P4 Population of California

You have two sources of data available for this project:

- Data from the United States Summary on Population and Housing Unit Counts (http://www.census.gov/population/www/censusdata/pop-hc.html)
  On this page, go to heading **3. Population and Housing Counts: 1790 - 1990, from CPH-2-1.** Click on **2. Population: 1790 to 1990** to download an Adobe Acrobat file named table 16.pdf, which contains the relevant data.
- Current Population Reports (http://www.census.gov/prod/2/pop/p25/p25-1131.pdf)

Your goal in this project is to find a model for the number of people living in California.

1. Exponential Model (using the data for 1850 - 1930)

Overview: The first model you will consider is an exponential or Malthusian model, which we have discussed in class. The basic assumption for this model is that the change in population (births - deaths) is proportional to the size of the current population.

a) Enter the data for the years 1850-1930 from the United States Summary on Population and Housing Unit Counts and graph it. (Associate 1850 with n = 0.)

b) Define the meaning of your input and output variables and indicate their units. (Make sure these units agree with the way you entered the data in part a).)

c) Use the paradigm of **new = old + change** to set up the iterative model equation.

d) The iterative model equation derived in part c) contains a parameter, the growth factor, which also shows up in the general solution. In order to predict California's population, you need to estimate this growth rate. Use the data for the years 1850-1930 to come up with a value for the growth rate per time unit (as chosen for your input variable).

e) Using the growth rate derived in part d), state the general solution for the analytic model derived in part c).

f) Fit an exponential model to the data for the years 1850-1930, using the function **ExpoFitGraph**. From the function **ExpoFitFunc**, read off the growth factor per time unit. Compare this to the value you derived in part d). Are they similar? very different?

g) You have now derived two different models: One model was derived from assumptions (in part e)) and the other using a least squares fit (in part f)). Compare these two models using the function **FitComp**, and decide which one is the better model.

h) Using the best model found in part g), predict the population of California for the year 1990. How does your prediction compare to the actual data?

i) As you have seen in part h), the model prediction is not very good. Can you explain the reasons? What are, in general, the limitations of an exponential model?

j) The data seems to show some irregularities. Can you think of historical reasons for these anomalies?

2. Logistic Model (using the data for 1850 - 1990)

Overview: As you have seen in part 1, there are some serious limitations of the exponential model. You will now adapt the model to (hopefully) derive a more realistic model. This revised model follows the ideas of Verhulst. Instead of assuming that the change in population is proportional to the size of the current population (which would lead to unlimited growth), he assumed that there was an upper limit to how large the population could grow, the so-called carrying capacity L. Calling the difference between carrying capacity L and current population the "unused growth potential", he postulated that

> **the change in population is <u>jointly proportional</u> to the <u>current population</u> and the <u>unused growth potential</u> (*)**

a) Enter the full data set (1850 - 1990) and plot it.

b) Fit an exponential model to the larger data set, using **ExpoFitGraph** and **ExoFitFunc**. Compare this exponential fit to the one you derived in part 1 f). (How well do these functions fit the data, how do the functional expressions from **ExpoFitFunc** differ?)

c) From the graph determine what type of function should be fitted to the data to get a better fitting model. Give reasons for your answer (using shape of the graph context, numerical tests if available).

d) You will now derive analytically the logistic model. Translate the verbal statement (*) into a mathematical expression using the fact that a quantity x is jointly proportional to quantities y and z, if there is a constant c such that $x = c \cdot y \cdot z$. Using the paradigm **new = old + change**, state the iterative model equation for the logistic model.

e) To use the logistic model derived in part d) for predictions, the value of the carrying capacity L needs to be determined. Use the data or its graph to estimate (if possible) where the population size will level off.

f) Fit a logistic model to the larger set of data (1850 - 1990). Compare the logistic fit with the exponential fit of part 2 b) using **FitComp**.

g) Using the function **LogisticFitFunc**, predict the population of the State of California in the years 1995 - 2025 (every five years). How do these predictions compare to the values in Table 1 (Current Population Reports)? What other factors may be included in the Census Bureau model that have not been addressed by your model?

h) Finally, describe a situation for which a (possibly more sophisticated) model on the population size of California may be used.