MATH 442: Mathematical Modeling

Lecturer: Dr. Jean Marie Linhart
http://www.math.tamu.edu/~jmlinhart/m442

Project 2: Population Modeling (P2)
Writing 3 (W3)

Instructions You are to work together in groups of 3 to complete the following project. You each have to do write your own Matlab. You are to split the writing tasks and critique each others work, bringing everything together into one coherent report at the end. Keep track of who does what and how each person contributes to this assignment; this will also be handed in!

References on these mathematical techniques are plentiful. Your job is to explain your topic in your own words. You will probably use references to review the topics. You should cite this reference (or these references) in your bibliography. If you quote from a reference, you must use quotes and a citation. If you paraphrase closely or take a problem from a reference, you use a citation. Avoid excessive quotation and paraphrasing. To accomplish this, as much as possible do not look directly at a reference while writing up your explanation, so that you use your own words. If you need to, make an outline of the major ideas and then refer back to your outline while you are writing.

Things to check before handing in your project:

1. Have you done everything you were asked to do?
2. Is the math right? If not, get it fixed.
3. Is what you are trying to explain clear?
4. Do your graphs clearly show the conclusions you reach?
5. Are the spelling, grammar, and English usage correct and concise?
6. Do you think this is A, B or C level work?
7. Run the document through a spell checker!
8. Have the Writing Center or some other good writer read your document and critique it; then edit your document taking those comments into account.

By the end of this project, you and your partners should have a thorough understanding of curve fitting, population models, and how to evaluate the models.

Population Modeling We are going to investigate a variety of mathematical models for population growth in this project.

Partner 1 will write the introduction and will hand in MATLAB code that makes all graphs and analysis for the US population. Partner 1 should **not** have written the introduction for Project 1.

Partner 2: will write the methods section and a results and conclusions section for his/her population dataset.

Partner 3: will write the results and conclusions section for the US population dataset and for his/her population dataset.

The goal is for each person to do two pages of writing, not including equations or graphs. The group as a whole is responsible for writing the abstract and a general conclusions section.

You and your partners will model the US population data from section 1.9 of *A Concrete Approach to Mathematical Modeling* **plus** whatever additional US census data that has been published since the book was published. Your group gets to pick two other datasets to model: a US state or a foreign country or a city (US or foreign). Choose these wisely: history, like World War 2, played some havoc in population figures in Europe and in the Pacific. There may be no overlap in the class, everyone must choose something different. You will use your incredible internet and library reference skills to find as much population data from your two choices as possible. You want to go back as far as you can; with a city in France, for example, you would want to go back to at least 1800, preferably earlier. If you were to pick Alaska or Hawaii, you would not be able to find as much data (and you may have to do more explanation with your curve fits). Each of you is responsible for the data collection, programming and analysis for your choice. Your partners may assist, but you bear primary responsibility for one of the three datasets and its analysis.

Population data must come from a reliable source, e.g. http://census.gov, not Wikipedia.

You will fit three growth models to the three data sets you choose using least-squares curve fitting. The three models are:

1. An exponential growth model
$$P = Ae^{rt}$$
2. A logistic growth model
$$P = \frac{L}{1 + Ae^{-rt}}$$
3. A Gompertz model
$$P = Le^{-be^{-rt}}$$

You will use these models to fit the population data. The error used to evaluate the model is the R^2 value given by the **coefficient of determination**.

$$R^2 \equiv 1 - \frac{\mathrm{SS_{err}}}{\mathrm{SS_{tot}}}$$

If the data is $\{y_i\}$, with mean $\bar{y}$, SS_{err} is the sum of the squared errors and $\text{SS}_{\text{tot}} = \sum_{i=1}^{N}(y_i - \bar{y})^2$.

Your modeling goals are:

I. Determine which model has the best fit overall for the data sets. You cannot do this "by eye". You must have a numerical measure for best curve fit.

II. Using as much data as you can get before 2000 find the best fit using only this data, and predict the population 2010 (or another similar year). Determine which model(s) has (have) the best prediction(s). The error associated with the prediction is a (signed) relative error:

$$\frac{P(t_i) - y_i}{y_i}$$

III. Determine if one model clearly better than the others.

The general conclusions section should contain two tables, one of the goodness-of-fit data, a second of the population predictions and accuracy. Tie it all together with a concluding paragraph.

You should indicate in your report who was primarily responsible for what, so that I can evaluate the individual part of your grade. Your final report should include:

(a) A 150-250 word abstract summarizing your paper. Include the main conclusions of the paper.

(b) An introduction. In the introduction you should:

   i. Explain the theoretical difference between the exponential model and the logistic/Gompertz model with respect to population. To do this you will probably want to introduce the growth rate; make sure this is defined consistently for all three models. (Logistic and Gompertz are very similar; they can be discussed together.)

   ii. Does one model make more theoretical sense than the others for population modeling in any the cases you are looking at? If you can, predict which you think will give the best fit and give reasons why. Having a reason to believe something, but finding out later that the data doesn't support your argument **will not result in a loss of points. This is good!** This is the scientific method at work. We make hypotheses and sometimes they are not validated by our experiment.

(c) A methods section explaining how you got your population data, how you will fit the three models, and how you will evaluate the models and their predictions.

i. Show how to put the exponential, logistic and Gompertz equations into the form where we can can do a linear least-squares curve fit with the data. In the case of the logistic and Gompertz, we need to also make an initial assumption of the maximum population, and this is an initial fit, that is later refined using MATLAB 's `lsqcurvefit` routine.

ii. Tell how your group evaluates the curve fits (everyone should use the same method.)

iii. Tell how your group makes and evaluates the predictions (everyone should use the same method).

(d) A results/conclusions section for each dataset in which you include

i. One graph of the data and the the model fits, a second graph of your model fits omitting the final data point for the year 2010, and the model predictions for those data points. You should be able to clearly read the title, axis labels, and axis titles on your graphs in your report.

ii. Tables with the constants you find that are associated with your model fits, the errors in the model fits, and the predictions and errors in predictions. Two example tables:

Fit	Bryan TX data; Population in thousands		
Exponential	A = 0.503	r=0.026	
Logistic	L = 10.17	A = 37.39	r = -0.0421
Gompertz	L=29.2	b = -5.04	r = -0.0122

Bryan TX data				
Fit	R^2 error	Population 2010	Prediction 2010	Percent error
Exponential	0.95	65,660	76,394	16.35%
Logistic	0.998	65,660	65,480	-0.27%
Gompertz	0.998	65,660	68,036	3.62%

iii. Any criticisms/comments about the population models based on your data and/or "uncommon sense". A knowledge of history can be helpful here. (I.e. you should write the conclusions about your data.)

(e) A general conclusions section for the entire project which incorporates

i. Table(s) with

- The error associated with each model (the R^2 value).
- The predictions for each model and the error associated with the predictions. where $P(t)$ is your best exponential, logistic or Gompertz fit and y_i is the data point you are trying to fit.

Example:

Bryan TX data				
Fit	R^2 error	Population 2010	Prediction 2010	Percent error
Exponential	0.95	65,660	76,394	16.35%
Logistic	0.998	65,660	65,480	-0.27%
Gompertz	0.998	65,660	68,036	3.62%
Some Other data				
Fit	R^2 error	Population 2010	Prediction 2010	Percent error
Exponential	0.95	65,660	76,394	16.35%
Logistic	0.998	65,660	65,480	-0.27%
Gompertz	0.998	65,660	68,036	3.62%

ii. An evaluation of which model(s) performed best. This table above should allow an easy comparison.

iii. Analyze the meaning of the statement "these population models have too much memory to produce accurate predictions."

iv. Criticisms and potential improvements for these population models. What do these models not take into account?

(f) A bibliography where you cite your sources of information.

(g) Each partner submits his/her own Group Work Assessment with this assignment. You do not need to collaborate on these.

(h) Each partner individually submits a reflection, which can be in paragraph form or a list of what are the most important things you learned doing this project. Include some explanation as to why you selected these things. You may have learned some math, some MATLAB , some LaTeX, something about writing, something about working with others, etc.

Each of you will hand in a copy of your report along with the group work assessment and reflection. Submit your MATLAB files on eLearning. Submit your LaTeX files on eLearning. You will use TurnItIn on eLearning to submit the PDF of your final report. Include the group work assessment as an appendix to your report.

W3 Grading This assignment is worth 50 points. See the outline above for what is required in the report sections. Your audience is a student who has had calculus and differential equations, perhaps not remembering much differential equations aside from what an ODE and an initial condition is. Introduce ideas in a logical order. Explain in simple, friendly language. If you use jargon (technical terminology), explain it. Be consistent in what terminology you use throughout. You and your partners should strive to be consistent in the terminology you use in your report.

P2 Grading

- Abstract: 5 points

- General conclusions: 5 points
- Matlab code: 25 points

  The code should produce the following

  a) A graph of the data and exponential, logistic and Gompertz curve fits.
  b) The R^2 error for each of these curve fits.
  c) A graph of the data and exponential, logistic and Gompertz curve fits used for predictions (omitting the final data point)
  d) The predictions (numbers) and the relative error for each prediction.

  The person writing the introduction will have code that analyzes the US Data, the others will each analyze the US data and one other data set.

  Code should be well written. Names should make sense, organization should be logical. Code should not produce an excess of output. Each individual student hands in his/her own code. Get help from a partner or others as needed, but do not copy. Put your name in a comment at the top of your code.

- Overall appearance of your report: 5 points.
- Group work assessment: 5 points.
- Reflection on what you learned: 5 points.