Principles of Calculus Modeling: An Interactive Approach by Donald Kreider, Dwight Lahr, and Susan Diesel
Exercises for Section 1.7

1. (1 pt)
Setup

This is the setup section that looks at the mathematics behind modeling the data.

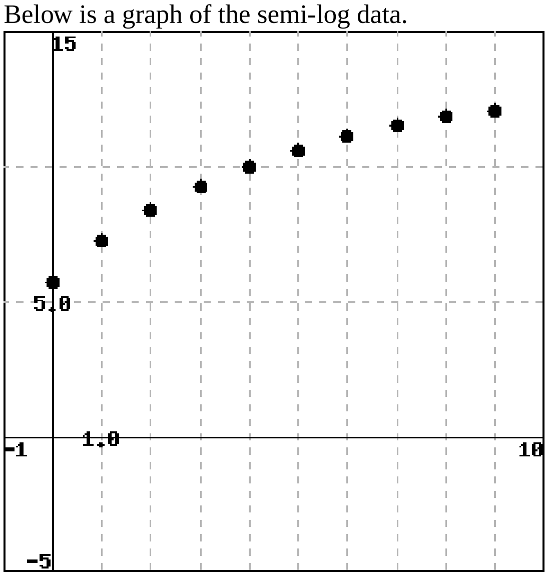
Recall that the spread of the AIDS virus in the United States during the first ten years following its discovery in 1981 is given in a table published by the Centers for Disease Control:

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

Your interest is in finding a function that fits the data points and passes smoothly between them. The candidates are an exponential function, of the form $f(x) = Ce^{kx}$, and a cubic polynomial, of the form $c(x) = a_0 + a_1x + a_2x^2 + a_3x^3$.

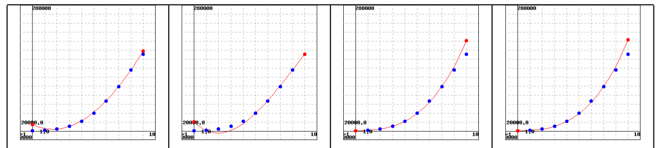
An equivalent way of describing the exponential function is $y = e^b e^{mx}$. The *semi-log data* is the set of points $(x, \ln y)$. If the original data are described by the exponential $y = e^b e^{mx}$, then the semi-log data are described by the line $y = b + mx$.

Below is a graph of the semi-log data.



Do you believe an exponential function will fit the data well? You may want to address this question in your Interpretation and Summary write-up in problem 4.

Consider the four graphs below. Each shows the data in the table, and each includes a possible cubic curve that passes through 4 of the 10 points. The points used are listed below each graph. Because of the magnitudes of the data, the years are renumbered so that 1982, the first year of the study, becomes year 0, 1983 becomes year 1, and so on.



2,3,7,8	1,5,8,9	1,2,5,7	0,3,4,7

The coefficients of the cubic polynomials graphed above, not necessarily in the same order in which the graphs are displayed, are:

A. a0 = 14383.6, a1 = -12699.667, a2 = 3942.35, a3 = -57.583
B. a0 = 20216.393, a1 = -26034.799, a2 = 7497.402, a3 = -304.996
C. a0 = 461.417, a1 = 384.858, a2 = 289.983, a3 = 237.742
D. a0 = 295, a1 = 504.94, a2 = 178.25, a3 = 251.738

The goal of fitting a curve to a set of data is to come close to the points with the curve. The sum of squared errors

$$\sum_{i=0}^{9} (y_i - c(x_i))^2$$

is a measure of how good the fit is. The desire is to make this sum as small as possible.

To find a best-fitting least squares cubic polynomial, or line, computers rely on finding partial derivatives and solving simultaneous equations. The resulting curve or line may not pass through any of the data points, even though the sum of squared errors is as small as possible. These techniques are beyond the scope of the course. However, in the next section you will be expected to use such computer tools without regard to the method the computer employs.

Give the sum of squared errors for each set of coefficents. Remember that the years are numbered 0 through 9.

A: ____________________
B: ____________________
C: ____________________
D: ____________________

Which polynomial has the best fit, determined by the smallest sum of squared errors?

___

---

2. (1 pt)
Thinking and Exploring

Find the least squares exponential function and cubic polynomial that fit the data. The method used in this problem is solving a matrix equation. You may choose to use alternate methods, including use of a computer program, to find the answers.

Year	renumbered year (i)	No. of AIDS cases
1982	0	295
1983	1	1374
1984	2	4293
1985	3	10211
1986	4	21278
1987	5	39353
1988	6	66290
1989	7	98910
1990	8	135614
1991	9	170851

Part 1

As stated in problem 1, finding the best-fitting exponential function for the data is equivalent to finding the best-fitting line $y = b + mx$ for the points $(x, \ln y)$. Minimizing the sum of squared errors

$$\sum_{i=0}^{9}(\ln(y_i) - (b + mx_i))^2 = \sum_{i=0}^{9}(\ln(y_i)^2 - 2(b + mi)\ln(y_i) + (b+mi)^2$$

where we have replaced x_i by i, gives rise to a system of 2 linear equations in the two unknowns b and m:

$\sum_{i=0}^{9} b + \sum_{i=0}^{9}(mi) = \sum_{i=0}^{9}(\ln y_i)$

$\sum_{i=0}^{9}(bi) + \sum_{i=0}^{9}(mi^2) = \sum_{i=0}^{9}(i \ln y_i)$

Simplify as necessary to fill in the missing numbers in the following two linear equations. Enter your answers with 6 decimal places of accuracy.

__________ b + 45 m = __________

__________ b + __________ m = 494.513972

Solve the system for the unknowns b and m. Enter your answers with at least 6 decimal places of accuracy.

b = ____________________

m = ____________________

Compute the sum of squared errors $\sum_{i=0}^{9}(y_i - e^b e^{mi})^2$.

____

Part 2

Use the applet or Maple to determine the coefficients of the best-fitting cubic polynomial.

$a_0 =$ ____________________

$a_1 =$ ____________________

$a_2 =$ ____________________

$a_3 =$ ____________________

---

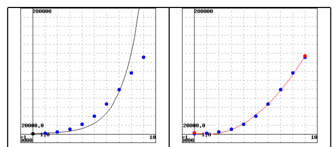
3. (1 pt)

Using the Model to Predict Future Cases

Year	renumbered year (i)	No. of AIDS cases
1982	0	295
1983	1	1374
1984	2	4293
1985	3	10211
1986	4	21278
1987	5	39353
1988	6	66290
1989	7	98910
1990	8	135614
1991	9	170851

Having found the best-fitting exponential function and the best-fitting cubic polynomial, you must determine which provides a better mathematical model for the data. Not only should the model fit the existing data well, it should be the best predictor of future AIDS cases.

Graphed below are the best-fitting exponential function and the best-fitting cubic polynomial as determined by the least-squares method from problem 2, and the data for the years 1982 through 1991. As in previous problems, the years are renumbered 0 through 9.



exponential | cubic

Which model appears to fit the data best?

___

A. the exponential function
B. the cubic polynomial

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. What does the best-fitting exponential function predict for the number of cases in each of these years? Use the exponential function computed by the applet.

Exponential:

Predicted cases in 1998 = ____________________

Predicted cases in 1999 = ____________________

Predicted cases in 2000 = ____________________

What does the best-fitting cubic polynomial predict for the number of cases in each of these years?

Cubic:

Predicted cases in 1998 = ____________________

Predicted cases in 1999 = ____________________

Predicted cases in 2000 = ____________________

What is the percentage error for each year?

Exponential:

percentage error in 1998 = ____________________

percentage error in 1999 = ____________________

percentage error in 2000 = ____________________

Cubic:

percentage error in 1998 = ____________
percentage error in 1999 = ____________
percentage error in 2000 = ____________

---

4. (1 pt)

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?

To submit your answer, use the Email instructor button below. Include your email address, and enter your report in the Feedback box. When you are satisfied with your composition, click the Send Feedback button. **Your write-up should be no longer than one page.**

When you are done, return to this screen and complete the Affirmation below.

Affirmation: Even though I may have discussed the CSC project with other people, I have written up this CSC report by myself and on my own. No sharing of electronic files or notes has been involved.

Please type your name in the answer box, just as it appears in the WebWorK database and on the problem list screen, and click the Check Answers button.

____