the asymptotes to make the graph look better.


Microsoft Excel - asymptote.xls

	A	B	C	D	E	F	G	H	I	J	K	L
1	Negative Domain		Middle Domain		Positive Domain		Vertical asymptote		Vertical asymptote		Horizontal asymptote	
2	x	y	x	y	x	y	x	y	x	y	x	y
3	-2.05	-10.1235	-1.96	12.37374	2.05	10.12346	-2	-10	2	-10	-5	0
4	-2.06	-8.45649	-1.95	9.873418	2.06	8.456486	-2	0	2	0	0	0
5	-2.07	-7.26571	-1.94	8.20643	2.07	7.265707	-2	10	2	10	5	0
6	-2.08	-6.37255	-1.93	7.015631	2.08	6.372549						
7	-2.09	-5.6778	-1.92	6.122449	2.09	5.677805						
8	-2.1	-5.12195	-1.91	5.427678	2.1	5.121951						
9	-2.15	-3.45382	-1.9	4.871795	2.15	3.453815						
10	-2.2	-2.61905	-1.85	3.203463	2.2	2.619048						
11	-2.25	-2.11765	-1.8	2.368421	2.25	2.117647						
12	-2.3	-1.78295	-1.75	1.866667	2.3	1.782946						
13	-2.35	-1.54351	-1.7	1.531532	2.35	1.543514						
14	-2.4	-1.36364	-1.65	1.291585	2.4	1.363636						
15	-2.45	-1.22347	-1.6	1.111111	2.45	1.223471						
16	-2.5	-1.11111	-1.55	0.970266	2.5	1.111111						
17	-2.6	-0.94203	-1.5	0.857143	2.6	0.942029						
18	-2.7	-0.82067	-1.4	0.686275	2.7	0.820669						
19	-2.8	-0.72917	-1.3	0.562771	2.8	0.729167						
20	-2.9	-0.6576	-1.2	0.46875	2.9	0.657596						
21	-3	-0.6	-1.1	0.394265	3	0.6						
22	-3.1	-0.55258	-1	0.333333	3.1	0.552585						
23	-3.2	-0.51282	-0.9	0.282132	3.2	0.512821						
24	-3.3	-0.47896	-0.8	0.238095	3.3	0.478955						
25	-3.4	-0.44974	-0.7	0.19943	3.4	0.449735						
26	-3.5	-0.42424	-0.6	0.164835	3.5	0.424242						
27	-3.6	-0.40179	-0.5	0.133333	3.6	0.401786						
28	-3.7	-0.38184	-0.4	0.104167	3.7	0.381837						
29	-3.8	-0.36398	-0.3	0.076726	3.8	0.363985						
30	-3.9	-0.3479	-0.2	0.050505	3.9	0.347904						
31	-4	-0.33333	-0.1	0.025063	4	0.333333						
32	-4.1	-0.32006	6.38E-16	-1.6E-16	4.1	0.320062						
33	-4.2	-0.30792	0.1	-0.02506	4.2	0.307918						

Figure 3.4: Excel spreadsheet showing the divisions of the domain for graphing $g(x)$, including lines for the asymptotes.

We choose to plot the graph on the interval $[-5, 5]$. Figure (3.4) shows the numbers that are entered to produce a graph of $g(x)$ in this example. **Columns A** and **B** contain the x and y values for the function with $x \in [-5, -2]$. It starts at $x = -2.05$, since we only plan to plot y values between -10 and 10. Small steps between x values of only -0.01 are used for the first 5 entries, then the stepsize is increased to -0.05 for the next 8 entries, and finally, the stepsize is increased to -0.1 for the remaining x values until $x = -5$. This allows for a total of about 40 points that we evaluate the function on this interval. (You generally want to take 40-60 function evaluations to obtain a good curve. Also note that when plotting in Excel, you should use the straight line between points option for most function evaluations. The curved line option gives good curves, but students often take too few points with this option and the Excel polynomial curve fitting can give the wrong curve!) After entering these values, **Chart Wizard** is invoked to produce the curve for $x \in [-5, -2]$.

The next step is to add the other portions of the curve. Figure (3.4) shows the middle portion of the graph listed in **Columns C** and **D** and the right portion of the graph listed in **Columns E** and **F**. Again the stepsize is variable with very small steps near the asymptotes and larger steps where the curve is flatter. Once again about 40-60 function evaluations were used. These sections are added to the graph sequentially by first clicking on the chart produced when graphing **Columns A** and **B**, which results in the corners of the Chart to have small black squares indicating the Chart is active. Next you click on the **Chart** label at the top of the Excel spreadsheet and select the option **Add Data...** A window appears titled **Add Data** with the cursor blinking in a window saying **Range**. You can type in the range of your data or more easily (and more accurately), you simply highlight the data that you are interested in adding to the graph. For example, to add the data for the middle part of this graph, we highlight **C3** through **D61**. Next we click **OK**, and a new box appears with the title **Paste Special**. The default of Add series as **New Series** and **Values (Y) in Columns** should be left alone, but it is very important to check the box **Categories (X Values) in First Column**. When you click **OK** in this box, then you should see the new portion of the graph added to the previous portion of the graph. After repeating this process on **Columns E** and **F** to get the right portion of $g(x)$, we have our graph of the function. However, at this point, the function will not look very good as is standard in Excel.

The next part of this problem is to add the asymptotes, which we see in Figure (3.4) in **Columns G** to **L**. We put three points lying on each of the asymptotes, then entered these lines much as we did for the other pieces of the curves described in the previous paragraph. For example, in **Column G** we simple put $x = -2$, then in **Column H** used the y values at -10, 0, and 10 to produce a graph of the vertical asymptote at $x = -2$. Similarly, the vertical asymptote at $x = 2$ and the horizontal asymptote at $y = 0$ are entered. (Note that Excel has not found these for you!) At this point the graph will appear as seen in Figure (3.5). This is clearly not a very good looking graph.

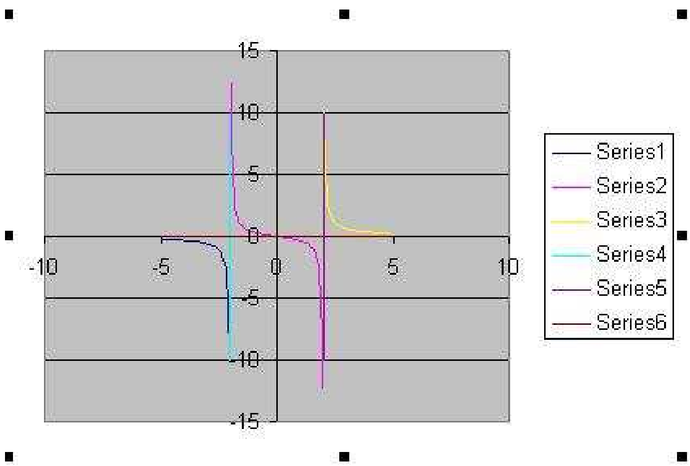


Figure 3.5: Initial Excel graph of $g(x)$ before adding labels and other details.

The final portion of this problem is the cleaning up and making the graph look good, including labeling. We begin by going to the Menu line at the top of the page, and under **Chart** we select the **Chart Options**category. We add a title to the graph and label the x and y axes. We add **Gridlines** and delete the **Legend**. After enlargening the graph, we adjust the domain (x) to $-5 \le x \le 5$ and the range (y) to $-10 \le y \le 10$. Recall this is done by right clicking on the appropriate axis and selecting **Format Axis**, then selecting **Scale**. (You can also modify your font at this stage.) By double clicking on each portion of the graph, we make each section of the graph have the same color and line thickness. We selected a dark blue medium width solid line for $g(x)$ and a medium width dashed red line for the asymptotes. The resulting final is shown in Figure (3.6).

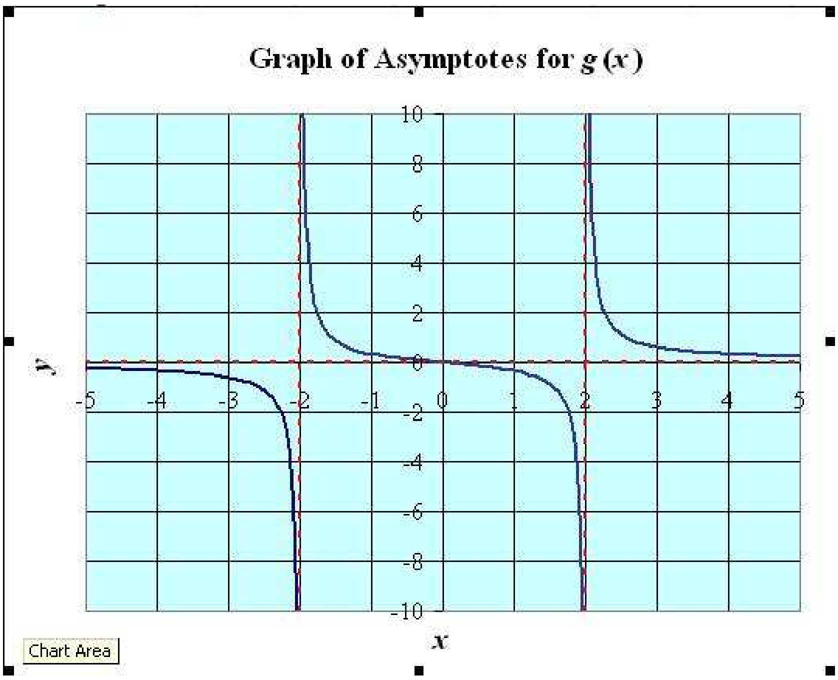


Figure 3.6: Excel graph of $g(x)$ clearly showing the asymptotes.

3.3 Computer Laboratory Exercises

1. (C1) a. Consider the following quadratic function,

$$f(x) = x^2 - 4x - 3.$$

Find the x and y-intercepts and the vertex for $f(x)$. Give both the exact solution and decimal solution (with 4 significant figures) for the x-intercept. Use Excel to graph this function, choosing a domain that includes the x and y-intercepts and the vertex for $f(x)$.

b. Now consider these functions, $f(x)$ and $g(x)$. Find the x and y-intercepts for both functions. Give the slope of the line. Find all points of intersection between the curves:

$$f(x) = 2 - x \qquad \text{and} \qquad g(x) = x^3 - 2x^2 - 3x + 3.$$

For all the intercepts and points of intersection give the decimal value to 4 significant figures. Use Excel to graph these functions. Choose a domain such that the graph clearly shows all the points you identified above.

2. (D1) a. Consider the following linear and rational functions,

$$f(x) = 4 - x \qquad \text{and} \qquad g(x) = \frac{2x}{6 - x - x^2}.$$

Find the x and y-intercepts both for $f(x)$ and $g(x)$. For $g(x)$, find any vertical or horizontal asymptotes.

b. Use Maple to graph these functions for $x \in [-6, 6]$ with the range restricted so that $-10 \leq y \leq 10$. Copy this graph into your Word document.

c. Use the graph and Maple to help you find all points of intersection, giving both the x and y values (give the decimal value to 4 significant figures).

3. (C3) In 1913, Carlson [1] studied the growth of a culture of yeast, *Saccharomyces cerevisiae* . Over time this culture levels off, but its initial growth is exponential or Malthusian . A Malthusian growth model is given by

$$P_{n+1} = P_n + rP_n.$$

(This model will appear later in this Lab Manual for more detailed study.) Even though a population grows continuously we take discrete measurements every time n for n an integer number such that $n = 0, 1, 2, ...$ Simply put, the population at the next time $(n + 1)$ is equal to the population at the current time (n) plus some growth term, which is simply proportional (r) to the current population. Thus, we have a growth function

$$g(P) = rP.$$

Below is a table from Carlson's data showing the population and the rate of growth at that particular population

a. Plot these data. Use Excel's Trendline on the data points to find the best straight line passing through the origin. (Note you will need to use the option in Trendline of setting the y-intercept = 0.) What is the slope of the line that best fits through the data?

b. The accompanying text for this Lab Manual has a section (Function Review and Quadratics [2]) that examines a linear model for mRNA synthesis. For the linear model (passing through the origin) given above, we can

Population (P)	$g(P)$ (# per hour)
18,300	10,700
47,200	23,900
119,100	55,500
174,600	82,700

readily find the sum of squares function . Consider a data point $(P_i, g(P_i))$. The error between this data point and our model is given by

$$e_i = |g(P_i) - rP_i|.$$

Thus, $e_1 = |10,700 - 18,300r|$. Similarly, you can find e_2, e_3, and e_4. The sum of squares function is given by

$$J(r) = e_1^2 + e_2^2 + e_3^2 + e_4^2.$$

Find the expression for the quadratic function of the slope of the model, r (in simplest form). Sketch a graph of this quadratic for $r \in [0.3, 0.7]$.

c. Find the r-coordinate of the vertex and compare this to the slope Trendline gives you in Part a.

4. Characterizing the properties of enzymes is a very important endeavor in the biological sciences. Many enzymes follow Michaelis-Menten kinetics and have the following form for their reaction rates:

$$R([S]) = \frac{V_{max}[S]}{K_m + [S]},$$

where $[S]$ is the concentration of the substrate that the enzyme is catalyzing, $R([S])$ is the rate of production of the product, V_{max} is the maximum rate of the reaction (and depend on the enzyme concentration used, and K_m is the Michaelis-Menten constant, which characterizes a particular enzyme.

Below are data from Schmider, et. al. [3] for cytochrome P450 mediated demethylation of amitriptyline (AMI) to nortriptyline (N) by human liver microsomes.

[AMI] (μM)	N formation nmol/min/mg
5	0.20
10	0.45
15	0.60
25	0.80
50	1.35
100	2.17
200	2.68
260	2.89
330	3.05
500	3.12

From this table, we let $[S] = $ [AMI] and $R([S]) =$ N formation. We want to find the best values of V_{max} and K_m to characterize the enzyme cytochrome P450.

a. One of the easiest methods for finding the constants V_{max} and K_m is using the Lineweaver-Burk plot as discussed in [2]. To do this, you take the data and create a new table with the values $x = 1/[S]$ and $y = 1/R([S])$. Create this new table in Excel, then plot the data y vs x. Use Trendline to find the best fit to this straight line. The y-intercept has the value $1/V_{max}$ and the slope has the value K_m/V_{max}. Use this information to find V_{max} and K_m.

b. With this information, plot the original data with the Michaelis-Menten reaction formula for $[S] \in [0, 500]$. Find the prediction from the Michaelis-Menten reaction formula for the production of nortriptyline (N) when [AMI] is $[S] = 15, 50, 100, 200,$ and 500. Calculate the percent error between the model and the actual data. Find all intercepts and any asymptotes for this function for $[S] \geq 0$.

c. When a nonlinear least squares fit is applied to the data above, a better model is given

$$R([S]) = \frac{3.738[S]}{80.63 + [S]}.$$

With this information, plot the original data with this Michaelis-Menten reaction formula for $[S] \in [0, 500]$. Find the prediction from this Michaelis-Menten reaction formula for the production of nortriptyline (N) when [AMI] is $[S] = 15, 50, 100, 200,$ and 500. Calculate the percent error between the model and the actual data. Find all intercepts and any asymptotes for this function for $[S] \geq 0$. Which model would you say is better and why?

5. This problem is probably best done using Maple. Consider the quadratic function

$$f(x) = x^2 - 3x - 5$$

and the rational function

$$g(x) = \frac{20x}{1.4 + x}.$$

(You will probably want to graph these functions on the interval $x \in [-10, 10]$ with the range restricted to $y \in [-50, 50]$.)

a. Find the x and y-intercepts for both of these functions. Find the vertex of the quadratic function, $f(x)$. Give both x and y values. List any asymptotes (vertical and horizontal) for the rational function, $g(x)$.

b. Find all points of intersection between the graphs of $f(x)$ and $g(x)$.

3.4 References

[1] T. Carlson Über Geschwindigkeit und Grösse der Hefevermehrung in Würze. Biochem. Z. **57**: 313-334, 1913.

[2] J.M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[3] Schmider J., Greenblatt D. J., Harmatz J. S., Shader R. I., (1996) Enzyme kinetic modelling as a tool to analyse the behaviour of cytochrome P450 catalysed reactions: application to amitriptyline N-demethylation, *Brit. J. Clin. Pharmoco.* **41**, 593-604.

Chapter 4

Special Functions

Most problems in Biology are nonlinear. It follows that a number of different functions are needed to interpret biological data. This section shows how to enter a number of different functions into Excel and Maple to use for modeling biological problems.

4.1 Square Roots, Logarithms, and Exponentials

In solving the quadratic equation, we obtain square roots . To enter the square root function in Excel, one types **sqrt**. For example, if you enter **=sqrt(2)** in any Excel cell, then the resulting number that appears is **1.414213562**. Excel is not case sensitive, so one can enter sqrt, SQRT, or Sqrt and the correct response will be given. Maple is case sensitive, so one must enter **sqrt** for the square root function in Maple. Maple treats this function slightly differently, so entering **sqrt(2);** in Maple gives the response $\sqrt{2}$. However, entering **sqrt(2.);** in Maple gives the response **1.414213562**. (Maple handles integers and decimals slightly differently.) One can readily convert the integer values to decimal values in Maple with the **evalf** command, which is demonstrated below in the paragraph on the exponential function.

Maple and Excel differ on their handling of the logarithm function. Most computer languages use the natural logarithm, so typing $\log(x)$ is the same as typing $\ln(x)$. However, Excel is a business oriented program, so taking a logarithm in Excel defaults to the logarithm base 10, *i.e.*, $\log(x) =$

$\log_{10}(x)$ in Excel. To get logarithm base 10 in Maple one types **log10**. The natural logarithm in Excel is obtained by typing **ln**. Be careful when graphing functions with the logarithm to avoid letting the domain allow the value zero, as the logarithm function has a vertical asymptote at $x = 0$ and is undefined for $x \leq 0$. The fill down function in Excel works with all special functions, including the logarithm function, so graphing is handled for logarithms much in the same way as we showed in the previous chapters of this lab manual.

For the logarithm in Excel, if we type **=log(2)** in a cell, the result is **0.30103**, while typing **=ln(2)** yields **0.693147**. For Maple, typing **log(2);** yields the response ln(2), while typing **log(2.);** or **ln(2.);** gives the answer **0.693147181**. Typing **log10(2.);** gives the answer **.3010299957** in Maple.

Finally, we need to introduce the exponential function. In almost all computer languages, including Excel and Maple, the exponential function, e^x, is entered as **exp(x)**. (Note that there is **NO** carat between the 'p' and the '(x).' This is the most common error made by students in any lab using the exponential function.) Thus, in Excel we type **= exp(5)** and the result is **148.4132**. Again, we could type exp, EXP, or Exp in Excel and have the same result. In Maple, we type **exp(5.);** and the result is **148.4131591**. Typing **exp(5);** in Maple produces e^5. We can still get the decimal value by using the Maple command **evalf**. Thus, by entering **evalf(%)** immediately following Maple's response of e^5, then we obtain **148.4131591**.

4.2 Allometric Models and the Power Law

Biological problems often have a nonlinear relationship between different measured quantities, such as the weight of an animal and the amount of food that it consumes. **Allometric models** find a **power law** relationship between the measurements, which can be useful in predicting other related quantities, such as food consumption of a different animal than the one measured with the data. Below we describe how to use Excel to find an allometric model for the best Olympic times in the rowing event based on the number of oarsmen.

Example 1: A power law expression relating the number of oarsmen (n)

to the winning time in the Olympics (T) is given by

$$T = kn^a,$$

where k and a are constants to be determined. Refer to Chapter 6 of the text for more examples and information on Allometric models [1]. Below is a table of the Olympic rowing event and the best times for the event. Create

Event	# of Oarsmen n	Winning Time T (sec)	Country/Year
Single Scull	1	404.85	Switzerland/1996
Double Scull	2	371.49	Great Britain/2004
Cox-less Pair	2	380.09	Great Britain/1996
Cox-less Four	4	350.44	Great Britain/2004
Eight	8	319.85	United States/2004

a graph of these data and find the best power law model. Also, create a log-log plot of the data and the model.

Solution: If the natural logarithm is taken on both sides, then the allometric (power law) model has a straight line fitting the logarithms of data, giving

$$\ln(T) = \ln(k) + a \ln(n).$$

Finding the values of k and a is simply a matter of finding the slope a of this straight line and the intercept $\ln(k)$, which is easily done.

Excel finds power law models very easily. Highlight the data and create a graph. Next we select Excel's Trendline, and one of the options is the **Power** , which gives the best Power Law fit to the data. After selecting this option, we check the additional option, **Display equation on chart**. After some modifications to make the graph look better (font and color changes and modifying the default x and y variables of Excel to the variables in our problem), we produce the graph in Figure 4.1.

Excel can easily create **semi-log** or **log-log** plots. In our example above, we can create the **log-log** plot by simply activating the graph as usual. Next we right click on the axis that we want to change and select **Format axis**. A new window appears, and we select the scale folder in which we check the **Logarithmic scale** option . This changes the scale of the chosen axis. If

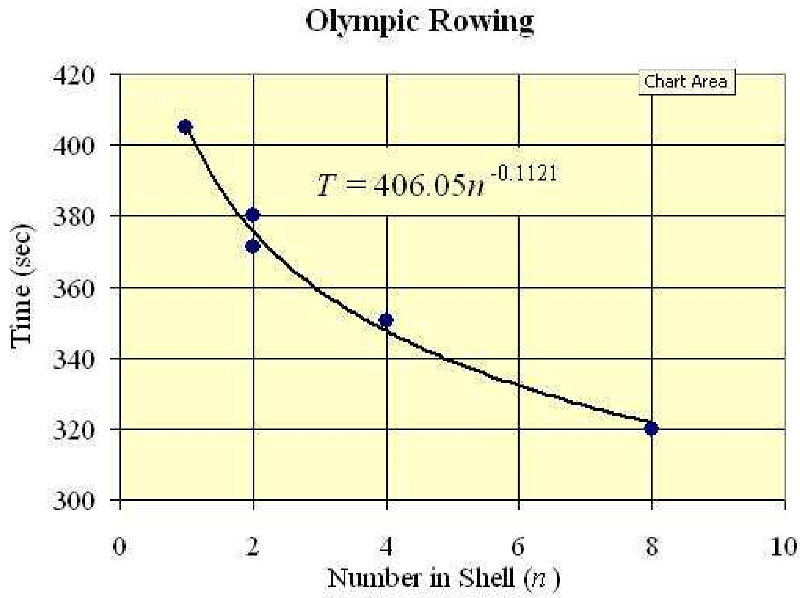


Figure 4.1: Allometric model for the winning times of Olympic rowing events.

we only change the y-axis, then we produce a **semi-log** plot. If the data falls on a straight line, then an exponential model is the best fit to the data. If we change both the x and y-axes, then we produce a **log-log** plot and a power law or allometric model is most appropriate. In Figure (4.2) we show the model of the Olympic rowing events on a log-log plot.

4.3 Computer Laboratory Exercises

1. (C2) a. Iodic Acic (HIO_3)—indexIodic Acid is a weak acid with an equilibrium constant $K_a = 0.2$. Recall from Chemistry that for a weak acid, HA, the dissociation constant K_a satisfies the relation

$$K_a = \frac{[H^+][A^-]}{[HA]},$$

which implies that for a solution of HA with a normality (formality) of x that,

$$K_a = \frac{[H^+][H^+]}{x - [H^+]}.$$

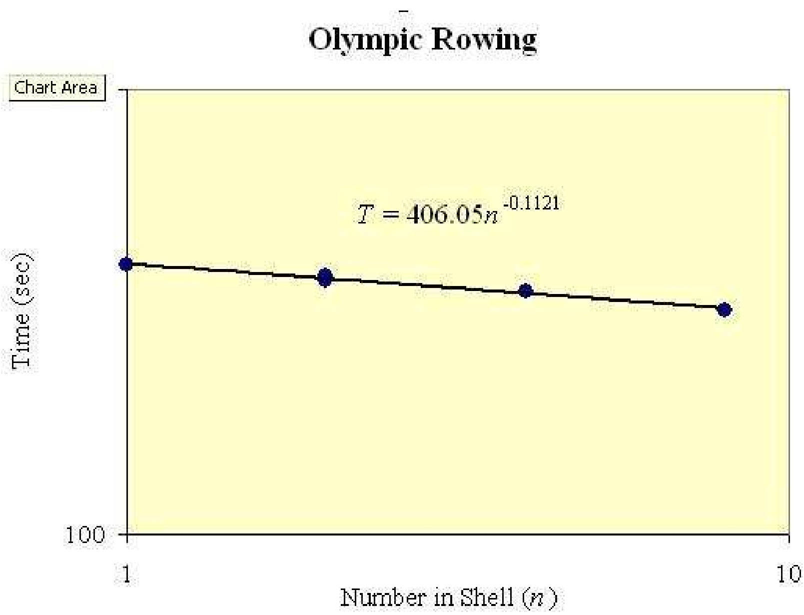


Figure 4.2: Allometric model for the winning times of Olympic rowing events on a log-log scale.

(See Chapter 4 in [1] for other examples in how to obtain K_a.) This expression can be converted into a quadratic equation in $[H^+]$ and solved with only the positive solution making sense. Find an expression for $[H^+]$ as a function of the normality, x, of the weak acid solution. This means that you need to solve the quadratic equation in $[H^+]$ using the quadratic formula, leaving x as a variable in this formula. Write the expression for $[H^+]$ in your lab report. (You should use Equation Editor in Word to write this expression.)

b. Plot a graph of the $[H^+]$ as a function of the normality x for $x \in [0.001, 2]$. Be sure to label your axes. What is the $[H^+]$ for a 0.5N solution of Iodic acid?

c. The pH of a solution is given by pH $= -\log_{10}([H^+])$. Plot a graph of the pH as a function of the normality x for $x \in [0.001, 2]$. Be sure to label your axes. What is the normality of the solution that has a pH of 1? (Hint: This question is most easily solved using Maple's **fsolve** routine.)

2. (E1) In this problem we explore the exponential, logarithmic, and power functions.

a. Consider the functions:

$$f(x) = e^x \qquad \text{and} \qquad g(x) = x^4.$$

Find all points of intersection (x and y values). (There are three of them.) Create two graphs in Excel or Maple of $f(x)$ and $g(x)$ for $x \in [-2, 2]$ and $x \in [-10, 10]$. (These graphs will help narrow the range where the points of intersection are occurring to use Maple's **fsolve** as we have done before.) Determine the intervals where $f(x) < g(x)$ and $f(x) > g(x)$. Which function dominates for large values of x?

b. Now consider the functions:

$$f(x) = \ln(x) \qquad \text{and} \qquad g(x) = x^{\frac{1}{7}}.$$

Create two graphs in Excel or Maple of $f(x)$ and $g(x)$ for $x \in [0, 5]$ and $x \in [0, 10^{10}]$. Find all points of intersection. (There are two of them.) Determine the intervals where $f(x) < g(x)$ and $f(x) > g(x)$. Which function dominates (is larger than) for large values of x?

3. (D2) This problem relates to Kepler's Third Law. In this problem you will use the power law to determine the period of revolution about or distance from the sun for all planets given information about some of the planets. Let d be the mean distance ($\times 10^6$ km) from the sun and p be the period of revolution in days about the sun. You are given the following data [3] concerning four of the planets:

Planet	Distance d	Period p
Mercury	57.9	87.96
Earth	149.6	365.25
Mars	227.9	687.0
Jupiter	778.3	4337

a. The power law expression relating the period of revolution (p) to the distance from the sun (d) is given by

$$p = kd^a,$$

where k and a are constants to be determined. Use the power law under Excel's Trendline to best fit the data above. Plot the data and the best power law fit, then have Excel write the formula on your graph. How well does the graph match the data?

b. The information given above shows that an allometric (power law) model has a straight line fitting the logarithms of data, giving

$$\ln(p) = \ln(k) + a\ln(d),$$

for the formula above. In the table above, take the logarithm of the Distance ($\ln(d)$) and the logarithm of the Period ($\ln(p)$). Use Excel's scatter plot and linear fit under Trendline to see how this fits the data. Plot a graph of the logarithm of the data and the best straight line fit to these data. Show the formula for this straight line on your graph. Compare the coefficients obtained in this manner to the ones found in Part a. How well does the graph match the data?

c. Use the power law found in Part a to complete the table below:

Planet	Distance d	Period p
Venus	108.2	
Saturn		10,760
Uranus	2871	
Neptune	4497	
Pluto		90,780

d. The Jet Propulsion Laboratory has an excellent website for astronomical data. Go to their website at http://pds.jpl.nasa.gov/planets/welcome.htm to obtain actual data on the distance and period on the planets listed in Part c, then find the percent error between your calculations in the table above and the actual values. (Note that 1 AU = 149.6×10^6 km.)

4. (D3) A collection of dogs were measured and weighed producing the following table of data [2]:

Length (cm) from nose to anus	Body Weight (gm)	Surface Area (cm^2)
51	3390	2320
100	25930	9106
62	5350	3284
76	10150	5070
74	5450	3815

a. The first model examines Weight (w) as a function of the Length (u). Use Excel's Trendline with the power law to find the best fit to the data for a model of the form:

$$w = au^k.$$

In your lab report write the best fit coefficients, a and k, found by Excel. Show a plot of this best fit curve with the data. What are the appropriate units for the coefficient a? From a biological perspective, briefly explain why the coefficient k has the value it does when fitting the data.

b. Repeat the process in Part a. for the surface area data graphing the surface area (s) as a function of length (u). In this case the model satisfies the allometric model:

$$s = au^k,$$

where the coefficients a and k differ from Part a. What are the appropriate units for this coefficient a? Do not forget to include a biological explanation along with your graph of the curve and the data.

5. (E2) Currently there is a debate on the importance of preserving large tracts of land to maintain biodiversity. Many of the arguments for setting aside large tracts are based on studies of biodiversity on islands. In this problem you apply the power rule to determine the number of species of herpetofauna (amphibians and reptiles) as a function of island area for the given Caribbean islands. You are given the following data [2]:

Island	Area (mi^2)	Species
Redunda	1	3
Montserrat	33	10
Jamaica	4,411	38
Cuba	46,736	97

a. Let N be the number of species and A be the area of the island, then the power law expression relating the number of species to the area of the island is given by

$$N = kA^a.$$

Use the power law under Excel's Trendline to best fit the data above. Plot the data and the best power law fit, then have Excel write the formula on your graph. How well does the graph match the data?

b. As in the previous problem, we want to fit a straight line to the logarithms of the data. From the allometric model above, we obtain the formula:

$$\ln(N) = \ln(k) + a\ln(A).$$

In the table above, take the logarithm of the Number of Species ($\ln(N)$) and the logarithm of the Island Area ($\ln(A)$). Use Excel's scatter plot and linear fit under Trendline to see how this fits the data. Plot a graph of the logarithm of the data and the best straight line fit to these data. Show the formula for this straight line on your graph. Compare the coefficients obtained in this manner to the ones found in Part a. How well does the graph match the data?

c. From your calculations above give estimates to fill in the table below.

Island	Area (mi^2)	Species
Saba	5	
Puerto Rico		40
Saint Croix	80	
Hispaniola		88

d. How important is maintaining a large tract of land to the maintenance of biodiversity based on this model? What does the model predict is required in increased preserved area to double the number of species supported by the environment? (Give a numerical value for the factor multiplying the area of land to achieve this doubling of species.) Write a short paragraph explaining your results as best you can.

6. (E3) The data below came from the Allegheny National Forest in Pennsylvania [4]. The issue was whether either the diameter or the height of a tree accurately predict the volume of wood in the tree. Using the data below, you are to see if there exists a meaningful relation between these variables. Thus, you want to find the volume as a function of either diameter or height. The volume is measured in board feet, the diameter is in inches, and the height is in feet.

a. The reference for this model suggests a simple linear model, so use Excel's Trendline to find the best line through the data. Graph the data and model for volume as a function of diameter, then repeat the process for volume as a function of height. Give the formula for the best straight

Diameter	Height	Volume
8.6	65	10.3
10.7	81	18.8
11.0	75	18.2
11.4	76	21.4
12.0	75	19.1
13.3	86	27.4
14.5	74	36.3
16.0	72	38.3
17.3	81	55.4
18.0	80	51.5
20.6	87	77.0

line through each of the data sets. Which graph seems to have the better predictive ability? Why is this what you would expect based on the biology of trees? What happens with both models as the diameter or height gets close to zero?

b. In this part of the problem only use the relationship between volume and the one variable that you showed in Part a. was the better predictor. Use the power law under Excel's Trendline to best fit the one set of data that best predicts the volume. Plot the data and the best power law fit, then have Excel write the formula on your graph. Can you provide any explanation for the power that you have obtained?

c. Have Excel plot a log-log plot of the data and the Trendline that you found in Part b. (This is done by editing the graph and selecting logarithmic scales for both the x and y axes, which is easily done by double clicking on the axes.) Do the data roughly fall on a straight line in this plot?

4.4 References

[1] J.M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[2] J. Mazumdar (1989), An Introduction to Mathematical Physiology and Biology, Australian Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge.

[3] Jay M. Pasachof, (1993) Astronomy: From the Earth to the Universe, Fourth Edition, Saunders College Publishing.

[4] E. K. Yeargers, R. W. Shonkwiler, and J. V. Herod (1996), An Introduction to the Mathematics of Biology, Birkhauser, Boston.

Chapter 5

Discrete Dynamical Models

Many population models are based on discrete dynamical systems . These models use the census of populations at distinct times with some functional dynamics determined by the particular biological system. An Excel spreadsheet with its pull down feature updates cells based on what occurred in the cell above, which is basically what discrete dynamical models do.

5.1 Discrete Malthusian Growth Models

There are a number of different discrete population models. A discrete model uses discrete time steps with the new population at time $n+1$ based on some function of the population and possibly the time at n. Thus, if a population at time n is given by P_n, then the population at time $n+1$ satisfies the equation

$$P_{n+1} = f(n, P_n)$$

for some function, f (also referred to as the *update function*). The example below illustrates how several discrete population models can be simulated using Excel.

Example 1: Suppose that the census of a population gives the following information, where n is the time in weeks:

Time n	0	2	5	8	12	15	20
Population P_n	20	22	24	26	29	32	34

We compare these population data to a Malthusian growth model given by

$$P_{n+1} = 1.05P_n,$$

to a nonautonomous discrete Malthusian growth model given by

$$P_{n+1} = (1.05 - 0.002n)P_n,$$

and a logistic growth model given by

$$P_{n+1} = P_n + 0.05P_n\left(1 - \frac{P_n}{50}\right).$$

The Malthusian growth model assumes a 5% growth each week, while the nonautonomous discrete Malthusian growth model assumes that initially there is a 5% growth per week, but this growth rate is declining by 0.2% each week. The logistic growth model also assumes an initial growth rate of 5% growth per week, but crowding factors from the population decrease its growth rate by $0.001P_n$ per week (for more information refer to Chapter 7 of the text book [2]).

Solution: We want to create an Excel spreadsheet that shows the simulations for each of these models, then overlays the data for comparison. The reader may obtain a copy of this Excel worksheet at

`www-rohan.sdsu.edu/~jmahaffy/courses/labs/discrete.xls`

As noted above, the fill down spreadsheet function of Excel works extremely well for this type of modeling.

We begin the spreadsheet by entering our labels for the relevant columns for the simulation. The time n is placed in Column A with the initial time set at $n = 0$ placed in **A3** as shown in Figure (5.1). In **A4**, we enter **=A3+1** to have Excel update this cell to one time unit higher than the cell above it. We use the fill down function as discussed earlier in this manual to fill in the times from $n = 0$ to $n = 20$ in **Column A**. **Column B** shows the simulation of the discrete Malthusian growth model. We enter the initial population P_0 in the cell **B3**, that is type **20** in **B3**. In **B4** we enter the formula for the Malthusian growth model. This is simulated by typing **=1.05*B3** in **B4**, then simply fill down from there to match the times in Column A. The resulting values give the populations at each corresponding time.

Column C shows a simulation of the nonautonomous discrete Malthusian growth model, which has a component of time that it must get from **Column A**. From the formula for this model, the new population is updated from knowledge about the previous population and the previous time. Again we enter the initial population P_0 in the cell **C3**, that is type **20** in **C3**. It follows that the entry for **C4** is given by **=(1.05-0.002*A3)*D3**, where **A3** gives the previous time and **D3** is the previous population. The model is simulated by filling down from there to again match the corresponding times.

Column D shows a simulation of the Logistic growth model, which is an autonomous model, meaning that the new population only depends on the earlier population and is independent of the time. The model given above begins with the same growth rate as the Malthusian model, but it has a carrying capacity of 50, so as time progresses, this population levels off at $P = 50$. Again we enter the initial population **20** in **D3**. From the formula for this model, it follows that the entry for **D4** is given by **=D3+0.05*D3*(1-D3/50)**, where **D3** is the previous population. The model is simulated by filling down from there to again match the corresponding times.

The next step is to highlight the first 4 columns and use **Chart Wizard** to create a graph. These are models, so we want to use the line option in **XY (Scatter)** plot. The appropriate labels are added and options to make the graph look better. Next edit the graph to change the scales of the axes and define the legend labels. Legend labels are readily changed by selecting **Source data** under **Chart** on the menu. By selecting the **Series** option, one can easily create appropriate labels for the graph.

The data are added by again clicking on the graph window, then going to the **Chart** on the menu bar and selecting **Add data** . Next we are careful to select **New Series** and **Categories X values**. Finally, since these last values were data, we clicked on these points (either right click or double click) and edited the style to change from line to data points. Figure (5.1) shows the spreadsheet with the simulated models in the first few columns, the data, and the graph that was produced.

5.2 Excel Solver

After creating a mathematical model, it is important to fit the model to experimental data. An earlier section showed how to compute the least

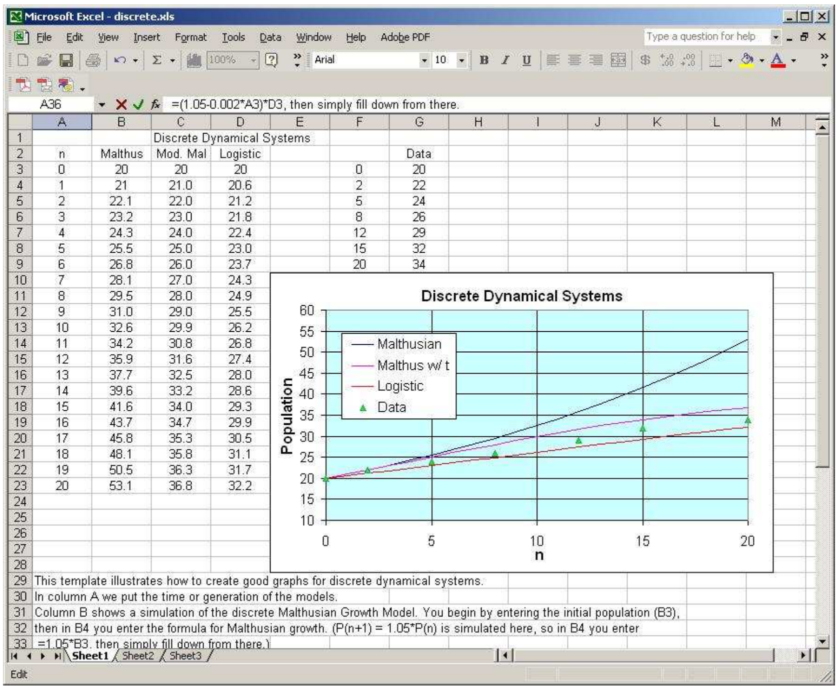


Figure 5.1: Excel spreadsheet for simulating several population models.

squares best fit of a model to experimental data. The best fit is found by finding the minimum least squares fit to the data. Excel has an **Add-In** feature called **Solver**, which provides a means to fitting models to experimental data. Below we provide an example of data for a growing culture of bacteria, which can be modeled by the logistic growth equation.

Example 2: The bacterium *Staphylococcus aureus* is a fairly common pathogen that can cause food poisoning. In Table 6.2 are data for one experiment from the laboratory of Professor Anca Segall in the Department of Biology at San Diego State University (experiments by Carl Gunderson). Standard growth cultures of this bacterium satisfy the logistic growth pattern given by the equation:

$$P_{n+1} = P_n + rP_n\left(1 - \frac{P_n}{M}\right) = P_n + f(P_n),$$

where $f(P_n)$ is the growth in each time interval. Here a normal strain is grown using control conditions and the optical density at a wavelength of 650 nm (OD_{650}) is measured to determine an estimate of the number of bacteria in the culture.

t (hours)	OD_{650}	$f(p)$
0	0.035	0.004
0.5	0.039	0.03
1	0.069	0.041
1.5	0.11	0.06
2	0.17	0.059
2.5	0.229	0.032
3	0.261	0.027
3.5	0.288	0.021
4	0.309	0.018
4.5	0.327	0.02
5	0.347	

a. Use the data above to find the best quadratic growth function $f(p)$ from Excel's polynomial fit function.

b. Simulate the logistic growth model with the growth function found above and using the initial data at $t = 0$ as a starting point. Compare this simulation to the time series data above.

c. Once again simulate the logistic growth model with parameters r, M, and P_0. Use Excel's **Solver** to find the least squares best fit to the data for *Staphylococcus aureus* given in Table (6.2).

Solution: a. The growth function, $f(p)$, is given in the last column in Table (6.2). From the logistic growth equation above, we see that this growth function has the form

$$f(p) = rp\left(1 - \frac{p}{M}\right) = rp - \frac{r}{M}p^2,$$

which is a quadratic in p. We begin by copying the table into an Excel spreadsheet with Column A being the time, t, Column B being the population measured in OD_{650}, and Column C being the growth per half hour. Note that there is one less entry in Column C as it is the difference between successive cells in Column B.

To find the best quadratic function $f(p)$, thus find the best values of r and M, we plot **Columns B** and **C**. Note that the quadratic function $f(p)$

does not have a constant term, which means that $f(p)$ must pass through the origin. The next step is to use Trendline in Excel, which can be added by either right clicking on a data point or going to the menu under **Chart** and selecting **Add Trendline**. When the Trendline menu appears, we begin by choosing the **Type** of curve to be a **Polynomial** of **Order 2**. Next we select the **Options** tab, then check the categories **Set intercept = 0** and **Display equation on chart**. After clicking on the **OK**, the quadratic function appears on the graph. We find that Excel finds the best fitting function to be

$$f(p) = -1.8439p^2 + 0.6238p.$$

It follows that the value for r is 0.6238 and $M = r/1.8439 = 0.3383$. The graph of the growth data and the best fitting quadratic through the origin and these data is shown in Figure (5.2).

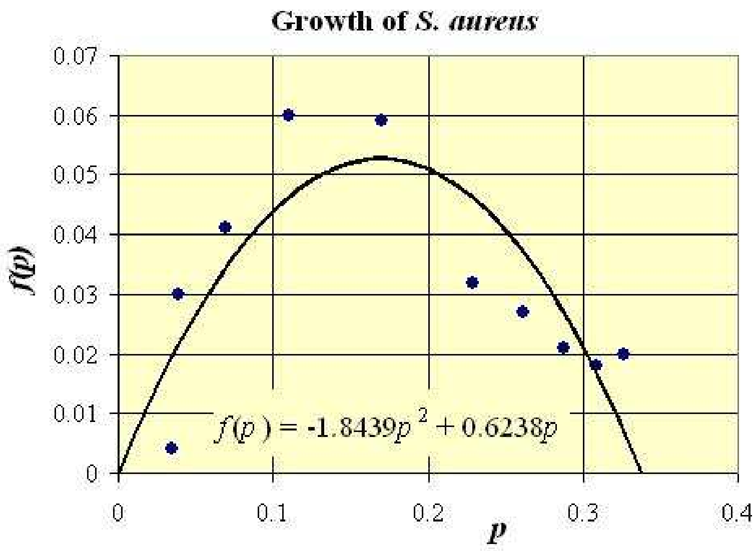


Figure 5.2: Fitting a quadratic growth function to the *S. aureus* growth experiment for Example 2.

b. With the information from Part a., we have the discrete logistic growth model given by the equation:

$$P_{n+1} = P_n + 0.6238P_n \left(1 - \frac{P_n}{0.3383}\right).$$

This is readily simulated in **Column D** in the following manner. After labeling the column in the first cell, we enter the initial value $P_0 = 0.035$ in **cell D2**. Next we enter
= D2 + 0.6238*D2*(1-D2/0.3383)
in the **cell D3**. We fill this down in the usual manner until reaching **cell D14**, which extends the model to $t = 6$ (so also extended the time in Column A).

The next step is graphing the data and the model. This is easily done by using the mouse to **highlight** Columns A, B, and D. (You highlight Columns A and B, then hold the Ctrl key while highlighting Column D.) We invoke **Chart Wizard** in Excel, selecting the **XY Scatter** as usual with grids and no labels. The data points are left as data points (the default option), while the model is modified after the graph is created by right clicking on the model points and removing the data point markers and choosing an appropriate line type. After formatting the title and axis labels to have the correct fonts (italics, subscripts, Times New Roman, etc.) and adjusting the t domain, we obtain the graph seen in Figure (5.3).

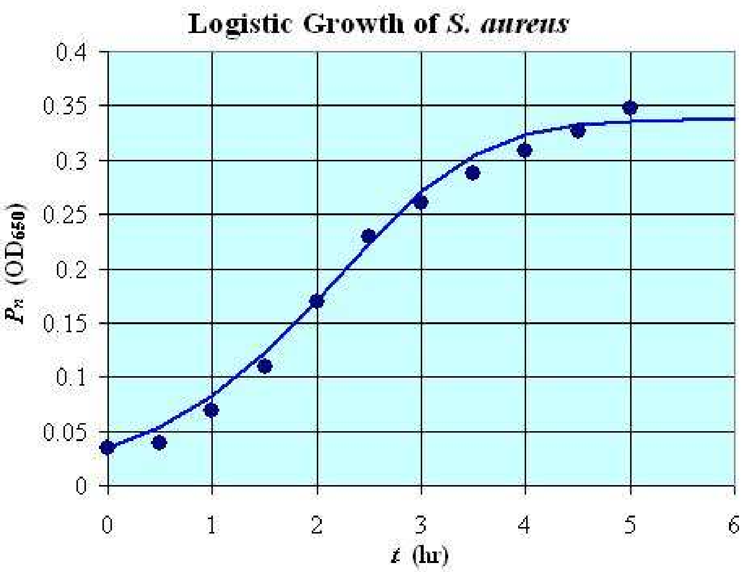
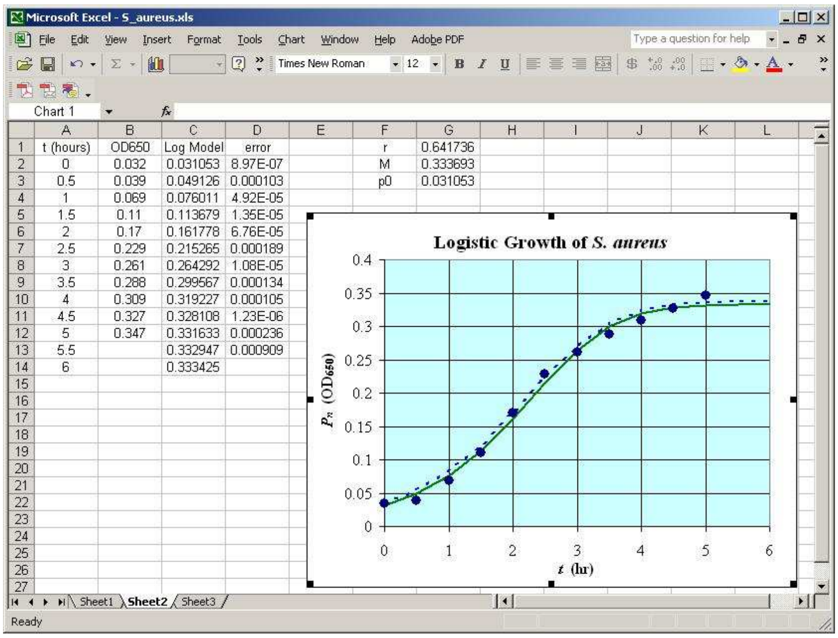


Figure 5.3: The simulation of the discrete logistic growth model for the *S. aureus* growth experiment for Example 2.

c. For the last part of this problem, we are going to introduce the **Solver** feature in Excel. This provides a powerful tool for fitting models to experimental data. Figure (5.4) shows the form of the spreadsheet that creates the best fitting discrete logistic growth model for the data in Table (6.2).



	A	B	C	D	E	F	G
1	t (hours)	OD650	Log Model	error		r	0.641736
2	0	0.032	0.031053	8.97E-07		M	0.333693
3	0.5	0.039	0.049126	0.000103		p0	0.031053
4	1	0.069	0.076011	4.92E-05			
5	1.5	0.11	0.113679	1.35E-05			
6	2	0.17	0.161778	6.76E-05			
7	2.5	0.229	0.215265	0.000189			
8	3	0.261	0.264292	1.08E-05			
9	3.5	0.288	0.299567	0.000134			
10	4	0.309	0.319227	0.000105			
11	4.5	0.327	0.328108	1.23E-06			
12	5	0.347	0.331633	0.000236			
13	5.5		0.332947	0.000909			
14	6		0.333425				

Figure 5.4: Excel spreadsheet for finding the least sum of squares best fit to the data for Example 2 for *S. aureus*.

From Figure (5.4), it can be seen that we opened a second sheet, where we copied the data from the first part of this problem. Columns A and B contain the time series data from the experiment. The next step is to enter our parameters r and M and the initial condition P_0 with the labels in **F1-F3** and the first guesses at the values in **G1-G3**. A reasonable first guess is $r = 0.62$, $M = 0.34$, and $P_0 = 0.035$. After entering these cells, we highlight **F1-G3**, then go to the top menu under **Insert** and select **Name** and **Create**. Check the box **Left column** in the **Create Names** box and select **OK**. This names our parameters for use in the model.

The next step in the modeling process is to enter the model in **Column C**. We enter the initial value in cell **C2**. This is most easily done by clicking on the cell and typing **= G3** (or typing "=" and then clicking on **G3**. The result is that 0.035 should first appear in **C2**. Next we click on

C3 and enter the model equation
= C2 + r_*C2*(1-C2/M).
You can enter the **C2**, **r_**, and **M** by simply clicking on the appropriate cells when you encounter the variable in the equation. The model is simulated by filling down from **C3** to **C14**.

To find the best fitting model, we need to find the sum of square errors between the model and the data, which we enter in **Column D**. In **D2**, we enter **= (B2 - C2)ˆ 2.** This is filled down to **D12**. In **D13**, we enter the sum of the square errors **= SUM(D2:D12)**. With the initial values entered above, this sum of square errors has the value of **0.001352**.

We need to find the smallest possible value of this sum of square errors by varying r, M, and P_0. To accomplish this we need to use Excel's **Solver**. If this Solver has not been used yet, then you need to begin with the Excel menu **Tools** and go to the **Add Ins** and check **Solver**. Now **Solver** should appear in the **Tools** list. To find the smallest possible value of the sum of square errors, we click on the cell **D13**, then select **Solver** from the **Tools** menu. The Solver window should appear with the **Set Target Cell** reading **$D$13**. On the next line, we select **Min** with a value of **0**. The next step is to enter in the window for **By Changing Cells** the values **$G$1:$G$3**. (This is most easily accomplished by highlighting the cells **G1-G3** after clicking in the window for By Changing Cells.) The final step in the process is to click on the button **Solve** in Solver. This should give the least sum of square errors In this case, the least sum of square errors is **0.000908991**, and the best possible parameter values are $r = 0.64174$, $M = 0.33369$, and $P_0 = 0.031053$. Figure (5.5) shows the best fitting discrete logistic model for this example as a solid line and the simulation from Part b appears as the dashed line.

In the next example we consider the population of the U. S. in the 20^th century from census data and find three discrete models that simulate the data. These models show a variety of computer techniques and show how different models behave.

Example 3: Table (5.2) gives the U. S. census data in the 20^{th} century with the population given in millions.

a. Find the growth rate over each decade and compute the average growth rate for the century, r. Graph the growth rates along with the average

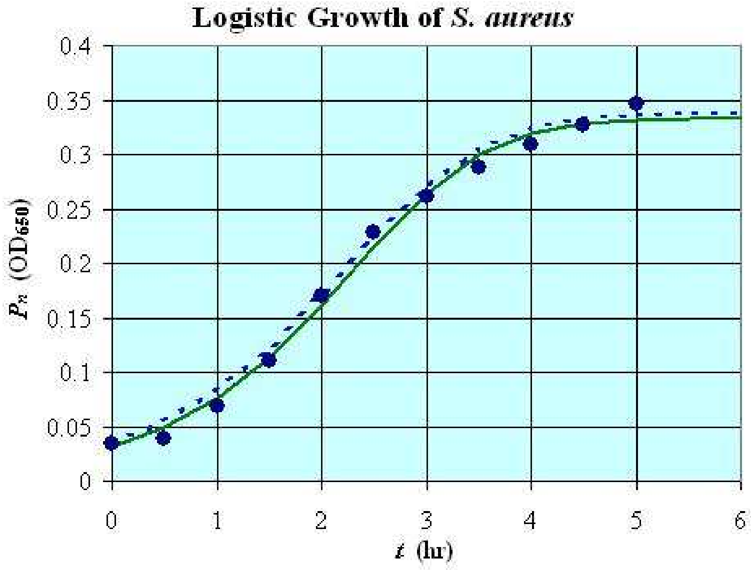


Figure 5.5: The least sum of square best fit to the *S. aureus* growth experiment for Example 2.

growth rate, then use Excel's Trendline to find the best linear fit to the growth rate.

b. Let $P_0 = 76.0$ million be the initial population for all discrete model simulations. Take the average growth rate, r, for the century and simulate the discrete Malthusian growth model

$$P_{n+1} = (1+r)P_n,$$

where P_n represents the population for each decade in the 20^{th} century. Find the percent error predicting the population in 2000, and use this model to predict the population in 2050 and 2100.

c. Create a nonautonomous discrete Malthusian growth model of the form

$$P_{n+1} = (1 + k(t_n))P_n,$$

where $k(t_n) = a\,t_n + b$ and $t_n = 1900 + 10\,n$. The constants a and b are the best linear fit to the growth rate found in Part a. Simulate this model for every decade in the 20^{th} century. Find the percent error predicting the population in 2000, and use this model to predict the population in 2050 and 2100.

Year	Population
1900	76.0
1910	92.0
1920	105.7
1930	122.8
1940	131.7
1950	151.3
1960	179.3
1970	203.3
1980	226.5
1990	248.7
2000	281.4

d. Create a logistic growth model of the form

$$P_{n+1} = P_n + r\,P_n\left(1 - \frac{P_n}{M}\right),$$

where r and M are constants representing the growth rate and carrying capacity for this model. Use Excel's Solver to find the least squares best fit of these parameters to the data in Table (5.2). Simulate this model for every decade in the 20^{th} century. Find the percent error predicting the population in 2000, and use this model to predict the population in 2050 and 2100. What does this model predict about the long term behavior of the population of the U. S.?

e. Graph all three models above. Discuss their fit to the data and predicts for the future.

Solution: a. We begin by opening an Excel spreadsheet and copying the data in **Columns A** and **B**. The complete spreadsheet used for this problem is shown in Figure(5.6).

To determine information about the growth of the population between decades we re-enter the years 1900-1990 in **Column I**. In **Column J** we find the growth rates between the decades. The growth rate, r_n, is determined by

$$r_n = \frac{P_{n+1}}{P_n} - 1,$$

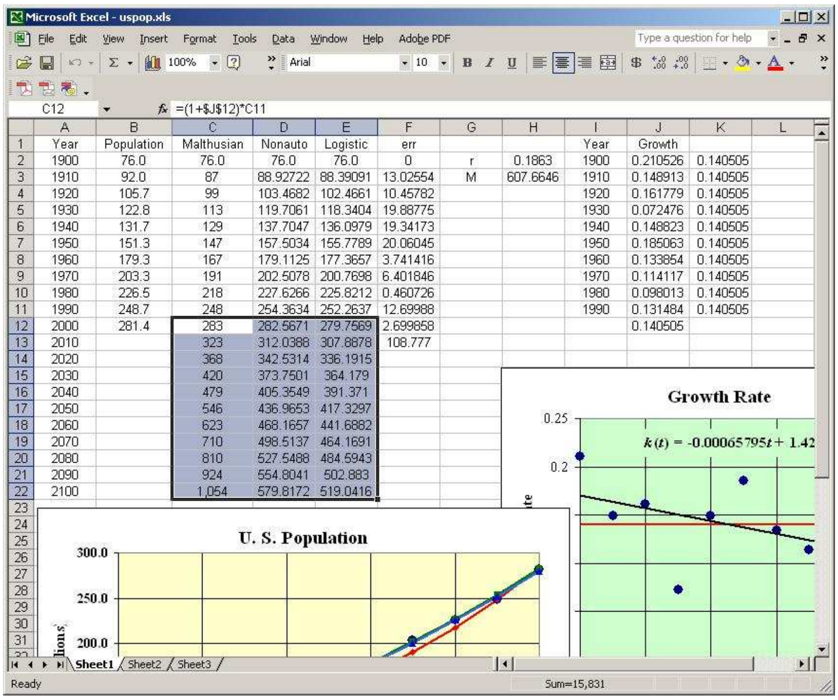


Figure 5.6: The Excel spreadsheet used for Example 3, where all calculations are show of this example.

so the growth rate in 1900 is found by taking the population in 1910 and dividing by the population in 1900 and subtracting one. Thus, in **J2** we enter = **B3/B2 - 1**, then fill down this formula to 1990, the last growth rate available with the data. In **J12**, we compute the average growth rate for the century. The average in Excel is computed here by typing in **J12** the expression = **AVERAGE(J2:J11)** or alternately, while in **J12**, right click on the **SUM symbol** in the menu bar and select **Average** and Excel will automatically select the cells **J2:J11**. The resulting average growth rate is $r = 0.140505$. For graphing purposes, we copy this value in **Column K**. Next we highlight the cells **I2:K11** and invoke **Chart Wizard**, selecting as usual the **XY (Scatter)** option and taking the default style, since the elements of **Column J** are data points. We change the data points from **Column K** into a line and make the usual changes in the domain and other style changes.

Next Excel's Trendline is used to find the best linear fit to the growth data. In order to get good results, it is very important that after you use Trendline to find the best straight line through the growth data, you double click (or right click) on the equation (be careful here). When the **Format Data Labels** menu pops up, you can choose **Number**, then you select the way you want your numbers displayed. You should use the **Scientific** notation option. Its this equation that will determine the behavior of your nonautonomous model, so you need four significant figures for an accurate model. The best linear growth function is given by

$$k(t_n) = -0.00065795\, t_n + 1.4202,$$

where $t_n = 1900 + 10\,n$. The graph of the growth rate with the best linear growth function is given in Figure (5.7).

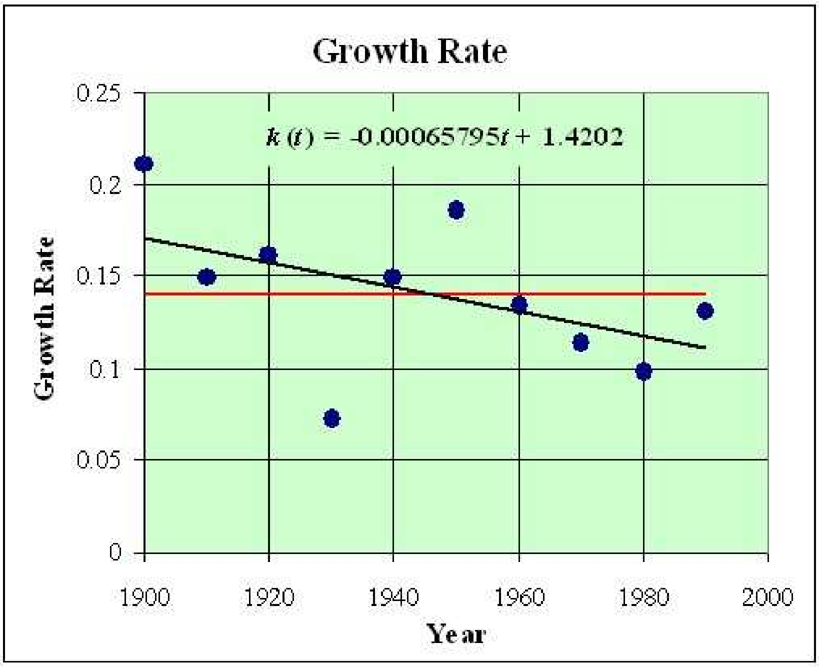


Figure 5.7: The graph showing the growth rates from the data and the best linear fit to the growth rates, which are used in the nonautonomous Malthusian growth model.

b. From Part a, we have the discrete Malthusian growth model is

$$P_{n+1} = 1.1405\, P_n.$$

This is easily simulated by letting $P_0 = 76.0$ (which is in **C2**), then inserting **= 1.1405*C2** in **C3** and filling down until 2100. The population in 2000 is predicted to be 283.0 million, which is only a 0.57% error from the actual census value. We compute the percent error by typing **=100*ABS(B12-C12)/B12**. This model predicts that the populations in 2050 and 2100 are 546.1 and 1,053.8 million, respectively, which are undoubtedly too high.

c. From Part a, we have the discrete nonautonomous Malthusian growth model is given by

$$P_{n+1} = (2.4202 - 0.00065795\, t_n)P_n,$$

where $t_n = 1900 + 10\, n$. This is easily simulated by placing $P_0 = 76.0$ in **D2**, then inserting **= (2.4202 - 0.00065795*A2)*D2** in **D3** and filling down until 2100. The population in 2000 is predicted to be 282.3 million, which is only a 0.41% error from the actual census value. This model predicts that the populations in 2050 and 2100 are 437.0 and 579.8 million, respectively.

d. The discrete logistic growth model is given by

$$P_{n+1} = P_n + r\, P_n \left(1 - \frac{P_n}{M}\right),$$

where r and M are constants representing the growth rate and carrying capacity for this model. We begin this model simulation by guessing values of $r = 0.21$ and $M = 400$, which come from the highest growth rate seen in the data from Part a and an arbitrary carrying capacity that is above the census value in 2000. From Figure (5.6), we place the names of the parameters in **G2** and **G3**. The values are placed in **H2** and **H3**. As usual, we name these variables by going to the main menu under **Insert** and **Name**, then selecting **Create** and taking the default setting.

Now the logistic growth model is easily simulated by placing $P_0 = 76.0$ in **E2**, then inserting **= E 2+ r_*E2*(1 - E2/M)** in **E3** and filling down until 2100. The next step is to compute the sum of square errors, which is done in **Column F**. In cell **F2**, we compute the individual error by typing **= (B2 - E2)^2**, then filling down to **F12**. In **F13**, we sum the square errors in **F2:F12**, then we apply Excel's **Solver** as we did in the previous example. Highlight **F13**, then invoke **Solver** by going to the main menu under **Tools**. We select **Min**, and in the box for **By changing cells** we select the cells **H2:H3**. After clicking on the button to **Solve**, we have the least sum of square error for this model, which is **108.78**. This changes the parameters to $r = 0.1863$ and $M = 607.66$.

The population in 2000 is predicted to be 279.8 million, which is a 0.58% error from the actual census value. This model predicts that the populations in 2050 and 2100 are 417.3 and 519.0 millions, respectively. These values are lower than either of the previous two models. Since the carrying capacity $M = 607.7$, this model predicts that the population of U. S. will level off at 607.7 million in the future.

e. A graph of all the models is presented in Figure (5.8). It is readily apparent that all three models simulate the data quite well. It is not surprising that the worst model is the simple Malthusian growth model, yet it still follows the data quite closely. Its maximum error is 6.7% in 1960. The discrete nonautonomous Malthusian and logistic growth models are extremely close to the data points for all of the 20^{th} century, which makes it hard to determine which model might be better. From a mathematical perspective, the logistic growth model is slightly better as its sum of square errors compared to the census data is only 108.8, while the sum of square errors for the nonautonomous Malthusian growth model is 133.9. Still these values are close. From a biological perspective, each model has its merits and faults to argue why it might be the better model. The logistic model is how most simple organisms behave, which is to say that crowding factors from the population limit growth independent of time. However, human populations have technology, which is a time varying factor. Improved medicine and better education, especially of women, results in better family planning and smaller families, which in turn slow the growth of the population. This argues for the time varying growth seen in the nonautonomous Malthusian growth model. These are questions that should be studied in more detail in more advanced ecology or mathematical modeling classes.

5.3 Maple and Linear Discrete Models

Maple has the ability to solve simple discrete dynamical models. In this section, a basic model for breathing an inert gas is solved using Maple, showing the necessary commands.

Example 4: The average healthy adult male exchanges about 16% of his lung capacity with each breath. The normal concentration of helium (He) in the atmosphere is only about 5 ppm (parts per million). Suppose that an in-

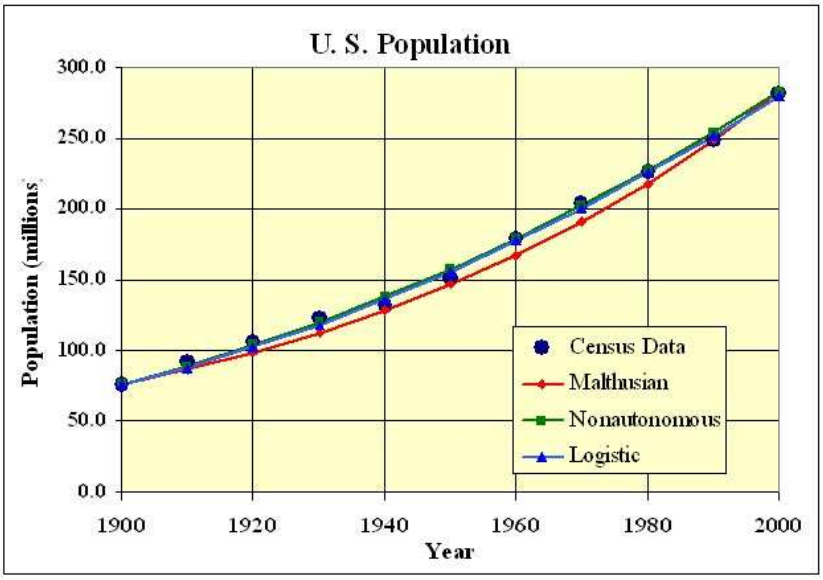


Figure 5.8: The graphs of the census data of the U. S. in the 20^{th} century with discrete Malthusian, nonautonomous Malthusian, and logistic growth models simulated.

dividual starts breathing an enriched mixture that brings the concentration of He in his lungs to 400 ppm, then he breathes normally regular atmospheric air. Show a graph of the concentration of He in his lungs for the first 10 breaths and give the general solution for the concentration of He in his lungs.

Solution: The general breathing model for this individual is given by the equation:

$$c_{n+1} = 0.84c_n + 0.8.$$

Details for deriving this model can be found in the accompanying text by Mahaffy and Chávez-Ross[2]. This model is easily simulated in Excel by putting the breath number in **Column A**, then starting with $c_0 = 400$ in the cell **B2**, we enter in **B3** the formula = 0.84*B2 + 0.8 and pull down to find the concentration of He in the first 10 breaths. A graph of the first 10 breaths is shown in Figure (5.9).

The breathing model can be solved exactly using Maple's fsolve (recurrence relation solver). The Maple commands for solving this problem are shown below.

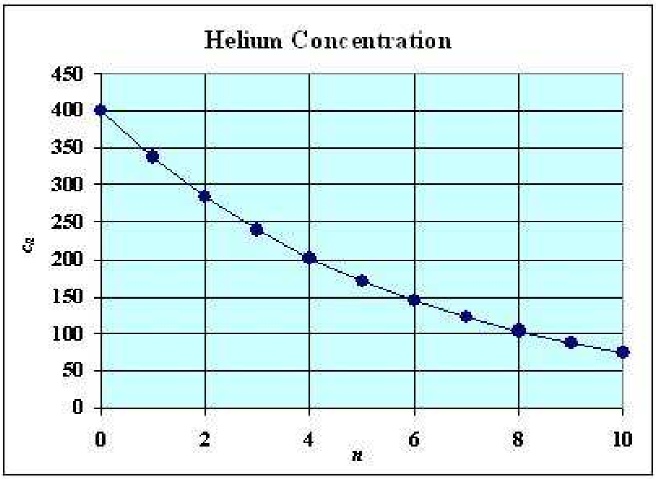


Figure 5.9: Graph showing the concentration of helium in an adult male breathing normally after having breathed an enriched helium mixture.

```
> eqn := c(n+1) = 0.84*c(n) + 0.8;
> rsolve({eqn, c(0) = 400}, c);
```

The output of the rsolve command is

$$395\left(\frac{21}{25}\right)^n + 5$$

5.4 Computer Laboratory Exercises

To simplify listing of web addresses below, we note that the beginning of the web address begins with the following:

`www-rohan.sdsu.edu/~jmahaffy/courses`

which will be given as **...** in the text below.

1. (F2) The population of Canada [3] was 24,070,000 in 1980, while in 1990 it was 26,620,000. The population in Kenya [3] was 16,681,000 in 1980, while it was 24,229,000 in 1990.

a. Over a limited range of years, the population P_n of most countries can be estimated using the Malthusian growth law, which is given by:

$$P_{n+1} = (1+r)P_n,$$

where n is the number of years since 1980 with P_0 the population in 1980. Find the general solution for this equation, writing an expression for the population P_n for each of these countries with the appropriate values of both P_0 and r from the data given to you. What does the value of r represent? (Don't forget that the data are at 1980 and 1990, while n is in years.)

b. Find how long it takes for each of their populations to double.

c. Find when the population of Canada is equal to the population of Kenya. Graph the populations of both countries between 1980 and 2030.

d. Assuming the populations continue to grow according to the Malthusian growth law above, then determine the populations of these countries in the years 2000, 2050, and 2100. Create a table showing these values.

2. (I3) The shape of a cell affects its surface area to volume ratio. This can be significant in the cell's ability to absorb nutrients or survive toxins. You are given that the volume of a sphere and cylinder are $\frac{4}{3}\pi r^3$ and $\pi r^2 h$, respectively, where r is the radius and h is the height. The surface area for a sphere and a cylinder are $4\pi r^2$ and $(2\pi rh + 2\pi r^2)$, respectively.

a. Complete the following table, which examines cellular geometry. Note

Organism	Shape	Diam mm	Height mm	Volume mm^3	Surface mm^2	S.A.:Vol. Ratio
Mycoplasm	Sphere	0.3	–			
Coccus	Sphere	1.5	–			
E. coli	Cylinder	0.75	4.0			
Yeast	Cylinder	5.0	8.0			
Diatom	Cylinder	20	60			

that the diameter and not the radius is given.

b. Suppose that the *Coccus* bacteria and *E. coli* satisfy the discrete Malthusian growth equation

$$Pn + 1 = (1 + k)P_n, \qquad P_0 = 1000,$$

where P_0 is the initial population and the doubling time for the population is 25 min. Find the value of k and write the general solution, then determine how long it takes for the total surface area of each of these growing populations to reach 1 m^2. (Recall that 1 μm $= 10^{-6}$ m.)

c. Assume the same population dynamics as given above. Determine how long it takes for each of the populations to grow to where their volumes occupy 1 cm^3. (Recall that 1 cm $= 10^{-2}$ m.)

d. Michael Crichton in the Andromeda Strain (1969) states that "A single cell of the bacterium *E. coli* would, under ideal circumstances, divide every twenty minutes... [I]t can be shown that in a single day, one cell of *E. coli* could produce a super-colony equal in size and weight to the entire planet Earth." The diameter of the Earth is 12,756 km, so assuming it is a perfect sphere, determine how long it takes for an ideally growing colony of *E. coli* (doubling every 20 min with the volume you computed above) to equal the volume of the Earth. (Don't forget that 1 km = 1000 m. Also, you have to find a new value of k and start with $P_0 = 1$.) How does your answer compare to the statement of Michael Crichton?

3. (F4) Using data from the U. S. census bureau, the table below presents the population (in millions) for France. This lab has you repeat for this country the modeling effort that we performed in class for the U. S.

Year	Population
1950	41.83
1960	45.67
1970	50.79
1980	53.87
1990	56.74
2000	59.38

a. Find the growth rate for each decade with the data above by dividing the population from one decade by the population of the previous decade and subtracting 1 from this ratio. Associate each growth rate with the earlier of the two census dates. Determine the average (mean) growth rate, r, from the data above. Associate t with the earlier of the dates in the growth ratio, and use Excel's Tendline to find the best straight line

$$k(t) = a + bt$$

through the growth data. Graph the constant function r, $k(t)$, and the data as a function of t over the period of the census data. It is very important that you click on the Trendline equation and reformat the coefficient b so that it has more significant figures (obtain 4 significant figures for a and b).

b. The Discrete Malthusian growth model is given by

$$P_{n+1} = (1 + r)P_n.$$

where r is computed in Part a. and P_0 is the population in 1950. Write the general solution to this model, where n is in decades. Use the model to predict the population in 2020 and 2050.

c. The revised growth model is given by

$$P_{n+1} = (1 + k(t_n))P_n.$$

where $k(t_n)$ is computed in Part a. and P_0 is again the population in 1950. Simulate this nonautonomous discrete dynamical model from 1950 to 2050. (Note that $t_n = 1950 + 10n$.) Use the model to predict the population in 2020 and 2050.

d. Create a table listing the date, the population data, the predicted values from the Malthusian growth model, the nonautonomous dynamical model, and the percent error between the actual population and each of the predicted populations from the models from 1950 to 2000. What is the maximum error for each model over this time interval? Use Excel to graph the data and the solutions to the each of the models above for the period from 1950 to 2050. Briefly discuss how well these models predict the population over this period. List some strengths and weaknesses of each of the models and how you might obtain a better means of predicting the population.

e. The growth rate of the nonautonomous dynamical model goes to zero during this century for France. At this time, this model predicts that the population will reach its maximum and start declining. Use the growth rate $k(t)$ to find when this model predicts a maximum population, then estimate what that maximum population will be.

4. (F1) A Malthusian growth model for the U. S. population can be found at the website

`.../s00a/math121/labs/labf/q1v1.htm`

At that website there is an applet that allows the user to adjust the growth rate, r, and the range of years to try to fit a discrete Malthusian growth model. This problem examines a version of that applet in more detail. In this problem you will be given a couple ranges of years to try to fit with a Malthusian growth model, then you will find the least squares fit to the census data by adjusting the parameter, r. Finally, you will use your model to compare to other census data.

a. Take the applet found at the website

`.../s00a/math121/labs/labf/q1v1.htm`

and adjust the data range to go from 1800 to 1880. Next adjust the parameter, r, until you obtain the smallest value of the sum of squares error (the least squares fit to the data). Write this value of r and the sum of squares error in your lab report and give the discrete Malthusian growth model for this set of data. Note that the initial value, P_0, of the data agrees with the population at 1800. Use Excel to graph the data and the solution to your model for the range from 1800 to 1900.

b. Use your model to predict the population in 1860, 1890, and 1900. Find the percent error between the actual census data and these predictions. Write a short paragraph describing how well this model works on the predictions for these dates, and briefly describe any discrepancies that you observe on the graph between the model and the data and put these errors in the context of what you know about U. S. history.

c. Repeat the process from Parts a. and b. but use the range 1840 to 1920 with the initial population, P_0, of the data agrees with the population at 1840. Use the model to predict the populations in 1870, 1930, and 1940. Which range of data provides the better information for predicting next two decades and why?

5. (G1) This problem examines discrete Malthusian and logistic growth models, which are appropriate for studying simple organisms over limited time periods. The Malthusian growth model is given by the equation:

$$B_{n+1} = B_n + rB_n = (1+r)B_n,$$

where n is the time in minutes and r is the rate of growth. The Logistic growth equation is given by

$$B_{n+1} = B_n + rB_n\left(1 - \frac{B_n}{M}\right),$$

where M is the carrying capacity of the population.

a. Begin with a simulation of the Malthusian growth model starting with 1000 bacteria (or $B_0 = 1000$). Assume that the growth rate $r = 0.024$/min. Write an expression for the number of bacteria at each min. Simulate this dynamical system, then create a table with the number of individuals at $n = 1$, 3, and 5 hr (60, 180, and 300 min). How long does it take for this population to double?

b. Next we examine a population of bacteria that satisfies the logistic growth law. Start again with $B_0 = 1000$ bacteria, but use a growth rate of $r = 0.029$/min. Assume that $M = 1,000,000$. Simulate this model for 300 min, then create a table with the number of individuals after $n = 1$, 3, and 5 hr (60, 180, and 300 min). How long does it take for these bacteria to double? (Note that in this case since you do not have a formula to find the doubling time, you will have to use your output from the simulation.)

c. On a single graph plot the populations of both bacterial cultures (Malthusian and logistic) for n from 0 to 300. (Be sure to use lines to represent these simulations and not points, labeling which line represents which model.) Use your data to determine the first time that the population from Malthusian growth model exceeds the one growing according to the Logistic growth model.

6. (G2) If c_n represents the concentration of the inert gas argon (Ar) in the lungs, then a mathematical model for breathing is given by the discrete dynamical model

$$c_{n+1} = (1 - q)c_n + q\gamma,$$

where q is the fraction of the lung volume exchanged with each breath and $\gamma = 0.0093$ (fraction of Ar in dry air) is the concentration of Ar in the atmosphere. Normal breathing usually exchanges a volume of air, known as the tidal volume, V_i. The space remaining in the lung after exhaling from a normal breath is known as the functional residual volume, V_r. The fraction of air exchanged

$$q = \frac{V_i}{V_i + V_r}.$$

a. Assume that a normal subject breathes an enriched mixture of air that contains 10% Ar, so that $c_0 = 0.1$ (fraction of Ar in dry air). Suppose that the tidal volume is measured at $V_i = 520$ ml for this subject, while another measurement gives the functional residual volume, $V_r = 2400$ ml. Make a table and create a graph showing the concentration of Ar in the first 10 breaths. Determine how many breaths are required until the concentration of Ar drops below 0.01. (Hint: You may want to fill down for 40-50 breaths to be sure that the argon level drops below the desired value.)

b. A patient with emphysema is given the same mixture of Ar (so again $c_0 = 0.1$ (fraction of Ar in dry air). The tidal volume for this patient is measured at $V_i = 210$ ml. The concentration of Ar in the first breath in found to contain 0.0897 (fraction of Ar in dry air) for this patient or $c_1 = 0.0897$. Find the fraction of the lung volume exchanged q and the functional residual volume, V_r.

c. For the emphysema patient in Part b., use the value of q that you found to simulate the discrete lung model for 10 breaths. Make a table and create a graph showing the concentration of Ar in the first 10 breaths. Determine how many breaths are required until the concentration of Ar drops below 0.01.

d. What do these results tell you about differences between the breathing of a normal subject and a patient with emphysema?

7. (G3) This problem extends the Malthusian growth law to include immigration or emigration of the population.

a. Suppose a population of organisms satisfies the Malthusian growth law with immigration

$$A_{n+1} = rA_n + m_a,$$

where n is the number of years, $r = 1.15$ is the annual growth rate (15% per year), and $m_a = 200$ is the yearly number of immigrants. Suppose the initial population $A_0 = 100,000$. Simulate this model for 10 generations, $n = 1, ..10$. List the populations at $n = 1$, 2, 5, and 10, and graph your solution.

b. This equation would be difficult for you to solve exactly. However, with the help of Msple's fsolve command, we can solve this discrete dynamical system. The solution satisfies:

$$A_n = A_0 r^n + \frac{m_a(r^n - 1)}{r - 1}.$$

Verify this solution agrees with your results in Part a. at $n = 5$ and 10. Use this solution to determine how long it takes for the population to double., *i.e.*, find n such that $A_n = 2A_0$.

c. Suppose another population of organisms satisfies the Malthusian growth law with emigration

$$B_{n+1} = qB_n - m_b,$$

where n is the number of years, $q = 1.13$ is the annual growth rate, and $m_b = 200$ is the yearly number leaving the region. Suppose the initial population $B_0 = 200,000$. Again simulate this model for 10 generations, $n = 1,..10$. List the populations at $n = 1$, 2, 5, and 10.

d. Take into account the sign change for emigration (or use Maple) and find the solution to the Malthusian growth model with emigration. How long does it take for this population to double?

e. Use the solutions from Parts b. and d. to graph the populations A and B on a single graph for $n = 0,..,50$. Find how long it takes until Population A is equal to Population B (*i.e.*, find the value of n when the populations are equal) and give the population at that time. (Hint: To find when the populations are equal, use Maple's fsolve routine. Enter your two functions of n, A_n and B_n, then set them equal to each other and solve (using fsolve) for n. You will probably need to tell Maple to search for the solution for $n = 0..500$ in the fsolve command.)

8. (H1) Carlson (1913) [1] grew yeast in laboratory cultures and collected data every hour for 18 hours. The list below gives the population (p) at representative times (t) and the change in population over the previous hour, $f(p)$.

t	p	$f(p)$
1	9.6	8.7
4	47.2	23.9
7	174.6	82.7
10	441.0	72.3
13	594.8	34.6
16	651.1	4.8

The growth of this yeast population satisfies a quadratic discrete dynamical system or logistic growth model.

a. In the first part of this laboratory exercise you will use the data above to find the rate of growth of the yeast. In particular, you want to use Excel's Trendline polynomial fit of order two with the y-intercept set to zero through the data (last 2 columns)

$$f(p) = a_2p^2 + a_1p.$$

(Note that you ignore the times listed in the table when you find $f(p)$.) Since this is a quadratic equation, you can find the p-intercepts and the vertex. Write these values in your report. Show a graph of the data and the best quadratic $f(p)$ passing through the data.

b. The growth function that you found in Part a. can be used to simulate the growth of the yeast using a discrete logistic model. The dynamical system for the yeast population is given by the following model:

$$pn + 1 = p_n + f(p_n),$$

where $f(p_n)$ is the best quadratic function found above. Use your initial population ($p_0 = 9.6$) starting at $t = 1$ and simulate the growth for 20 hours (20 iterations). (Note you are NOT starting at $t = 0$.) List the populations at times $t = 5$, 10, 15, and 20. Plot both the data from your simulation and the data given in the table above. (Note that this time you need to use only the time data and the population data in the table.) Discuss how well your simulation matches the data in the table.

c. Compute the error between the model and the data at times 7 and 16. What does this model say happens to the population of yeast for large t? Find all equilibria for this model and discuss the stability of these equilibria.

d. Use Excel's **Solver** to find the least squares best fit to the discrete logistic growth model by varying the parameters r, M, and the initial condition P_0. Give the best fitting values to these parameters and show a graph of the best fitting model with the data.

9. (H2) The general discrete logistic growth model is given by the equation:

$$P_{n+1} = f(P_n) = P_n + rP_n\left(1 - \frac{P_n}{M}\right).$$

This problem explores some of the complications that can arise as the parameter r varies.

a. Let $M = 5,000$. The first step in studying this model is to find all equilibria (where the population stays the same). Determine the equilibria by solving

$$P_e = f(P_e) \qquad \text{for} \qquad P_e.$$

b. Let $r = 1.89$ with $P_0 = 2,000$. Simulate the discrete dynamical system for 50 generations. Make a table listing the population for every fifth generation (P_0, P_5, P_{10}, ..., P_{50}). Graph the solution of the dynamical system and write a brief description of what you observe in your solution.

c. Repeat the process in Part b. with $r = 2.1$ and $r = 2.62$. (You can make a separate table for these simulations or simply add these to your table in Part b with appropriate labeling.) Don't forget to write a description of these solutions and how they compare to each other and your solution in Part b. What behavior do you observe for the solution in relation to the larger of the two equilibria? (Hint: Observe the values closer to when $n = 50$, so if you see that the model oscillates between two values, then you would say that for this parameter value the model has "Period 2" and list the values it oscillates between.)

d. Find a parameter value that gives you an oscillation with period 3. This always implies that the dynamical system has gone through chaos. Show your simulation that gives the period 3 oscillation. (Hint: You may want to use the applet at

`.../s00a/math121/lectures/logistic_growth/logistic.html`

search for the period 3 behavior.)

10. (H3) Since the new census numbers are out, we want to highlight the modeling of the U. S. census. This question examines three models for studying the population of the U. S. during the 20^{th} century. Below is a table of the U. S. census data for the 20^{th} century.

Year	Population	Year	Population	Year	Population
1900	75,994,575	1940	131,669,275	1980	226,545,805
1910	91,972,266	1950	151,325,7985	1990	248,709,873
1920	105,710,620	1960	179,323,175	2000	281,421,906
1930	122,775,046	1970	203,302,031		

a. The average growth rate for the 20^{th} century is 14.15%. Use the discrete Malthusian growth model

$$P_{n+1} = (1 + r)P_n$$

with $P_0 = 75,994,575$ and $r = 0.1415$ to simulate the population from 1900 to 2020. Make a table showing the actual population, the simulated values, and the percent error between the model and the actual data for the years 1930, 1950, 2000. What is the maximum percent error (in absolute value) for this model (over the range simulated) and when does it occur? Compute the average percent error (in absolute value) between the actual data and the model for the dates from 1910 to 2000.

b. Throughout U. S. history, immigration has played an important role. During the 20^{th} century, it has been tightly regulated and maintained a relatively constant value. Suppose that the immigration rate is $\mu = 3,200,000$ people per decade. The discrete Malthusian growth model with immigration is given by

$$P_{n+1} = (1+r)P_n + \mu,$$

where $P_0 = 75,994,575$ and $r = 0.1150$. Simulate this model from 1900 to 2020, and make a table showing the actual population, the simulated values, and the percent error between the model and the actual data for the years 1930, 1950, 2000. What is the maximum percent error (in absolute value) for this model (over the range simulated) and when does it occur? Compute the average percent error (in absolute value) between the actual data and the model for the dates from 1910 to 2000.

c. The two previous models grow without bound. One question is where the U. S. population will level off, and there are many estimates on what this might be. We studied the logistic growth model and found that it has this property of leveling off at the carrying capacity of the population. A study indicates that a good logistic growth model for the population of the U. S. in the 20^{th} century is given by

$$P_{n+1} = 1.21P_n - 4.773 \times 10^{-10}P_n^2.$$

Again let $P_0 = 75,994,575$ and simulate this model from 1900 to 2020. Make a table showing the actual population, the simulated values, and the percent error between the model and the actual data for the years 1930, 1950, 2000. What is the maximum percent error (in absolute value) for this model (over the range simulated) and when does it occur? Compute the average percent error (in absolute value) between the actual data and the model for the dates from 1910 to 2000. Compute the carrying capacity for this model. Also, determine how long, according to this model, it will be until the population reaches 90% of the carrying capacity.

d. Graph all three models and the census data on the interval 1900 to 2020. Looking at the three models above, determine which model you believe best predicts the population for the years 2010 and 2020. Which model do you believe is the best and why? Describe two ways that you could improve the best model to make a better prediction for either the 2010 census or determining the carrying capacity for the U. S.

5.5 References

[1] T. Carlson (1913), Über Geschwindigkeit und Grösse der Hefevermehrung in Würze. Biochem. Z. 57: 313-334.

[2] J. M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[3] Statistical Abstract of the U. S. (1995), The National Data Book, 115th ed., U. S. Department of Commerce, Bureau of the Census.

Chapter 6

Derivatives

The last chapter examined discrete population models, where population counts are measured at discrete times. As the time intervals of the population census shrink, one obtains the instantaneous rate of change of population. This is the derivative of the population function. This section uses a number of computer techniques for studying the derivative and applying it to practical problems. For more information on the topic refer to Chapters 9-11 of the text book [5].

The process of differentiation is very mechanical, which makes this operation well-suited for symbolic algebra programs. In this chapter we introduce a few Maple commands to show how Maple can very efficiently and accurately compute the derivative.

6.1 Differentiation using Maple

The derivative of a function at a particular point is described geometrically as the slope of the tangent line to the function at the given point. By creating a sequence of secant lines with the points on the curve becoming closer to a single point, one approximates the tangent line. The limit of the sequence of secant lines produces the tangent line, and the slope of this tangent line is the derivative at the given point. Formally, we find the slope of the tangent line at a point $(x_0, f(x_0))$ by looking at the secant line through the points $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$, then letting h tend to zero. The slope of this secant line is

$$m_h = \frac{f(x_0 + h) - f(x_0)}{h}.$$

This can be algebraically very difficult or impossible. The definition of the derivative is the limit of this expression as $h \to 0$, so creating a sequence of secant lines provides intuition to understanding the definition of the derivative of a function.

In the example below, we use the power of Maple's symbolic algebra to find the slope of the secant lines and simplify the process. We also demonstrate the Maple **diff** command to find the derivative very easily.

Example 1: Consider the function

$$f(x) = \frac{1}{(x+3)^3}.$$

a. Find the slope of the secant line through the points $(-2, f(-2))$ and $(-2+h, f(-2+h))$ for $h = 0.5$, 0.1, and 0.01, then determine the equations of the three secant lines (in slope-intercept form). Use Excel to graph $f(x)$ and the three secant lines for $x \in [-2.95, 0]$. Limit the range, so that y is in the interval $[-10, 20]$.

b. Find the general formula for the slope of the secant line through $x = -2$ and $x = -2 + h$ in simplest form. Find the slope of the tangent line by taking the limit as $h \to 0$. Give the equation of the tangent line at $x = -2$. What is the derivative of $f(x)$ at $x = -2$. Use Maple graph $f(x)$ and the tangent line for x in the interval $[-3, 0]$.

Solution: Our calculations center about the point $(-2, f(-2)) = (-2, 1)$. We can readily use an Excel spreadsheet to compute the function values and the slopes for each of the secant lines. In Column A, we insert the x-values -2, -1.5, -1.9, and -1.99. In Column B, we calculate the function values for each of the corresponding x-values giving $f(-2) = 1$, $f(-1.5) = 0.296296$, $f(-1.9) = 0.751315$, and $f(-1.99) = 0.970590$. In Column C, we compute the slopes

$$\begin{aligned}
\frac{f(-1.5) - f(-2)}{0.5} &= -1.40741, \\
\frac{f(-1.9) - f(-2)}{0.1} &= -2.48685, \\
\frac{f(-1.99) - f(-2)}{0.01} &= -2.940985.
\end{aligned}$$

Since the point slope form of the line satisfies $y - y_0 = m(x - x_0)$, it follows that the y-intercept, b, is given by $b = y_0 - mx_0$, which for this example is

$b = 1 + 2m$. It follows that the secant line for $h = 0.5$ is

$$y = -1.4074x - 1.8148,$$

and for $h = 0.1$ is

$$y = -2.48685x - 3.9737,$$

and for $h = 0.1$ is

$$y = -2.9410x - 4.8820.$$

In Figure (6.1) is a graph of $f(x)$ and the three secant lines computed above.

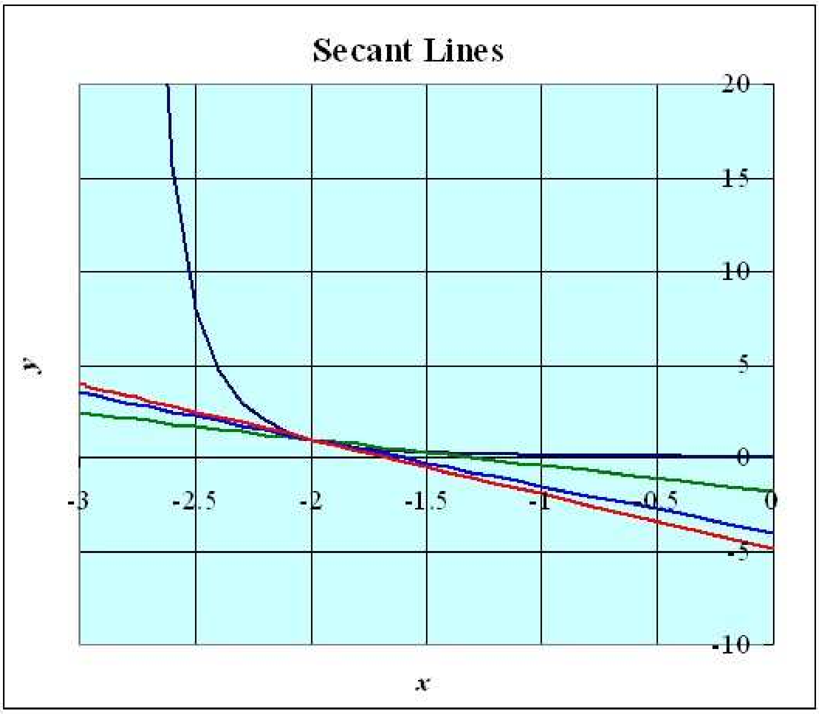


Figure 6.1: Graph of the function $f(x)$ and 3 secant lines.

b. We use Maple to find the general formula for the slope of the secant line in simplest form. First, you enter the function in the standard way for Maple:

```
> f : = x -> 1/(x+3)^3;
```

To which Maple responds

$$f := x \to \frac{1}{(x+3)^3}.$$

The slope of the secant line is found by the following:

```
> (f(-2+h) - f(-2))/h; simplify(%);
```

To which Maple responds

$$\frac{(1+h)^{-3} - 1}{h},$$

and

$$-\frac{3 + 3\,h + h^2}{(1+h)^3}.$$

The simplify command lets Maple find the easiest form of the expression that it chooses. (This may or may not be the best form to analyze.) You may be able to see from this expression the slope of the tangent line by letting $h = 0$.

Formally to find the slope of the tangent line, we let h tend to zero. This is the derivative of $f(x)$ at $x = -2$.

```
> m := limit(%, h = 0);
```

To which Maple gives the answer **m := −3**. The **m :=** is simply used to define the slope for future use. Maple has the **limit** command to take limits, which is mathematically a very difficult process. The **%** in Maple means to take the previous expression and place in the position with the %. (This can be a dangerous command to use, as it works on whatever you just did, not necessarily from the line above. If you do anything else, then use the cursor to return to a line with the %, then you may get the wrong result.)

Next we define tangent line (the **tline(x);** following the function definition has Maple print out the equation of the tangent line), then plot both the function and the tangent line. (You will need to adjust the x and y values in the plot command to get the desired graph for your problem.)

```
> tline := x -> m*(x - (-2)) + f(-2); tline(x);
```

Maples responds with

$$tline := x \to m(x+2) + f(-2),$$

$$-3\,x - 5.$$

Thus, the tangent line is

$$y = -3x - 5.$$

To plot the tangent line and the function using Maple (restricting the range of the y values), we enter the command

```
> plot({f(x), tline(x)}, x = -3..1, y = -10..20);
```

The last part of this example shows you how easily you can get a derivative using Maple. This is the way that many mathematicians actually do research when they need to accurately differentiate complicated expressions. To differentiate the function $f(x)$ given in this example, we type

```
> df := diff(f(x),x);
```

Maple gives the derivative

$$df := -3\,\frac{1}{(x+3)^4}.$$

We defined our expression for the derivative as **df** in our Maple command above to allow easy evaluation. To find the value of the derivative at $x = -2$, we type

```
> subs(x=-2, df);
```

and Maple gives us the answer -3.

A standard application of the derivative is finding a maximum or minimum of a function. The derivative is zero for smooth functions at any maxima or minima.

Example 2: Consider the function

$$f(x) = 0.3\,x^3 - 9.1\,x^2 + 3.7\,x + 42.2.$$

Find the minimum and maximum of this cubic polynomial.

Solution: The series of commands are to enter the function, differentiate the function, find when the derivative is zero, then use these values back in the original function. Often it is useful to make comments on a Maple worksheet for future reference. The symbol # is used in Maple to make comments. Anything following the # is ignored by Maple. Note also that you can produce a new line without Maple evaluating it by typing **Shift + Enter**. Below are a series of commands that show how to find the minimum and maximum of a cubic polynomial in Maple along with the Maple responses. Recall that the x-intercepts can be found using **fsolve(f(x) = 0, x);**

```
> f := x -> 0.3*x^3-9.1*x^2+3.7*x+42.2;
# This enters the function.
```

$$f := x \to 0.3\,x^3 - 9.1\,x^2 + 3.7\,x + 42.2$$

```
> df := diff(f(x),x);
#This differentiates the function and assigns it to df.
```

$$df := 0.9\,x^2 - 18.2\,x + 3.7$$

```
> xm := fsolve(df=0,x);
#This finds the x values at the extrema and assigns them to xm.
```

$$xm := .2053826275, 20.01683959$$

```
> f(xm[1]);f(xm[2]);
#This finds the y values at the previous x values found above.
```

$$42.57865834 \qquad -1123.802528$$

From Maple, it easily follows that this cubic polynomial has a relative minimum at $(20.017, -1123.8)$ and a relative maximum at $(0.2054, 42.58)$.

When the function is not a polynomial, it may be necessary to look at the graph of the function and estimate where a maximum, minimum, or point of inflection occurs to find the specific point. Below we work another example that includes an exponential function and proceeds to the second derivative to find points of inflection.

Example 3: Consider the function

$$f(x) = (x^2 - 5)\,e^{-0.1x}.$$

Find all relative extrema and points of inflection. Graph the function showing these critical points and give any asymptotes.

Solution: This problem is solved exclusively in Maple, and the Maple commands are given below. Comments and the Maple responses are shown to help understand the solution.

```
> f := x -> (x^2-5)*exp(-0.1*x);
# This enters the function.
```

$$f := x \to \left(x^2 - 5\right) e^{-0.1\,x}$$

```
> plot(f(x),x=-10..100,y=-10..100);
#This plots the f(x) over an appropriate domain and range.
```

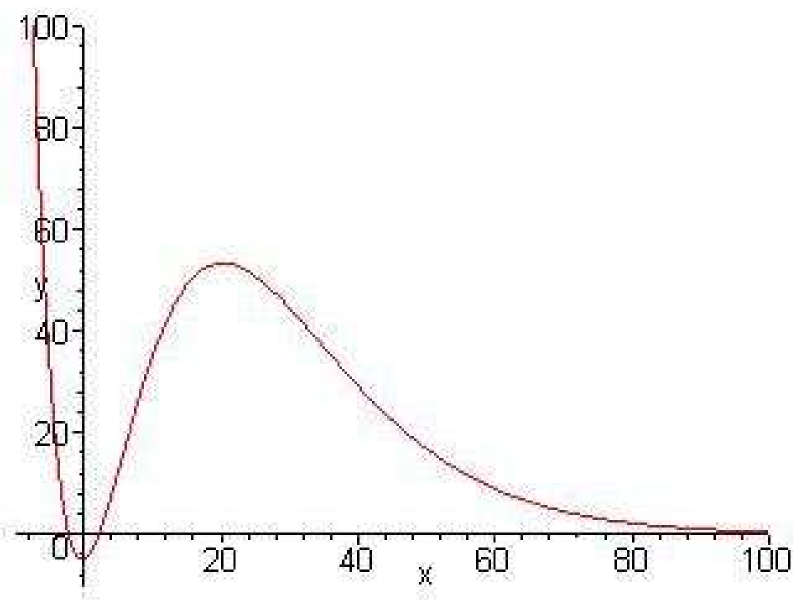


Figure 6.2: Graph of the function $f(x)$.

```
> df := diff(f(x),x);
#This differentiates the function and assigns it to df.
```

$$df := 2\,xe^{-0.1\,x} - 0.1\,\left(x^2 - 5\right) e^{-0.1\,x}$$

Note that sometimes the expression that Maple presents after differentiation is very complicated. The expression can occasionally be simplified by using the Maple commands **simplify(%);** or **factor(%);**. Also, we could graph the derivative by typing

```
> plot(df, x=-10..100,y=-10..100);
```

Note that **df** represents the expression for the derivative, which is slightly different that our functional representation of $f(x)$ in the previous plot command.

The next step is to find the extrema of $f(x)$, which is where the derivative is zero.

```
> xmin := fsolve(df=0,x=-10..10); f(xmin);
#This finds the x and y values of the minimum.
```

$$xmin := -.2469507660 \qquad -5.062503178$$

This minimum was found by looking at the graph of $f(x)$ and limiting where Maple searches for the x value of the minimum. The x value of the minimum is assigned to **xmin**, which is substituted into the function to obtain the y value. It follows that a relative (and absolute) minimum for $f(x)$ occurs at $(-0.24695, -5.0625)$.

```
> xmax := fsolve(df=0,x=10..50);f(xmax);
#This finds the x and y values of the maximum.
```

$$xmax := 20.24695077 \qquad 53.46575737$$

As with the minimum, we restricted our search of x for the relative maximum. Maple readily finds this relative maximum at $(20.247, 53.466)$.

Next we use Maple to find the second derivative, then use this to find the two points of inflection, which are visible on the curve where it changes concavity.

```
> sdf :=  diff(df,x);
#This computes the second derivative of f and assigns it to sdf.
```

$$sdf := 2\,e^{-0.1\,x} - 0.4\,xe^{-0.1\,x} + 0.01\,\left(x^2 - 5\right)e^{-0.1\,x}$$

```
> xp1 := fsolve(sdf=0,x=0..20);f(xp1);
#This finds the x and y values of the first point of inflection.
```

$$xp1 := 5.682178937 \qquad 15.45910694$$

```
> xp2 := fsolve(sdf=0,x=20..60);f(xp2);
#This finds the x and y values of the first point of inflection.
```

$$xp2 := 34.31782106 \qquad 37.91295621$$

It follows that the points of inflection occur at $(5.6822, 15.459)$ and $(34.318, 37.913)$. Finally, we use the limit command to find the horizontal asymptote to the right.

```
> limit(f(x),x=infinity);
```

This gives answer zero indicating a horizontal asymptote to the right of $y = 0$.

6.2 Qualitative Analysis of Discrete Models

In the previous chapter, we examined a number of discrete dynamical models. Excel provided a valuable tool for simulating these models. We learned about the **Solver** function in Excel that allowed us to fit parameters to the model. In the previous section, the derivative was applied to aid graphing by finding minima, maxima, and points of inflection. In this section we apply many of the computer techniques from before, including Excel's **Trendline** and **Solver**, and apply the derivative with the help of Maple to study the qualitative behavior of discrete dynamical systems.

Example 4: We want to return to Example 2 from the chapter on Discrete Dynamical Models. Often a biological study does not have the growth data, but simply population measurements at various times. Below we repeat the data on the bacterium *Staphylococcus aureus* with the table written in a slightly different form.

The general form for a discrete population model is given by

$$P_{n+1} = F(P_n),$$

where $F(P_n)$ is called the updating function for this population model and takes various forms. In this example, we examine two updating functions,

t (hours)	n	P_n OD_{650}
0	0	0.035
0.5	1	0.039
1	2	0.069
1.5	3	0.11
2	4	0.17
2.5	5	0.229
3	6	0.261
3.5	7	0.288
4	8	0.309
4.5	9	0.327
5	10	0.347

the logistic growth model and a Ricker's growth model, and the population is measured by the optical density at a wavelength of 650 nm (OD_{650}).

a. The discrete logistic growth model for the adult population P_n can be written

$$P_{n+1} = F(P_n) = rP_n - mP_n^2,$$

where the constants r and m are determined from the data. Plot the data P_{n+1} vs. P_n and find the best fit of the logistic growth model given above using Excel's Trendline.

b. Find the equilibria for this model. Write the derivative of the updating function. Discuss the behavior of the model near its equilibria.

c. Another common population model is Ricker's, which is given by

$$P_{n+1} = R(P_n) = aP_n e^{-bP_n},$$

where a and b are constants to be determined. This model has certain advantages over the discrete logistic growth model, especially for larger populations. (Ricker's model is always positive, while the logistic growth model can yield negative populations.) Use Excel's solver to find the least squares best fit of the Ricker's updating function to the given data by varying a and b. Once again plot P_{n+1} vs. P_n, using this updating function and show how it compares to the data and the quadratic logistic updating function.

d. Find the equilibria for Ricker's model. Write the derivative of the updating function, then discuss the behavior of these equilibria using this derivative. Simulate the discrete dynamical system using the discrete logistic growth model and Ricker's model comparing them to the actual data.

Discuss the similarities and differences that you observe between models and how well they work for this experimental situation. Use the equilibrium and stability analysis to predict what will happen to the population of *S. aureus* for large times (assuming experimental conditions continue)?

Solution: a. This problem differs from Example 2 of the previous chapter in that we are finding the updating function rather than a growth function for the discrete logistic growth model. The Excel spreadsheet begins by copying the table above into **Columns A-C**. In **Column D**, we insert the values for P_{n+1}, which are simply the values in **Column C** shifted up one cell, thus the entry in **C3** is placed in **D2**. Next the **XY (Scatter)** option of **Chart Wizard** is used to graph the data points from **Columns C** and **D**. Assuming the first row is used for labels, then **Cells C2** to **D11** are graphed. (The entry in **C12** has no pairing, so can't be graphed.) After these data points are graphed, Excel's Tendline with a second order polynomial is inserted. We have to check the options **Set intercept = 0** and **Display equation on chart**. In addition, we add the identity map to this graph, which is simply the line $P_{n+1} = P_n$. This becomes important for the stability analysis in Part b. Figure (6.3) shows the result of this graph. The best discrete logistic growth model is found to be

$$P_{n+1} = 1.6238\, P_n - 1.8439\, P_n^2,$$

which as can be seen, agrees with the model found in Example 2 of the Discrete Dynamical Models chapter.

b. The equilibria, P_e, for the general discrete population satisfy the equation $P_e = F(P_e)$. Thus, we have

$$\begin{aligned} P_e &= 1.6238\, P_e - 1.8439\, P_e^2, \\ 0 &= 0.6238\, P_e - 1.8439\, P_e^2 = 0.6238\, P_e(1 - 2.9559\, P_e). \end{aligned}$$

It follows that the equilibria are $P_e = 0$ and 0.3383. One can easily see that these are the points of intersection of the updating function with the identity map on the graph in Figure (6.3). The derivative of $F(P_n)$, which is easily computed by hand or using Maple,

```
> diff(1.6238*P-1.8439*P^2, P);
```

which gives the result

$$F'(P_n) = 1.6238 - 3.6878\, P_n.$$

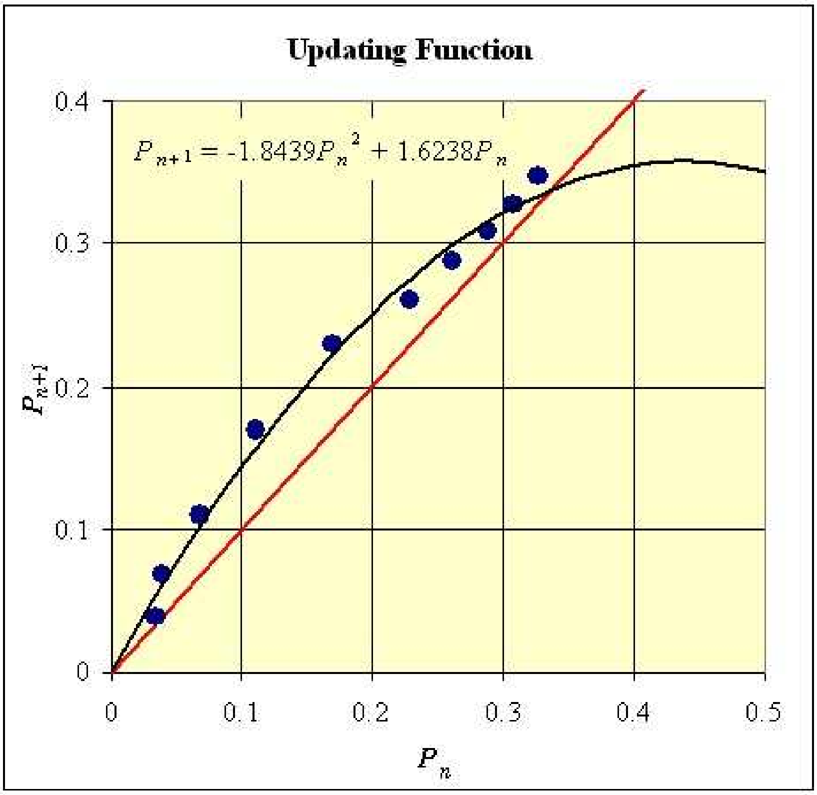
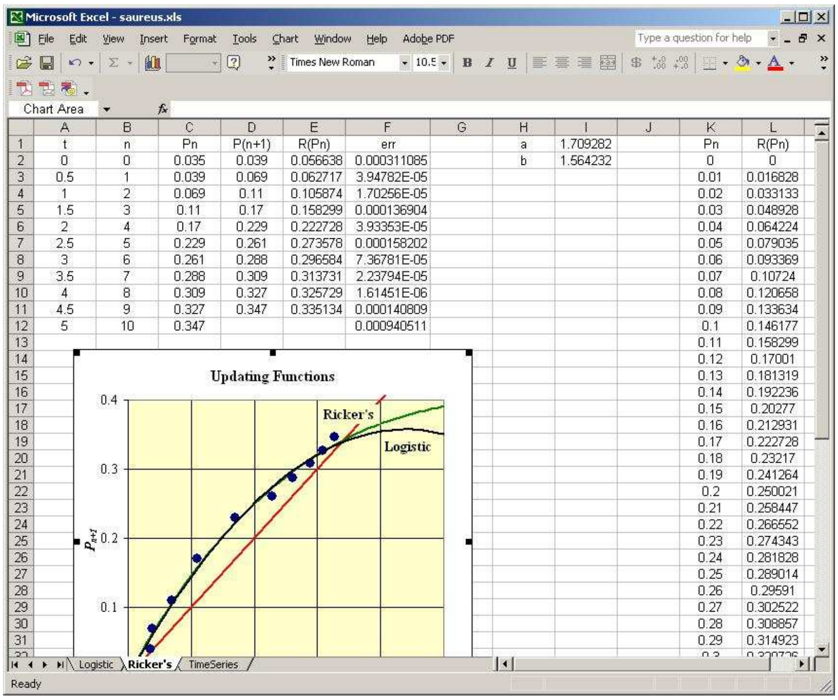


Figure 6.3: Graph of the logistic growth updating function and the data, P_{n+1} vs. P_n.

Thus, $F'(0) = 1.6238$, which is greater than one so the trivial equilibrium (extinction equilibrium) at $P_e = 0$ is unstable with solutions moving away from this point. In a good growth medium, one expects the population to grow, so this equilibrium should be unstable. At the other equilibrium, $P_e = 0.3383$, $F'(0.3383) = 0.3762$, which is less than one. It follows that this equilibrium is stable with solutions monotonically approaching this equilibrium. This equilibrium is the carrying capacity of the culture, and one expects biologically that a culture will grow to stationary phase and level off at this equilibrium value.

c. This part begins with a spreadsheet like in Part a for **Columns A-D**. Figure (6.4) shows the spreadsheet that we developed for this part of the problem. In **Column E**, we insert the Ricker's updating function,

$R(P) = aPe^{-bP}$. It has the parameters a and b, which we name in **Cells H1-I2**. (Recall this is done from the menu item **Insert** selecting **Name**, then **Create** and choosing the appropriate category.) As initial estimates, we entered $a = 1$ and $b = 0.1$. After naming the parameters we enter in **E2** the expression **=a*C2*EXP(-b*C2)**, then fill this down to **E11**. In **Column F**, we enter the square error between **Columns D** and **E**. Thus, in **F2**, we enter = **(D2 - E2)ˆ 2**, and fill this down to **F11**. In **F12** we sum the square error by entering **=SUM(F2:F11)**, which is also done by typing = and clicking on the summation symbol in the second row of the menu bar.



	A	B	C	D	E	F	G	H	I	J	K	L
1	t	n	Pn	P(n+1)	R(Pn)	err		a	1.709282		Pn	R(Pn)
2	0	0	0.035	0.039	0.056638	0.000311085		b	1.564232		0	0
3	0.5	1	0.039	0.069	0.062717	3.94782E-05					0.01	0.016828
4	1	2	0.069	0.11	0.105874	1.70256E-05					0.02	0.033133
5	1.5	3	0.11	0.17	0.158299	0.000136904					0.03	0.048928
6	2	4	0.17	0.229	0.222728	3.93353E-05					0.04	0.064224
7	2.5	5	0.229	0.261	0.273578	0.000158202					0.05	0.079035
8	3	6	0.261	0.288	0.296584	7.36781E-05					0.06	0.093369
9	3.5	7	0.288	0.309	0.313731	2.23794E-05					0.07	0.10724
10	4	8	0.309	0.327	0.325729	1.61451E-06					0.08	0.120658
11	4.5	9	0.327	0.347	0.335134	0.000140809					0.09	0.133634
12	5	10	0.347			0.000940511					0.1	0.146177
13											0.11	0.158299
14											0.12	0.17001
15											0.13	0.181319
16											0.14	0.192236
17											0.15	0.20277
18											0.16	0.212931
19											0.17	0.222728
20											0.18	0.23217
21											0.19	0.241264
22											0.2	0.250021
23											0.21	0.258447
24											0.22	0.266552
25											0.23	0.274343
26											0.24	0.281828
27											0.25	0.289014
28											0.26	0.29591
29											0.27	0.302522
30											0.28	0.308857
31											0.29	0.314923

Figure 6.4: Spreadsheet for the development of the Ricker's updating function.

The next step of the process is to let Excel find the least sum of square errors. This is done by clicking on **F12**, where the sum of square errors is computed. Next we go to the **Tools** on the menu bar and select **Solver**. In the menu box that pops up, we select **Min** and in the box **By Changing Cells** we highlight **Cells I1-I2** (the values of a and b). After pressing the **Solve** button, solver finds the smallest sum of square errors. In this case it gives $a = 1.7093$ and $b = 1.5642$, so the best fitting Ricker's function for these data are

$$R(P_n) = 1.7093\, P_n e^{-1.5642 P_n}.$$

A graph of this updating function with the logistic updating function is shown in Figure (6.5). It is clear that though this function is very different than the quadratic function of the discrete logistic model, these functions are extremely close on the interval between $P_n = 0$ and the equilibrium (the higher value where the updating functions intersect the identity map). The differences in these models are at higher populations, which is away from where there are data.

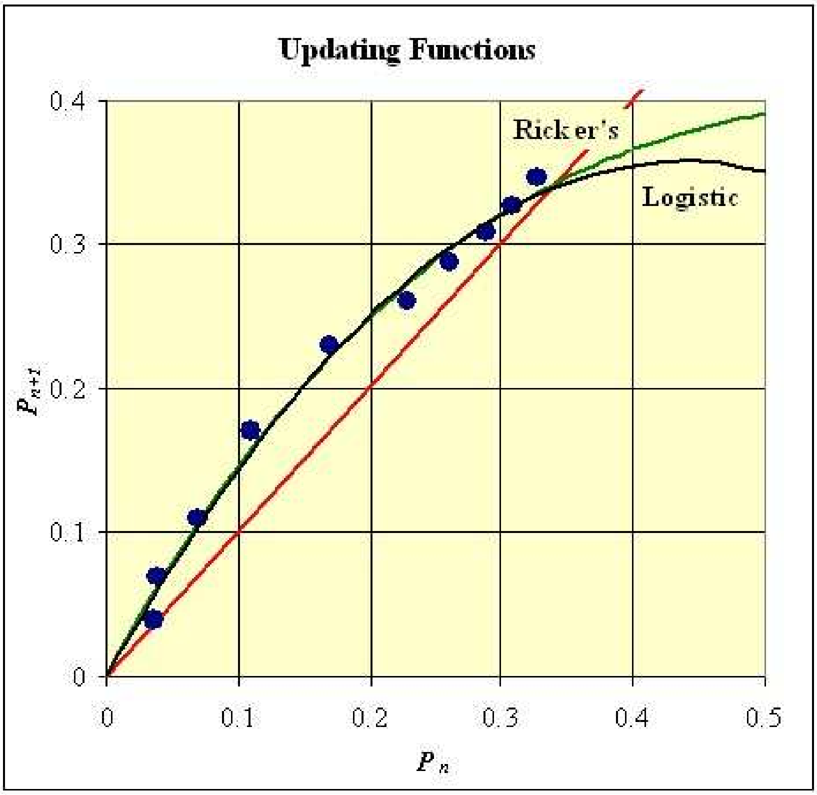


Figure 6.5: Graph of the logistic growth and Ricker's updating functions and the data, P_{n+1} vs. P_n.

d. The next step in this example is the qualitative analysis of the Ricker's model. We find the equilibria by solving

$$P_e = 1.7093\, P_e e^{-1.5642 P_e}.$$

One solution is clearly the trivial solution, $P_e = 0$. (Any closed population model should have this extinction equilibrium.) The other equilibrium satisfies $e^{1.5642P_e} = 1.7093$, which is found by taking logarithms of both sides. Maple can readily find the equilibria using the following commands:

```
> R:=P->1.7093*P*exp(-1.5642*P);
> solve(R(Pe)=Pe,Pe);
```

The other equilibrium is $P_e = 0.34272$, which is slightly higher than the one predicted by the discrete logistic growth model. The derivative of the Ricker's function is found using the product rule or by Maple with the command

```
> dR := diff(1.7093*P*exp(-1.5642*P), P);
# or if R(P) is defined above by dR := diff(R(P), P);
```

The result is

$$R'(P) = (1.7093 - 2.67368706\,P)\,e^{-1.5642\,P}.$$

It follows that the equilibrium at $P_e = 0$ is unstable, since $R'(0) = 1.7093$, which is greater than one. With Maple's help, we evaluate $R'(0.34272)$ at the other equilibrium, $P_e = 0.34272$, by typing

```
> subs(P=0.34272), dR); evalf(%);
```

Thus, $R'(0.34272) = 0.4639$, which is less than one. It follows that this equilibrium is stable with solutions monotonically approaching this equilibrium. Thus again, this equilibrium is the carrying capacity of the culture, and solutions approach this population asymptotically.

Figure (6.6) shows the time series graph of the data and both models. Both models simulate the data very well, and it would be hard to decide which model fits the data better. From the stability analysis, these models predict that over a long period of time this culture of *S. aureus* would stay in a stationary phase with an OD_{650} of approximately 0.34.

6.3 Continuous Logistic Growth Model

The discrete logistic growth model is very easy to simulate, but there is no solution to the model (unlike the solution found by Maple in Section 5.3). However, the continuous version of this model, which results from letting

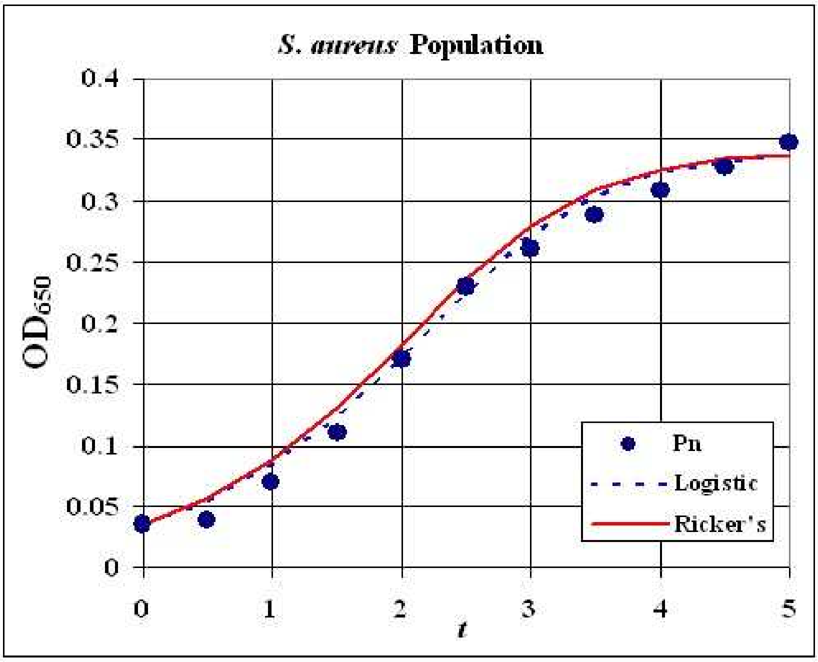


Figure 6.6: Graph of the discrete logistic and Ricker's growth model and the data, P_n vs. t.

the time step between P_{n+1} and P_n go to zero, gives a differential equation that has a solution that we present in the example below.

Example 4: Because of the difficulties noted above for the discrete logistic growth model, most biologists use the continuous logistic growth model for their studies of populations. This model is used very extensively and can be written with the following formula

$$p(t) = \frac{P_0 M}{P_0 + (M - P_0)e^{-rt}},$$

where P_0 is the initial population, M is the carrying capacity of the population, and r is the Malthusian growth rate (early exponential growth rate) of the culture. Consider the population data for the bacterium *Staphylococcus aureus* presented in Table (6.2) in the previous example.

a. Use Excel to find the best values of parameters P_0, M, and r. Include the sum of squares error. Create a graph showing both the data and the logistic growth function, $p(t)$. If the value of r gives the Malthusian growth rate for low populations (P_0e^{rt}), then use this to determine the doubling time for this culture of yeast. With the best fitting parameters, find the actual doubling rate of the logistic model from its initial population.

b. The growth rate for a culture can be found by taking the derivative of the population function. Differentiate the logistic growth function $p(t)$ with the parameters found in Part a. Create a graph of the derivative of the logistic growth function, $p'(t)$.

c. The turning point of the population or the mid-log phase for this culture of bacteria is where the growth of the culture is at a maximum. (This is also the point of inflection for the original logistic growth function, $p(t)$.) Find when the logistic growth function reaches the turning point by finding the maximum of the derivative of the logistic growth function, $p'(t)$. Write the time of the turning point, the maximum growth that you find, and the population of the culture at this time.

Solution: a. This example starts with an Excel spreadsheet, where we insert the data in **Columns A** and **B** with appropriate headings in the first row, such as t in **A1**. (For graphing purposes later, we extend the time units in *Columns A* an additional hour.) We highlight **Column A** and go to the **Insert** on the main menu and select **Name** and **Create**, taking the default setting to have our times labeled t. In **Column E**, we add the labels **P0**, **r**, and **M**, then in **Column F**, we input reasonable guesses. (Note that Excel changes the label **r** to **r_** for unknown reasons.) We selected $P_0 = 0.035$ (the initial value in the data), $r = 1.6$ (chosen from our work in Example 4 on the discrete logistic growth model), and $M = 0.35$ (a value slightly higher than the last value in the data). Once again, we **Name** these parameters from the **Insert** item on the main menu.

In **Column C**, we insert the formula for $p(t)$. Thus, in **C2**, we type **=P0*M/(P0+(M-P0)*EXP(-r_*t))** and fill this formula down to **C14**. (We filled down two cells beyond the data for graphing purposes later.) In **Column D**, we compute the sum of square errors between the model and the data, so in **D2** we insert the formula = **(B2 - C2)^2**, then fill this down to **D12**. In **D13**, we type **=SUM(D2:D12)** or simply type = and select the symbol Σ on the second line of the main menu and the default gives us the desired sum.

The next step in the process is computing the least sum of square errors, which again uses Excel's **Solver**. We highlight **D13**, then invoke **Solver** from the **Tools** listing on the main menu. As before, we check **Min**, then click in the window below **By Changing Cells** and highlight the parameter values in **F2:F4** (the values of P_0, r, and M). Finally, we click on the **Solve** button, and we find that **D13** changes to **0.0005245**, which is the least sum of square errors for this model with the given data. The parameters change from their initial values to $P_0 = 0.025414$, $r = 1.22818$, and $M = 0.34568$. In Figure (6.7) we see the graph of the data and the best fitting model.

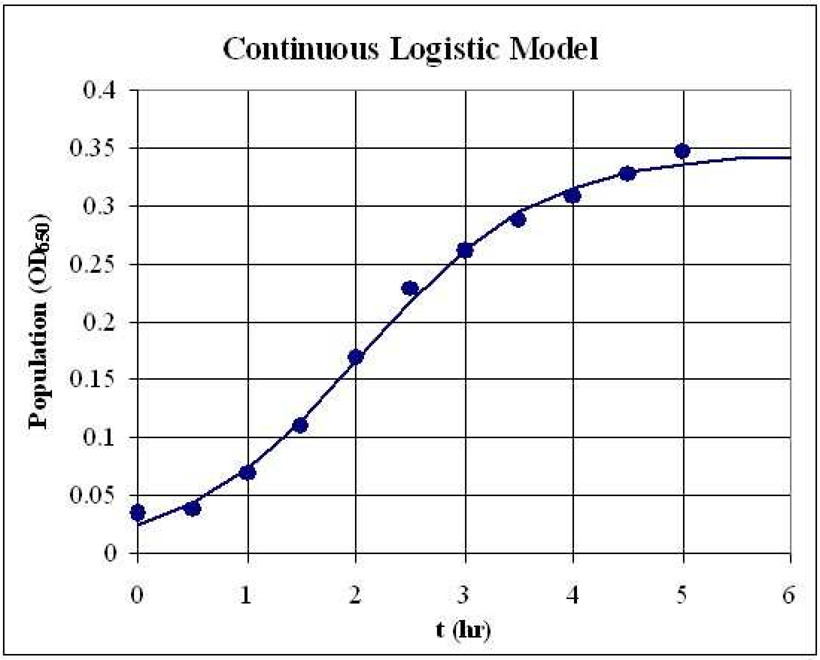


Figure 6.7: Graph of the continuous logistic growth model and the data, $p(t)$ vs. t.

To find the doubling times, we can solve the algebraic expressions. For the Malthusian growth model with the best fitting parameters, we need to solve

$$\begin{aligned} m(t_d) &= P_0 e^{rt_d} = 2\,P_0, \\ e^{rt_d} &= 2, \\ t_d &= \frac{\ln(2)}{r} = 0.56437. \end{aligned}$$

Though this is not difficult, it becomes much more challenging to solve the expression for the continuous logistic growth model doubling, $p(t_d) = 2\,P_0$. Below we show how the doubling times are easily found using Maple.

```
> m := t -> P0*exp(r*t);p(t);
# This enters the Malthusian growth model.
> p := t -> P0*M/(P0+(M-P0)*exp(-r*t));
# This enters the continuous logistic growth model.
> P0 := 0.025414335; r := 1.228178578; M := 0.345683182;
# This assigns the best fitting parameter values.
> fsolve(m(t)=2*P0,t);
# This finds the doubling time for m(t) (Malthusian model)
> fsolve(p(t)=2*P0,t);
# This finds the doubling time for p(t) (logistic model)
```

The results of Maple are the same as above for the Malthusian growth model with the doubling time being $t_d = 0.56437$ hr, while Maple computes the doubling time for the logistic growth model to be $t_d = 0.63169$ hr. The latter time is longer because of the declining growth rate with increasing population.

b. The logistic growth function, $p(t)$, is given by

$$p(t) = \frac{0.008785}{0.02541 + 0.3203\, e^{-1.2282\, t}}.$$

Its derivative is readily found using the Maple command

```
> diff(p(t), t);
```

giving the result

$$p'(t) = \frac{0.003456\, e^{-1.2282\, t}}{(0.02541 + 0.3203\, e^{-1.2282\, t})^2}.$$

Figure (6.8) shows a graph of the logistic growth function with the best parameters found in Part a and its derivative.

c. Figure (6.8) shows that the derivative of the logistic growth model has a maximum, which can be seen matches the point of inflection for the culture of *S. aureus*. The maximum of the derivative is found by computing the zero of the second derivative. The Maple commands that calculate this zero are shown below. Assume that the function $p(t)$ is entered as shown in Part a.

```
> dp := diff(p(t), t); # Computes the derivative of p(t).
> sdp := diff(dp, t); # Computes the second derivative of p(t).
> tp := fsolve(sdp=0,t); # Finds the zero of p"(t).
```

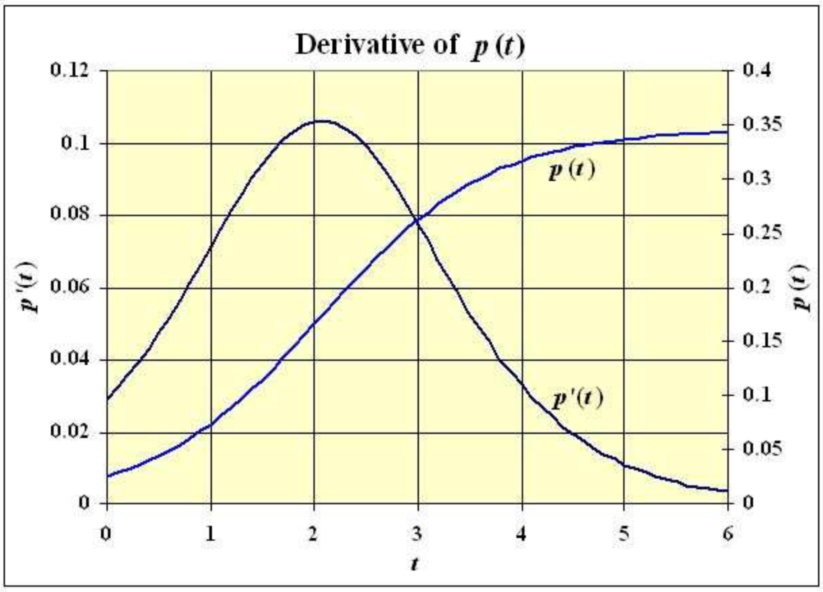


Figure 6.8: Graph of the continuous logistic growth model and its derivative.

This last command assigns to the variable **tp** the zero of the second derivative of $p(t)$, $p''(t)$, which Maple finds to be **tp := 2.06309**. It follows that the turning point or mid-log phase for this culture of *S. aureus* is $t_p = 2.06309$ hr. This value is substituted into $p(t)$ and $p'(t)$ using Maple as follows:

```
> p(tp);
> subs(t=tp, dp); evalf(%);
# The second command gives the decimal answer for the growth rate.
```

The turning point of the population occurs with a population of $p(2.06309) = 0.17284$ OD_{650}, and the maximum growth rate is $p'(2.06309) = 0.10614$ OD_{650}/hr.

6.4 Computer Laboratory Exercises

To simplify listing of web addresses below, we note that the beginning of the web address begins with the following:

```
www-rohan.sdsu.edu/~jmahaffy/courses
```

which will be given as **...** in the text below.

1. (I1) A ball is thrown vertically and data are collected at various times in its flight. Assume that air resistance can be ignored, then the height of the

Time (sec)	0.5	1.0	1.5	2.0	2.5	3.0
Height (m)	7.3	12.1	14.6	14.4	11.8	7.0

ball satisfies the quadratic equation:

$$h(t) = v_0 t - \frac{gt^2}{2},$$

due to gravity. (Note: There is no constant term as we are assuming that the height of the ball is zero at $t = 0$.)

a. Use the Excel's Trendline to find the best constants v_0 and g that fit the data in the table and write that equation in your report. (Remember that when you are using Trendline, you must decide if your graph passes through the origin. Does this one?) Graph both the quadratic function and the data. Find the time that your model predicts the ball will hit the ground. Also, find how high the ball goes, and find the time that it reaches this highest point.

b. The average velocity between two times t_1 and t_2 is given by the formula:

$$v_{ave} = \frac{h(t_2) - h(t_1)}{t_2 - t_1}.$$

Use Excel's spreadsheet capabilities to create tables showing the average velocity of the ball based on the t_1 and t_2 values given below.

t_1	1	1	1	1	1	1
t_2	2	1.5	1.2	1.1	1.05	1.01

t_1	2	2	2	2	2	2
t_2	3	2.5	2.2	2.1	2.05	2.01

The velocity of the ball at a given time is the derivative of the height function at that time. Compute the derivative of $h(t)$,

$$h'(t) = v(t).$$

t_1	2.5	2.9	2.95	2.99	2.999
t_2	3	3	3	3	3

Evaluate $v(1)$, $v(2)$, and $v(3)$. How do these values compare to the values of v_{ave} that you obtained for each of the tables above?

c. Associate v_{ave} with t_1, *i.e.*, let $v_{ave} = v(t_1)$. Once again compute v_{ave} for the table below, then graph $v(t)$ versus t using these data. Describe the graph that you have produced. Use Trendline (or any other method) to find the equation of this graph and find the v and t-intercepts. Compare this equation to the equation of the derivative $h'(t)$ that you obtained above.

t_1	0	0.5	1.0	1.5	2.0	2.5	3.0
t_2	0.01	0.51	1.01	1.51	2.01	2.51	3.01

2. (I2) Pediatricians monitor for normal growth of children by the annual measurement of height and weight. These are expected to increase annually, the growth curve paralleling a standardized curve. In the note introducing the idea of a derivative, there are data on juvenile heights from birth to age 18. Below is a table of both heights and weights for American girls in the 50th percentile.

a. Use the data from ages 4 through 18, together with the Trendline feature of Excel, to find a power law relationship between height and weight. Give a physiological explanation for this relationship.

b. Include the data back to birth (age 0). What happens to the power law? Why does this happen? (Think about the morphological changes between an infant and a small child.)

c. Create a graph of weight versus age, then create another graph of rate of change in weight versus age (much like the graphs seen in the text book [5]). Recall that the rate of growth (in height) was relatively constant over the ages 3 to 12. What happens with the rate of change in weight? Describe the graph for the rate of weight gain over the early years (0-3), the ages 3-12, then adolescence (13-18).

3. (J1) This problem investigates the concept of a derivative from the geometric perspective of limiting secant lines going to a tangent line. It also allows you to explore the rules of differentiation using Maple. (For the the-

age(years)	height(cm)	weight(kg)
0	50	3.4
0.25	60	5.4
0.5	66	7.3
0.75	71	8.6
1	75	9.5
1.5	81	10.8
2	87	11.8
3	94	15.0
4	102	15.9
5	108	18.2
6	114	20.0
7	121	21.8
8	126	25.0
9	132	29.1
10	138	32.7
11	144	37.3
12	151	41.4
13	156	46.8
14	160	50.0
15	161	52.3
16	163	56.4
17	164	57.7
18	164	58.6

oretical description please refer to Chapter 12 of the text book.)

a. Consider the function

$$f(x) = \frac{2}{(3-x)^2}.$$

We want to investigate the derivative of $f(x)$ at $x = 2$ by observing secant lines that pass through the points $(2, f(2))$ and $(2+h, f(2+h))$ for different values of h. The slope of the secant line is given by

$$m(h) = \frac{f(2+h) - f(2)}{h},$$

and this line always passes through the point $(2, f(2))$. Find the equations of the 4 secant lines (in slope-intercept form) using $h = 0.5$, 0.2, 0.1, and

0.05. Use Excel to graph $f(x)$ and the 4 secant lines for $x \in [1, 2.95]$. Limit the range, so that y is in the interval $[-10, 20]$.

b. Find the general formula for the slope of the secant line through $x = 2$ and $x = 2 + h$. The slope of the tangent line is found by taking the slope of the secant line and letting h tend toward zero. This becomes the derivative of $f(x)$ at $x = 2$. Find the derivative of $f(x)$, $f'(x)$, at $x = 2$. Find the equation of the tangent line at $x = 2$, then use Maple graph $f(x)$ and the tangent line for x in the interval $[0, 3]$. (Use the same limits on the range as in Part a.)

c. Differentiate the following functions using Maple. You are just about to learn a collection of differentiation rules, so you should be looking for patterns to help you in the future as we learn more about derivatives.

1. $f(x) = x^2 + 3x - 5$,
2. $f(x) = (x^2 + 3x - 5)^4$,
3. $f(x) = e^{-3x}$,
4. $f(x) = \ln(x + 4)$.

Write a brief description of what you observe upon taking the derivative in each of the above cases.

4. (J2) This problem was given to me by Professor Boyd Collier from his research in Ecology at San Diego State University. The study examines the oxygen consumption of the bug *Triatoma phyllosoma* after ingestion of a blood meal. In this experiment the bug matures through its fifth instar stage of development and begins its molt into the adult phase. This is a bug that causes major problems throughout South America. It spreads the deadly disease chagas when it obtains a blood meal at night from its sleeping victims. The poor rural houses may be so infested with thousands of these bugs that some of the occupants become anemic by losing up to 500 ml of blood per month.

a. Below are the data for the time, t (in hours), and the oxygen consumption, y (in ml of O_2/hr).

These data are most reasonably fit by a cubic equation. Follow the techniques you have used before to let Excel find the best fit to the data using the equation:

$$y = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

t (hr)	y ml O_2/hr
1	116.6
2	120.1
3	114.9
4	129.9
5	116.5
6	107.7
7	99.0
8	104.0
9	100.7
10	87.5
11	82.7
12	53.8
13	54.0
14	72.4
15	81.1

Have Trendline show the equation of the best cubic polynomial through these data. Show the plot of the data points and the best cubic polynomial through the data for $t \in [0, 15]$.

b. Differentiate the polynomial that you found in Part a., then determine where the derivative is zero. What is the correspondence of these numbers to the maximum and minimum values of oxygen consumption for t in the given interval? State where the function is increasing and where it is decreasing.

c. Recall that at $t = 0$, the bug has just finished eating a blood meal, and at $t = 15$, it molts to become an adult. With this information briefly present a biological explanation for why you see the regions of increasing oxygen consumption and decreasing oxygen consumption. (Hint: You might want to think about your own physiology after eating a meal. The molting stage could be related to you deciding to exercise at some time after your meal.)

5. (J3) With the death of Sonny Bono, there has been renewed interest in the Salton Sea. The Salton Sea formed from 1905-1907 when an engineering mistake plus heavy rains on the watershed of the then undammed Colorado River combined to break through a levee. The lake was originally freshwater but became saltier than seawater as there is no outlet and a lot of evaporation. The water going into it is fairly salty (leached from the agricultural

soils). The creatures in it are mostly marine, some introduced on purpose and some accidentally with the establishment of a sport fishery. If it weren't for the agricultural and municipal wastewater flowing into the Sea it would have dried up long ago. However this water also has a lot of fertilizers, which cause massive algal blooms and a large biomass of invertebrates and fish. People like to fish there because it is so easy to catch fish, but there are also large fish kills, which are NOT pleasing. There are many birds at the Sea (there is lots of food for them), but they also experience large die-offs at times.

There are not very many kinds of metazoan zooplankton in the Salton Sea, but often there is a high density present. Professor Debbie Dexter is currently studying marine invertebrates in the Salton Sea, and a presentation of some of her work can be found on the second floor of the Life Sciences building. Mary Ann Tiffany provided me with the background information above and the data below on some of the zooplankton.

The table below lists the number/liter of various zooplankton for data averaged over depth for station S-1 (in the center of the north basin of the Salton Sea at a depth of 14 meters). The first column lists the number of days after January 1, 1997 when the data were taken. The second column represents the rotifer, *Brachionus rotundiformis*, the third column represents the nauplius form of the barnacle (*Balanus amphitrite*) larvae, and the last column is the nauplius form of the copepod, *Apocyclops dengizicus*.

a. Graph each of these species in Excel using a logarithmic scale for the population. (Create three separate graphs.) Determine the season of the year when each of these species is most productive.

b. Take the natural logarithm of the data for Rotifers and Barnacles, then plot these against time. Use Excel's Trendline to find the best 4th order polynomial through the logarithm of the data versus time. When Excel gives you the equation of this 4th order polynomial, click on the formula and transform the coefficients using scientific notation with 2 decimal places.

c. Differentiate the polynomials found in Part b. and determine where the derivatives are zero. These should correspond to the three minima or maxima. Find when these minima and maxima occur and state the populations at these extrema. (Recall you have the logarithm of the populations, so you will need to exponentiate your answer.) Give the approximate dates of the minimum and the maximum that occur in the range of the data based on the polynomial fit that you found. (One should lie outside the range of the data.) Note that a maximum would be a good time for birds to feed at the Salton Sea.

Date	Rotifers	Barnacles	Copepods
21	0.045	4.466	0.060
34	0.047	2.04	0.063
53	0.073	0.7	0.102
78	0.167	0.573	0.251
106	51.785	0.295	0.093
154	182.403	0.035	45.687
175	372.655	0.031	50.25
199	295.288	0.035	14.56
225	802.128	0.039	59.539
249	532.203	0.031	21.629
277	33.723	0.031	2.992
311	9.245	0.056	4.551
329	0.93	0.149	1.65
371	0.047	0.491	0.144
402	0.081	1.618	0.159
423	0.178	1.925	0.097
454	5.826	0.408	0.08
479	299.183	1.923	0.08

6. (K1) Consider the function

$$g(x) = (4 - x^2)e^{-x^2}.$$

a. Graph $g(x)$ for $x \in [-5, 5]$. Also, graph this function for $x \in [1.9, 3]$ to observe behavior near one of the x-intercepts. Find all x- and y-intercepts (to at least 4 significant figures).

b. Find the derivative of $g(x)$ and write it in your lab report. Graph $g'(x)$ for $x \in [-5, 5]$ and $x \in [1.9, 3]$. Find the points (x and y values) where the maximum and minimum values of g occur. (Note that there are 3 extrema.)

c. Find the second derivative of $g(x)$. Determine when it is zero, which is where points of inflection occur. Find all points of inflection (x and y values) for this function. (There are 4 of them.)

d. Is this function odd, even, or neither?

7. (K2)In this problem use the power rule to determine the pulse (beats per min) as a function of the weight (kg) of the animal. You are given the

following data concerning six animals [1]:

Animal	Weight (kg)	Pulse
Mouse	0.017	450
Hamster	0.103	347
Guinea Pig	0.437	269
Goat	33	81
Man	68	65
Cattle	500	49

a. Let P be the pulse and w be the weight, then the power law expression relating the pulse to the weight is given by

$$P = kw^a.$$

Use the power law under Excel's Trendline to best fit the data above. Plot the data and the best power law fit and have Excel write the formula on your graph. (Note that you will be adding the model from Part c. to this graph, so wait to enter only one graph in your lab report.)

b. Determine the sum of squares error between the data and the model that you found in Part a. How well does the graph match the data? Find the percent error between the pulse given by the model and the actual data for each of the animals in the table above. (Assume that the weight in the table is accurate.) Which animal has the highest percent error and explain why you might expect this? Also, which animal has the lowest percent error and explain why this might be the case?

c. Use the applet at

`.../s00a/math121/labs/labk/q2v1.htm`

to find the nonlinear least squares best fit to the data. (There is a hyperlink at

`.../s00a/math121/lectures/allometic_modeling/nonlinlstsq.html`

discussing this nonlinear fit.) Minimize the sum of squares error, then write in your lab report the equation of the best model and the value for the sum of squares error. Find the percent error between the pulse given by this model and the actual data for each of the animals in the table above. Simulate this model and add its graph to the graph produced in Part a. (Note that this model has a vertical asymptote at $w = 0$, so recall how to make Excel graphs look good near a vertical asymptote.) Compare this model to the one given by Excel above. Which is the better model? Explain why differences arise.

d. Use both models to find the missing entries in the table below. Discuss which estimates are the best and which are the worst.

Animal	Weight (kg)	Pulse
Rat		352
Opossum		187
Swine	100	
Elephant	2500	

e. Write an expression for the derivative $P'(w)$ for each of the models. Is the function increasing, decreasing, or neither? Write a brief paragraph that describes how pulse and weight are related based on these models. [1]

8. (K3) Several of you are considering careers in medicine and biotechnology. Drug therapy and dose response is very important in the treatment of many diseases, particularly cancer. Since cancer cells are very similar to your normal body cells, their destruction relies on very toxic drugs. There are some very fine lines in certain cancer treatments between an ineffective dose, one that destroys the cancer, and one that is toxic to all cells in the body. At the base of many of the calculations for these treatments are simple mathematical models for drug uptake and elimination.

a. The simplest situation calls for an injection of the drug into the body. In this case, a fixed amount of the drug enters the body, then its quantity decreases exponentially as the drug is metabolized and excreted from the body. Suppose that for a certain patient, it is found that the amount of a certain drug in his body satisfies the equation

$$A(t) = 10e^{-kt},$$

where your patient has $k = 0.03$ (day^{-1}). Determine how long the drug is effective if it has been determined that the patient must have at least 3 mg in his body.

b. With new materials being developed, the drug can be inserted into polymers that slowly decay and release the drug into the body. This delivery system can prevent large toxic doses in the body and maintain the drug level for longer at therapeutic doses. Suppose that the amount of drug delivered by this new type of drug delivery system satisfies the model

$$B(t) = B_0(e^{-pt} - e^{-qt}),$$

where $B_0 = 14.3$ (mg), $q = 0.1$ (day^{-1}), andp $= 0.03$ (day^{-1}). The first decaying exponential is from the body metabolism, while the second one is from the polymer degradation. (It can be shown that this is the same amount of drug as delivered in Part a.) Once again assume that the patient must have 3 mg in his body to be effective. Over what time period (if any) is this therapy effective. Is this time period longer or shorter than your answer from Part a? (Hint: Finding when the level of the drug is at 3 mg uses the Maple command **fsolve(B(t)=3, t=A..B);**, where A and B are selected to include the region where $B(t) = 3$ by observation of the graph.)

c. On a single graph show both solutions, $A(t)$ and $B(t)$, for 60 days. Find what the maximum dose is in the body from the second treatment given in Part b. and when this occurs. Which treatment do you consider to be superior and why?

9. (L1) A. C. Crombie [2] studied *Oryzaephilus surinamensis*, the saw-tooth grain beetle, with an almost constant nutrient supply (maintained 10 g of cracked wheat weekly). These conditions match the assumptions of the discrete logistic model. The data below show the adult population of *Oryzaephilus* from Crombie's study (with some minor modifications to fill in uncollected data and an initial shift of one week).

Week	Adults	Week	Adults
0	4	16	405
2	4	18	471
4	25	20	420
6	63	22	430
8	147	24	420
10	285	26	475
12	345	28	435
14	361	30	480

The discrete logistic growth model for the adult population P_n can be written

$$P_{n+1} = f(P_n) = rP_n - mP_n^2,$$

where the constants r and m must be determined from the data.

a. Plot P_{n+1} vs. P_n, which you can do by entering the adult population data from times 2–30 for P_{n+1} and times 0–28 for P_n. (Be sure that P_n is on the horizontal axis.) To find the appropriate constants use Excel's

Trendline with its polynomial fit of order 2 and with the intercept set to 0 (under options). In your lab, write the equation of the model which fits the data best. Graph both $f(P)$ and the data.

b. Find the equilibria for this model. Write the derivative of the updating function. Discuss the behavior of the model near its equilibria. (Note that if P_e is an equilibrium point, then you can determine the behavior of that equilibrium by evaluating the derivative of the updating function $f(P_n)$ at P_e. (For more reference check Chapter 8 of the text book [5]). Simulate the model and show this simulation compared to the data from the table above (adult population vs. time). Discuss how well your simulation matches the data in the table. What do you predict will happen to the adult saw-tooth grain beetle population for large times (assuming experimental conditions continue)?

c. Another common population model is Ricker's, which is given by

$$P_{n+1} = R(P_n) = aP_n e^{-bP_n},$$

where a and b are constants to be determined. Use Excel's solver to find the least squares best fit of the Ricker's updating function to the given data by varying a and b. As initial guesses take $a = 2.5$ and $b = 0.002$. Once again plot P_{n+1} vs. P_n, using this updating function and show how it compares to the data (much as you did in Part a.

d. Find the equilibria for Ricker's model. Write the derivative of the updating function, then discuss the behavior of these equilibria using this derivative. (Give the value of the derivative at the equilibria.) Simulate the discrete dynamical system using Ricker's model. Show the graphs of the logistic and Ricker's models with the data. Compare these simulations with the data. Discuss the similarities and differences that you observe between models and how well they work for this experimental situation.

10. (L2) We have studied the discrete logistic growth model and seen the difficulties computing and analyzing populations using this model. As mentioned before, most biologists use the continuous version of the logistic growth model for their studies of populations. This model is used very extensively and can be written with the following formula

$$p(t) = \frac{P_0 M}{P_0 + (M - P_0)e^{-rt}},$$

where P_0 is the initial population, M is the carrying capacity of the population, and r is the Malthusian growth rate (early exponential growth rate) of the culture. Below is a table with data from Gause [3] on a growing culture of the yeast *Schizosaccharomyces kephir* (a contaminant culture of brewers yeast).

Time (hr)	0	14	33	57	102	126
Volume	1.27	1.7	2.73	4.87	5.8	5.83

a. Use Excel to find the best values of parameters P_0, M, and r, then write these values clearly in your lab report. Include the sum of squares error. Also, write the complete formula with the best parameter fit in your report. If the value of r gives the Malthusian growth rate for low populations (P_0e^{rt}), then use this to determine the doubling time for this culture of yeast. As noted above, M is the carrying capacity of the population. Give a brief biological interpretation of this parameter and describe what your value of M says about what happens to this experimental culture of yeast. Create a graph showing both the data and the logistic growth function, $p(t)$.

b. The growth rate for a culture can be found by taking the derivative of the population function. Differentiate the logistic growth function $p(t)$ with the parameters found in Part a. and write this formula in your lab report. Create a graph of the derivative of the logistic growth function, $p'(t)$.

c. The turning point of the population or the mid-log phase for this culture of yeast is where the growth of the culture is at a maximum. (This is also the point of inflection for the original logistic growth function, $p(t)$.) Find when the logistic growth function reaches the turning point by finding the maximum of the derivative of the logistic growth function, $p'(t)$. Write the time of the turning point, the maximum growth that you find, and the population (volume) of the culture at this time.

11. (L1) The growth of fish has been shown to satisfy a model given by the von Bertalanffy equation:

$$L(t) = L_0(1 - e^{-bt}),$$

where L_0 and b are constants that fit the data in previous chapters, it was shown that there is often an allometric model relating the weight and length of different animals. A model relating the weight of a fish as a function of its length

$$W(L) = kL^a,$$

where k and a are constants that fit the data.

Age (yr)	1	2	3	4	5	6	7	8	9	10
Length (cm)	83	155	208	249	279	302	319	332	342	349

a. Below are growth data for the Blue Marlin (*Makaira mazara*) [4].

Find the least squares best fit of the data to the von Bertalanffy equation above. Give the values of the constants L_0 and b (to at least 3 significant figures) and write the model with these constants. Find the intercepts and any asymptotes for the length of the Blue Marlin. Graph the data and the model.

b. Below are data on the length and weight for the Blue Marlin [6].

Length (cm)	110	135	160	165	175	190	210	225	255	300
Weight (kg)	13	41	50	53	62	95	119	173	248	426

Use Excel's Trendline (Power Law) to find an allometric model of the form above. Give the value of the constants k and a (to at least 3 significant figures) and write the model with these constants. Graph the data and the model.

c. Create a composite function to give the weight of the Blue Marlin as a function of its age, $W(t)$. Find the intercepts and any asymptotes for $W(t)$. Graph the weight of a Blue Marlin as it ages.

d. Find the derivative of $W(t)$ using the chain rule (Refer to Chapter 17 of the text book [5] to learn more about this rule). Also, compute the second derivative, then determine when this second derivative is zero. From this information, find at what age the Blue Marlin are increasing their weight the most and determine what that weight gain is. Graph $W'(t)$.

6.5 References

[1] P. L. Altman and D. M. Dittmer, (eds.) (1964) Biology Data Book , Federation of American Societies for Experimental Biology.

[2] A. C. Crombie (1946) On competition between different species of graminivorous insects, Proc. R. Soc. (B), **132**, 362-395. [1]

[3] G. F. Gause, (1934) Struggle for Existence, Hafner, New York.

[4] M. G. Hinton. Status of Blue Marlin in the Pacific Ocean. Website accessed 1/04.

[5] J.M. Mahaffy and A. Chavez-Ross, *Calculus: A Modeling Approach for the Life Sciences*, San Diego State University, Pearson Custom Publishing, U.S.A. (2004).

[6] J. H. Uchiyama and T. K. Kazama. Updated Weight-on-length relationships for pelagic fishes caught in the central north Pacific Ocean and bottomfishes from the Northwestern Hawaiian Islands, www.nmfs.hawaii.edu/adminrpts/PIFSC_Admin_Rep_03-01.pdf, (accessed 1/04)

Index