MATHEMATICAL MODELING: AN ALGEBRAIC TOOL FOR GUIDED-INQUIRY SCIENCE

Scott A. Sinex, Barbara A. Gage, and John L. McClure
Department of Physical Sciences
Prince George's Community College
Largo, MD 20774-2199

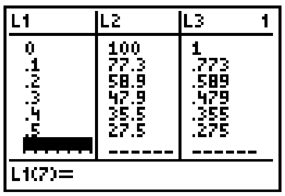
Mathematical modeling is the process of using graphical analysis to determine the symbolic relationships, or equations, which best describe a set of numerical data. Using algebra and the graphing calculator (or spreadsheet), we can analyze experimental data to discover many scientific relationships in chemistry and physics. The data can be collected manually or via a calculator-based laboratory (CBL) or computers with interfaces. The table below gives a sample of the relationships that can be discovered by modeling collected data.

Model	Experimental Relationships	
	without data transformations	with data transformations
Linear	Charles' Law Gay-Lussac's Law Hooke's Law	Lambert's Law Beer's Law
Quadratic	Free fall (accelerated motion)	
Exponential	Radioactive decay Light transmittance	Newton's Law of Cooling
Power	Boyle's Law Inverse Square Law for light Magnetic field strength	

In modeling the task is to identify and understand a system whose behavior can be explained by linear, quadratic, exponential, or power functions. Other types of relationships between variables can exist, such as logistic (populations) and sine (cycles), but are more complicated. Higher order polynomial models (cubic, quartic) also exist but are rare in nature. In this article, we illustrate elementary mathematical modeling using the TI-83 graphing calculator, and show how it naturally leads to the discovery of relationships. Science process and higher order thinking skills are automatically involved in modeling. Mathematical modeling, requiring the use of algebra and simple statistics, is a very useful tool for guided inquiry and a very important part of the "real world" scientific method. Key strokes on the TI-83 are shown with [ ] in bold, such as **[ON]**.

AN EXAMPLE OF MODELING DATA

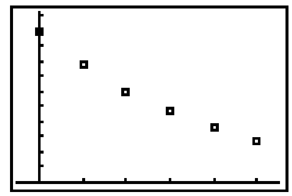
How does light behave when passing through a colored solution? To understand this system, let's examine some data from an experiment which measured light transmitted through a series of colored solutions of increasing concentration. The data are stored in lists in a TI-83. Concentration of the solution is in L_1 and the percentage of light transmitted (%T) is in L_2. The percent is converted to a fraction of light transmitted (T) and is stored in L_3 (L_2/100). Use **[STAT]** EDIT to enter the data. Is there a relationship between light transmitted and concentration?



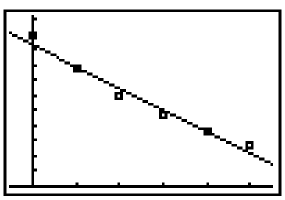
L1	L2	L3 1
0	100	1
.1	77.3	.773
.2	58.9	.589
.3	47.9	.479
.4	35.5	.355
.5	27.5	.275

L1(7)=

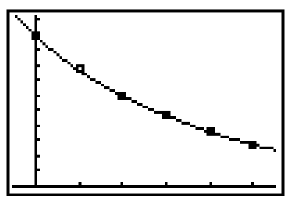
To address this question a scatter plot is generated. Use **[2nd] [STAT PLOT]** to set up the plot of L_3 against L_1, then press **[ZOOM][9]**. The light transmitted is on the y-axis and concentration is on the x-axis. The scatter plot shows a rather consistent decreasing relationship of the fraction of light transmitted as concentration increases.



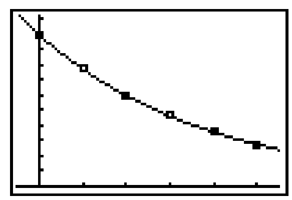
We can analyze the scatter plot by two approaches: regression analysis using the graphing calculator; or manually fitting a function. First, let's consider three different regression analyses of the same data, and compare the coefficients of determination, r^2 (DiagnosticON under **[2nd][CATALOG]**).



LINEAR
$r^2 = 0.9674$



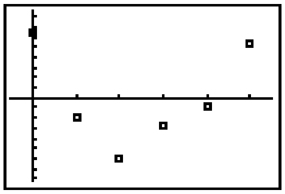
QUARTIC
$r^2 = 0.9992$

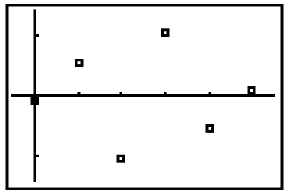


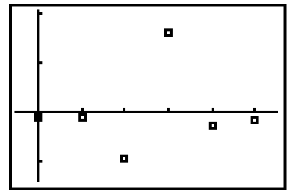
EXPONENTIAL
$r^2 = 0.9987$

How can we determine which model best fits the data? The coefficient of determination, r^2, for the linear model is 0.9674 or about 96.74% of the variation of light transmitted is explained by concentration, while 99.92% is explained for the quartic and 99.87% for the exponential. For the statisticians, an $r^2 > 0.841$ for n = 6 is significant at the 99% level. Which model is the best fit?

A simple graphical way to judge best fit is to plot the residuals, which are the differences from the actual y-value and the y-value obtained from the regression equation when the appropriate x-value is placed in the equation. The TI-83 automatically calculates and stores the residuals after a regression analysis is performed. The residuals, RESID, are stored under **[2nd] [LIST]** NAMES. A plot of the residual vs. the x-variable is done via **[2nd] [STAT PLOT]**, where x-variable is the x list and RESID is the y list. For the plots below, the y-axis is the residual and Yscl = 0.01 on all three.







LINEAR QUARTIC EXPONENTIAL

Best fit is judged when the residuals are at a minimum, and the plot is random and does not show any pattern. A pattern in the residuals plot indicates that the regression does not fit the data, especially near the extremes of the data. The LINEAR model shows a pattern in the distribution of the residuals! The QUARTIC model shows the smallest residuals, hence the best fit as also judged by using the coefficient of determination, r^2, above.

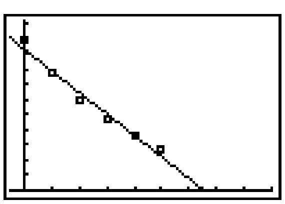
Since the sum of the residuals is always zero, another way to judge the goodness of fit is to minimize the sum of the squared residuals for best fit (SUM is found under **[2nd] [LIST]** MATH). This is sometimes referred to as the sum of the squared errors or SSE. For the three plots above:

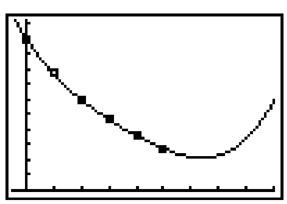
LINEAR	QUARTIC	EXPONENTIAL
SUM (LRESID^2) =	SUM (LRESID^2) =	SUM (LRESID^2) =
1.2×10^{-2}	2.8×10^{-4}	3.7×10^{-4}

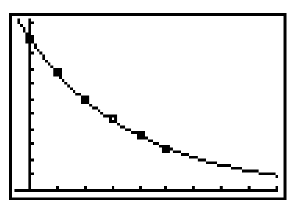
So, the quartic fit, with the smallest SSE value, indicates the best fit. Residuals provide an easy and understandable way to judge goodness of fit and help provide a simple explanation to describe regression analysis.

An important aspect of a model is how well it allows us to predict values. Predictions, especially beyond the range of data or extrapolations, are critical to evaluating the model. They should be experimentally verified if possible. What would you estimate the light transmitted for a concentration of 0.7?

Now open the **[WINDOW]** by setting Xmax = 0.9. This will allow you to see how the models work for predictions at higher concentrations.





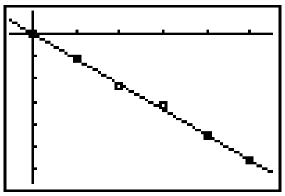


LINEAR goes negative! | QUARTIC goes up!!! | EXPONENTIAL matches your prediction!

LINEAR	QUARTIC	EXPONENTIAL
goes negative!	goes up!!!	matches your prediction!

A quadratic model would do about the same thing as a quartic, except the upturn is not as fast. Exploring the predictive capabilities of a model are crucial. In this case the quartic model breaks down at higher concentrations.

Now let's think about the manual function fit approach. For an exponential function, the data can be transformed to linear by taking the log of the y data, while for a power function the log of x and the log of y will produce a linear relationship. The transformed data yields the following: for exponential, $y = 10^{kx}$ (or $y = e^{kx}$ if ln is used) and for power, $y = x^k$ with k = slope in both cases.



Here is the graph of log T against concentration for our model above. A linear regression line is shown through the data. The slope is -1.12, so our equation of the original data would be given by:

$$T = 10^{-1.12c}$$

where T is the fraction of light transmitted and c is concentration.

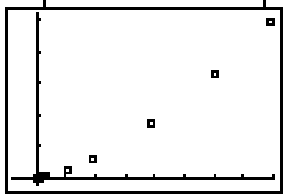
From a set of experimental data and a little guidance, we have discovered a mathematical relationship or model to describe the behavior of light transmitted through a series of varying concentrations of a specific solution. Understanding this relationship has importance in the chemical analysis of many substances.

DISCOVERY OF A RELATIONSHIP

Let's take a classical example of historical data to demonstrate modeling and guided inquiry. The table below lists the nine planets in our solar system along with their distances from the Sun, in AU (relative to Earth at 1 AU), and periods, in Earth years, or the time to revolve once around the Sun. This data is available in most any astronomy textbook.

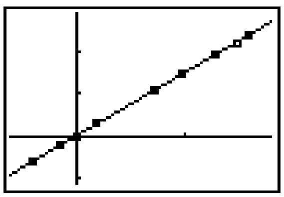
Planet	Distance, d (AU)	Period, P (years)
Mercury	0.3871	0.24084
Venus	0.7233	0.61515
Earth	1.0	1.0
Mars	1.5237	1.8808
Jupiter	5.2028	11.862
Saturn	9.5388	29.456
Uranus	19.191	84.07

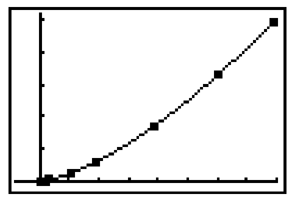
Neptune	30.061	164.81
Pluto	39.529	248.53



Is there a relationship between distance and period? Plot a graph of period on the y-axis against distance on the x-axis. Is the plot linear?

What type of relationship do you think exists? Try some regressions to decide on a fit.





After a few trials you probably discovered this plot is fit by a power function. The log-log plot is linear with a slope of 1.5. Thus the function $P = d^{1.5}$ fits the data.

log-log plot   power function

This is Kepler's third law of planetary motion discovered in 1619. In most astronomy books the standard form given is by $P^2 = d^3$ (the function above squared), where P is in Earth years and d in AU's. Kepler only had five planets to work with and no electronic devices to help with calculations.

Like Kepler, we now have discovered a mathematical model or an equation that relates period and distance. Next we want to test our model. How good is it? We want to make predictions and if possible verify them. In 1687, Newton verified Kepler's third law using the moons of Jupiter and then Saturn. The asteroid belt is located between the orbits of Mars and Jupiter at a typical distance of 3 AU. What should be the period of these asteroids?

THE PROCESS OF DEVELOPING A MATHEMATICAL MODEL

After experimentation, a scientist wants to examine the data collected for a pattern or mathematical relationship to explain the data. Based on what we have explored so far, let’s look at a general process to use when modeling a system.

Modeling data consists of four steps:

1. Generate a scatter plot of the data collected.

2. Pose the question: Is there a relationship between the two variables plotted? If so, can a regression analysis fit the data? Perform a number of different regressions on the data to see which fits best. Extend the **[WINDOW]** to see how the regression equation behaves at higher x-values. Fitting a mathematical function is an alternate to regression analysis.

3. Determine how well the regression or function fits. A number of simple methods to judge the goodness of fit can be used including:

   ‘ If you use the box symbol (**9**) on plots with the TI-83, when the curve passes through the box it darkens (**#**). Thus a simple, visible goodness of fit is determined by seeing how many data points fall on the curve.

   ‘ The **coefficient of determination**, r^2, is the proportion of the TOTAL variability of the y-data that can be explained by the x-variable. The limits of the coefficient of determination are 0#r^2#1. A value of 1 means all or 100% of the variation of y is explained by x.

   ‘ The **residual** is the difference from the actual y-value and the y-value obtained from the regression equation when the appropriate x-value is placed in the equation. It is the error between the measured y-value and that predicted by the model or regression equation.

$$\text{residual} = Y_{\text{actual data value}} - Y_{\text{regression equation}}$$

   Residuals, or errors as they are called by some authors, can be used with a mathematical function to judge fit. However, you would have to calculate the residuals in the list editor by taking the difference between the actual data value and the value computed from the manually fit function.

4. Use the model to make predictions. Do the predicted results make sense or does the model break down?

An important goal in science education is to provide students with the understanding and tools to pursue investigations of systems on their own. The computer with interfaces and probes or the CBL with probes and the graphing calculator have allowed rapid and extensive collection of data.

Using the graphing calculator or computer to process the data into a model provides the opportunity for students to develop analysis, synthesis, and interpretation skills and reinforces the critical link between science and mathematics in our world. Together data collection and mathematical modeling open doors for students to "discover" more about how everyday systems behave.

RESOURCES

D. Cathcart and T. Horseman, A Mathematical Modeling Course for Elementary Education Majors: Instructors' Thoughts, in *Journeys of Transformation*, Maryland Collaborative for Teacher Preparation, M. Gardner and D. Ayres, ed., UMCP, 1998, pp. 71-85.

D. Kalman, *Elementary Mathematical Modeling: Order Aplenty and a Glimpse of Chaos*, Mathematical Association of America, 1997. (An excellent introduction to modeling using difference equations.)

S. Sinex and B. Gage, *Mathematically Modeling Scientific Data: Using Your TI-83 Graphing Calculator to its Full Abilities in Science* presented at Chemistry Educators of Maryland (ChEMd) Workshop, 17 January 1998. (Step-by-step data analysis handout available from authors.)

Appeared in: MAST Rapper 23 (2), 24-33 (1998)