Lab 3: From Data to Models

One of the goals of mathematics is to explain phenomena represented by data. In the business world, there is an increasing dependence on models. We may want to represent sales vs. advertising expenditures, revenue vs. time, etc. so that a function can be used to analyze results. Businesses also may be interested in forecasting future results, another use of mathematical models. In many cases, analysis of a trend is not a linear analysis, but requires insight into the general problem and the selection of a more realistic nonlinear function. In this lab, we use the Modeler to find suitable functions and to initiate or reinforce your thinking along these lines.

There are some broad questions that you must consider in order to select a function to use in your model. Give these questions some thought. (Do not write answers.)

- Does the data indicate an increasing trend, decreasing trend, or both? Do you need a function that grows as your independent variable increases? Decreases?
- What is the expected long-term behavior? Do you need a function that levels off eventually, or one that increases (decreases) without bound?
- When you choose a function, how well does the function fit the data you have?
- Does a small error in your data destabilize your model? How sensitive or how stable is your model relative to small errors in the data? (This question will be examined in the lab.)

1. **Getting acquainted with the tool**

   Select the Function kit and open the Modeler tool. First, you see a graphing window, a button for selecting the data set by title and five "typical" graphs for functions at the bottom of the screen. The independent variable in many instances is time, \( t \). We have expressed the general form as a function of time here. The graphs represent the function types:

   - **linear functions** \( f(t) = mt + b \)
   - **quadratic functions** \( f(t) = at^2 + bt + c \)
   - **cubic functions** \( f(t) = at^3 + bt^2 + ct + d \)
   - **exponential functions** \( f(t) = ae^{bt} \)
   - **logistic functions** \( f(t) = \frac{kN_0}{N_0+(k-N_0)e^{-rt}} \)

1.1 Which of the functions in the list do not have any asymptotes?
1.2 Which of the functions have horizontal asymptotes?

1.3 Which of the functions have vertical asymptotes?

2. Microsoft Revenue

As an example, let us find a function to model Microsoft Revenue from the data set given. Select the data set and examine the graph of the data. We can see that it certainly is not linear.

As we study the data, an essential question comes to mind: Can Microsoft Revenue grow without bound? Can it grow without leveling off eventually or is there a limiting value? That is, is there a horizontal asymptote hiding out there in the future?

2.1 Make a decision about the question above and give a convincing argument. Pretend that you are talking to a client or a boss! You can always hedge your advice by saying that you are talking about the short term.

2.2 Let us see what the data indicates. If there is no limiting value, then we can try to fit an increasing piece of a quadratic, cubic or exponential function to the data. If there is a limiting value, the logistic function may give us the best fit.
Fitting a quadratic function:

- Select the \textbf{quadratic} function by clicking on "quadratic" (The graph is a parabola.)

- The sliders for \(a\), \(b\) and \(c\) appear. These are the coefficients in the function \(f(t) = at^2 + bt + c\) where \(t\) represents time. First use the slider to select \(c\), since \(c = f(0)\), the \(y\)-intercept. Then use the sliders for \(a\) and \(b\) to get the best visual fit for the data.

- To get a measure for the error, we take the sum of the square differences between the \(y\)-values for the data point and the corresponding point on the curve: \(\text{square error} = \sum (y - f(t))^2\). (Squaring the difference insures that all the values are positive and there is no canceling effect.)

Cubic and exponential functions are handled similarly. However, the logistic function, whose graph is the shape shown on the button with horizontal asymptotes, may not be as familiar to you. It is used in many applications where there is some limiting value, for example the growth of a population in an environment with limited resources, the adoption of an innovation such as the Xerox machine, the saturation of an advertising campaign, etc. On the \textbf{Modeler}, \(N_0\) represents the initial value (e.g., the initial population), \(k\) is the constant of proportionality and \(r\) is the percent growth rate. \((r = b\) in the exponential function.\)

2.2.1 Fit the quadratic, cubic, exponential and logistic functions to the data using the methods described and give the results.

\begin{align*}
\text{quadratic function: } & a = \underline{\phantom{0}} \quad b = \underline{\phantom{0}} \quad c = \underline{\phantom{0}} \quad \text{Error:} \underline{\phantom{0}} \\
\text{cubic function: } & a = \underline{\phantom{0}} \quad b = \underline{\phantom{0}} \quad c = \underline{\phantom{0}} \quad d = \underline{\phantom{0}} \quad \text{Error:} \underline{\phantom{0}} \\
\text{exponential function: } & a = \underline{\phantom{0}} \quad b = \underline{\phantom{0}} \quad \text{Error:} \underline{\phantom{0}} \\
\text{logistic function: } & N_0 = \underline{\phantom{0}} \quad k = \underline{\phantom{0}} \quad r = \underline{\phantom{0}} \quad \text{Error:} \underline{\phantom{0}}
\end{align*}

2.2.2 With a little more information, have you changed your mind? Explain.
2.2.3 Is there a hidden asymptote for this data set? ______

2.2.4 Depending on your answer, which function(s) from the above is the most realistic?

3. Decreasing functions

3.1 In the menu of data, find an example in which the graph of the data indicates a decreasing function. What is the name of the data set?

__________________________________________________________________________

3.2 Fit the data to the curve that seems to describe the data most effectively. If you cannot decide, there is a worksheet attached on which you can keep track of several trials and the errors associated with these trials.

3.3 Explain why you selected the function you used to fit the data. (Discuss asymptotes, long term behavior, etc.)

3.4 After finding the curve of best fit, give the equation of the function with the parameter values that you discovered and give a rough sketch of the graph. (Be sure to label axes and scale.)
4. Your choice!

4.1 Select a data set of interest to you. Name the data set: ____________________

4.2 Select a function to fit the data points. If the fit is unsatisfactory, try another function.

4.3 How did you decide on the function to fit?

4.4 After finding the curve of best fit, give the equation of the function with the parameter values that you discovered and give a rough sketch of the graph. Again, label axes and scale.

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Function: _____________________________   Error:___________

4.5 What are the limits of the use of this model? Can you think of any circumstances that you convince you to change your model?
5. Further Exploration

Using this modeling approach to find a function that best fits the data is not perfect. For one thing, there are other functions which might provide a better fit. In addition, the measure of the error used in this lab involves the absolute value of the differences between data values and values generated by the function. This is not the common method for determining error. Rather, the usual method is called the "least squares regression" method, and is discussed in another lab. There are statistical procedures for discovering the goodness of fit, generally taught in a statistics course.

Generally, we try to find a simple function that approximates the data. If there are 4 data points, we can solve a system of equations and find the function \( f(t) = ax^2 + bx + c \) that fits the four points exactly. However, this is not the goal. The objective is to find a function which represents the information in the data, or the sense of the data -- linearity, exponential tendencies, etc., rather than to fit the given points exactly. This concept is discussed in more detail in statistics and in trend analysis.

5.1 Apple Revenue

One of the data sets, Apple Revenue, has a strange behavior. Things were going along very well for the company, and all of a sudden there was a drop in revenue. What do you think would cause such a drop?

One assumption in making projections is that things will continue as they are. In this way, a model based on past data will continue to hold in the future. One question, of course, is just how far into the future can one project. You may wish to discuss this issue with your professors in finance and economics.

In the case of the data given, you can define a function in two pieces: One section will represent the function from \( t = 0 \) to \( t = 11 \) and a different function for \( t > 11 \) to \( t = 14 \). Just try to fit a pair of linear equations to this data.

\[
f(t) = \begin{cases} \quad & 0 \leq t \leq 11 \\ \quad & 11 < t \leq 14 \end{cases}
\]

5.2 Intercepts and asymptotes

Before continuing to fit curve to data, Find the \( y \) -- intercept and horizontal asymptotes for the graphs of the exponential and logistic functions. This information will assist you in setting the sliders.

5.2.1 To do this, first consider the exponential function: \( f(t) = Ce^t \). We find the \( y \) -- intercept by substituting \( t = 0 \) into the function.

The \( y \) -- intercept is \( f(0) = \)

Use this information to get an initial setting for \( C \) using the sliders.

If \( r \) is positive, the function has no horizontal asymptote for values \( t \geq 0 \). Does it have a horizontal asymptote when \( r \) is negative?
Do you see how this information helps you to choose values of the parameters?

5.2.2 Repeat this exercise of finding the $y$—intercepts and horizontal asymptotes for the logistic function.

Use the information gained about the parameters to choose values of $N_0$ and $K$ in order to obtain a good fit for the graph.

5.3 Cyclic Behavior

There are other types of functions which can be used to model cyclic behavior. For example, in some parts of the country, sales of ski equipment and sales of bathing suits depend on the season. During certain seasons, sales of one of these might be high, and another season, be nonexistent. Can you think of a curve that exhibits cyclic behavior? If you have not studied trigonometric functions, you may wish to skip this section.

Find a data set that exhibits cyclic behavior and see if you can find a model for the data.
Modeling Worksheet

In the table below, you may include up to three functions to model a set of data. Below the function, include the total error for each function.

<table>
<thead>
<tr>
<th>Data Title:</th>
<th>Notes</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Function 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>