Math241

Project 1 (Chp 2) Honors Assignment

Complete the following:

This investigation focuses on the spread of the AIDS virus. As with most disease data, AIDS proliferated quickly in its initial stages but eventually the rate at which it spread declined significantly. AIDS poses an unusual challenge in modeling. Currently no inoculation or cure is available so the afflicted population decreases only through mortality. Unlike cold viruses transmission of AIDS requires more than incidental contact and has significantly more dire consequences. Therefore its spread is apt to be influenced more by public awareness and behavior (conscious or otherwise).
You will conduct this study in several stages, initially deriving regression models for both annual infection as well as cumulative growth over the first six years the disease was monitored. You will follow this with models based on more complete data and then revise these to accept additional relevant information.

Begin by referring to the HIV/AIDS Surveillance report for December 2001 (you can download the pdf from http://www.cdc.gov/hiv/stats/hasr1302.htm or from my website). Read the commentary beginning on page 5. On page 30 you will find Table 21, *AIDS cases and deaths, by year and age group, through December 2001*.

I. a) Using a spreadsheet or TI-83 enter data for the following sections from 1981 through 1987:

L_1: year (1981 = year 0)

L_2: Annual cases diagnosed (adolescent and adult, why do you suppose cases are separated from ≤ 13?)

Plot the data for annual cases versus time and find the following regression equations:

- Cubic
- Exponential
- Logistic

Record each regression formula and determine the function that you believe best fits these data. Explain your choice citing mathematical inference and logic. Discuss the following concepts in your analysis: According to your data, where should the function grow fastest/slowest? Does the regression model you chose match this? Beyond the domain of the data you used, does your model behave in a way consistent with the context (e.g. increasing/decreasing, remaining positive). Using this function, estimate the number of people infected by 2001. How does this compare with the actual recorded value?

b) Now add the following list of values: L_3: Cumulative cases diagnosed.
This is a running total of cases diagnosed up to a particular year.

Plot the data for cumulative cases versus time (L_3 vs. L_1) and find the following regression equations:

- Cubic
- Exponential
- Logistic

Again record each regression formula and determine the function that you believe best fits these data. Explain your choice as above. Using this function, estimate the total number of people infected by 2001. How does this compare with the actual recorded value?

II. a) Complete the tables $L_1 – L_3$ for 1981 through 2001 and repeat the exercises in (I). (If you prefer, you can get a more current estimate by downloading the data from 2001 – 2004, also on the CDC website, or from mine.

b) Discuss how predictions made using your model in part (I) might influence AIDS policy in the US in 1987. If you were making policy decisions during that period, what position would you adopt - conservative (don't panic the population, this disease is limited to a small segment of society) or more reactive (if we're not careful this could get out of control, it's better to err on the side of caution)? Which regression model would you choose to support your position?
Repeat this exercise for (II) - in light of current data, what is your position and what would you do mathematically to reinforce this position?
Finally, discuss the shortcomings and compensations inherent in using limited data to form inferences.

$\boxed{\text{III.}}$ Modeling is a process of approximation and refinement. You may have already observed that our regression formulas assume a perpetually infected population ignoring the mortality of those afflicted. In order to make our infected population data more accurate we should subtract the number who died for that year (again, refer to Table 21). Revise the cumulative data table to reflect this.

Produce the plot for the revised data vs. time and again generate the following regression equations.

- Cubic
- Exponential
- Logistic

Again record each regression formula and determine the function that you believe best fits these data. Explain your choice as above.

$\boxed{\text{IV.}}$ Using your cumulative distribution model from (III) answer the following questions.

- Determine (approximately) the rate of growth in 1981, 1991, and 2001
- Determine the average growth rate over the interval [1981, 2001] and remark on how well it approximates the rate at which AIDS spread during this period.
- Determine all times when the instantaneous growth rate roughly matches the average growth rate.
- What does your model predict the total US population living with AIDS will be in 2010?
- What does your model predict the rate at which the disease spreads will be in 2010?
- Remark on the accuracy of your model, how well do your results fit the data and how well do they predict future values?

$\boxed{\text{V.}}$ Return to your cumulative models in (I) and (II). Determine the growth rate of the disease for both in 1987. How does the instantaneous rate of change in 1987 compare with the average rate of change for [1981, 1987] in both cases? Comment on how these different results might influence public reaction.

$\boxed{\text{VI.}}$ Summarize your results in a paragraph commenting on the accuracy of each model subject to constraints of time, data available, new treatments and education, public policy, mathematics, and so on. Suggest ways in which improve these models and offer one such example.