**Case Study: Modeling with Elementary Functions**
©2002 Donald Kreider and Dwight Lahr

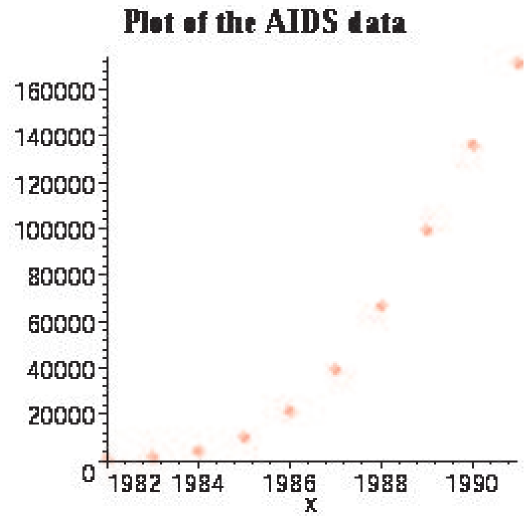
**Animation:** Modeling with Elementary Functions **To get you going on the Case Study!**

We close the first chapter on the elementary functions with a *Case Study in Calculus* (CSC). The purpose of a CSC is to consider a real application of calculus, with real data. In a CSC, we act like mathematicians working with scientists from other areas to answer questions in their fields. All of the answers involve using some of the calculus tools we are learning. One way to think of the CSCs is as extended homework problems. But they are more than that because they show the power of calculus at work in other disciplines. The concern we want to address in this CSC involves the AIDS data introduced earlier.

The spread of the AIDS virus in the United States during the first ten years following its discovery in 1981 is given in a table published by the Centers for Disease Control:

| Year | No of AIDS cases |
|---|---:|
| 1982 | 295 |
| 1983 | 1374 |
| 1984 | 4293 |
| 1985 | 10211 |
| 1986 | 21278 |
| 1987 | 39353 |
| 1988 | 66290 |
| 1989 | 98910 |
| 1990 | 135614 |
| 1991 | 170851 |

The table defines a function $AIDS(n)$, that gives the cumulative number of AIDS cases up to the year $n$. Here is a graph of the data.




Our interest is in finding a continuous function $f(x)$ whose graph approximates the data points and runs smoothly between them. Moreover, the function should be suitable for making future predictions. Such a function $f(x)$ is said to *model* the data. To obtain $f$, we have seen that mathematically a least squares fitting procedure is often the method of choice.

However, we want a function that both fits the points mathematically and is also based on known scientific principles. In the AIDS study, those principles are epidemiological. This is a good example of how mathematicians work with scientists in other fields to solve important problems.

But suppose we fit, for example, a cubic polynomial to the data points. And alternatively we fit an exponential function to the data points. The exponential appears appropriate because the graph of the AIDS data resembles the many examples of exponential population growth so often pictured in environmental publications. (We will study such population models ourselves later on.) How do we decide which fit is better?

There are of course two considerations. First, let's discuss the mathematical one. Because we are using least squares techniques to fit curves to the data, we are going to use the sum of the squares of the errors as a measure of the goodness of fit over the time period in question. That is, suppose we start with data points $(t_1, c_1), (t_2, c_2), (t_3, c_3), \ldots, (t_n, c_n)$ and we fit the data with the function $y = f(t)$. Then we will define the *sum of squared errors* over the period to be the sum of the individual squared errors at the data points, namely,

$$\sum_{i=1}^{n} (f(t_i) - c_i)^2$$

Moreover, we can use the sum of squared errors to develop a criterion for which fit is better.

**Mathematical Criterion for the Best Fitting Function**: One fitting function will be better than another if its sum of squared errors is smaller.

Squaring the errors does not allow the signed-errors to cancel one another in the summation; sqaring also has the effect of giving more weight to bigger errors than smaller ones.

Now that we have a mathematical criterion for deciding between fitting functions, we need to consider if there is a scientific one. Continuing with the AIDS example, one might be tempted to choose an exponential function because as we will see later, exponentials are associated with many population studies. However, it was reported in medical journals a few years ago that the epidemic does not seem to be following an exponential model. Indeed, it appears to be behaving very differently from other viral epidemics such as influenza epidemics. Rather, it appears to be best modeled by a cubic function, and several papers in the medical literature argue that this is to be expected for diseases like AIDS that have long latency periods and for which homogeneous mixing of the virus throughout the population is inhibited by variations in behavior of different groups of people. Thus, from a scientific standpoint, we should seek a cubic function that best fits the data.

And this brings us to the objective of this CSC.

**Objective**: The purpose of this CSC is to fit both a cubic polynomial and an exponential to the AIDS data and compare the results mathematically. We want to determine if in this case, the mathematical criterion of *best fit* gives the same answer as the epidemiological one. We also want to make predictions about AIDS cases in the future.

Before we can meet the objective, we need to know how to fit an exponential function to a data set.

**Fitting an Exponential by Least Squares**: Most Computer Algebra Systems such as Maple or Mathematica cannot fit an exponential function to data directly. Instead, we must fit a line to the semi-log data. That is, given the $(x, y)$ data, we fit a line to the $(x, \ln y)$ data. Suppose we carry out that process and obtain the fitting line $l(x) = mx + b$. Note that if $y = Ce^{kx}$, then $\ln y = \ln C + kx$. Thus, because we fit the semi-log data, we can determine the $C$ and $k$ of the exponential fitting function we seek: $\ln C = b$ and $k = m$. So, the exponential fitting function is

$$y = e^{mx+b} = \left(e^{b}\right) e^{mx}$$

**Back to the CSC**: There will be a CSC at the end of each section of the book. We have structured them all the same way, as reports that you will complete. The reports teach you how to analyse scientific modeling problems systematically. Every report will include a concluding written part because it is very important for mathematicians to be able to convey their findings and recommendations to a general audience. Below are the parts of the report that you will be completing.

**Setup**: The purpose of this part is for you to write a clear statement of the mathematics that will be applied to the problem. What is the nature of the information that is available? What are the steps that you should follow to analyze the information? What tools will you bring to bear in the analysis?

The setup phase in tackling a scientific problem is often called the *modeling phase*. Mathematical tools are chosen to apply to the problem. Often the setup involves deriving an equation or equations that express the physical aspects of the problem in mathematical language. It is these equations, then, that would be called the mathematical model for the physical problem. Once the physical problem has been thus expressed in mathematical terms one can apply the theories and techniques of mathematics toward finding solutions.

In our present AIDS study, you should indicate how to obtain a polynomial and an exponential fit of the data, and what steps you will take to carry out the objective.

**Thinking and Exploring**: Now it is time to think about the mathematics. Carry out the mathematical ideas that you outlined in the setup. What is the third degree best fitting polynomial? What is the best fitting exponential function? Which of these provides a better mathematical fit and why?

**Applet:** Fitting AIDS Data **Try it!**

In June, 2001 the Centers for Disease Control reported that there were 274,624 people in the US living with AIDS in 1998, 299,944 in 1999, and 322,865 in 2000. Would your model have predicted these numbers? What is the percentage error for each year?

**Interpretation and Summary**: Now that you have done the mathematics and explored the models, it is time to interpret and summarize the mathematical results in terms of the original objective.

Pretend that your synopsis is going to appear in the next issue of a magazine such as *Scientific American*. Include enough details so that a reader would learn what the major issues of the report are, and how you went about addressing them. What will you want to tell readers about your success with regard to the original stated objective of the investigation? Be sure to write in complete sentences using correct rules of standard English grammar. Make the write-up interesting and informative. Would the mathematical answers alone be sufficient to model the AIDS data? How good a predictor is your model of future AIDS cases? Is there further analysis that needs to be done?

**Exercises:** Problems **Check what you have learned!**

**Videos:** Tutorial Solutions **See problems worked out!**