MATH 2210: Project 2

Modeling US Population Growth

Sample Student 2

Introduction

In this project, I'll use least-squares approximation to calculate two different functions to estimate the United States population from the following United States Census data:

Year	1800	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000	2020
Population (Millions)	5.3	9.6	17.1	31.4	50.2	76.2	106.0	132.1	179.3	226.5	281.4	331.4

In my models, I will take the independent variable x to represent the number of decades since 1800, and I'll take the dependent variable y to be the population in millions.

Best-Fit Line

First, we wish to compute the best fit line for the data using least-squares approximation. That is, we wish to find coefficients r_0, r_1 for which the line $y = r_0 + r_1 x$ best fits our population data. To compute these coefficients, we use the normal equations $\mathbf{A}^\intercal \mathbf{A}\mathbf{r} = \mathbf{A}^\intercal \mathbf{y}$, with

$$\mathbf{A}^\intercal = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & 22 \end{bmatrix},$$

and $\mathbf{y}$ as the column vector of populations. The normal equations simplify to

$$\begin{bmatrix} 12 & 132 \\ 132 & 2024 \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \end{bmatrix} = \begin{bmatrix} 1446.5 \\ 24425.6 \end{bmatrix},$$

which have solution $r_0 \approx -43.191$ and $r_1 \approx 14.885$. Therefore, the equation of the best-fit line is

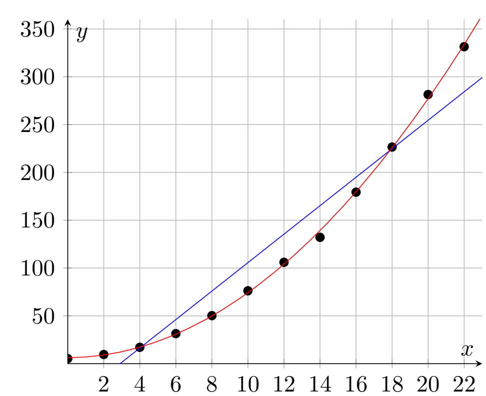
$$y = -43.191 + 14.885x.$$

Visualizing the Data

A graph of the population data is shown to the right. The best fit line has been superimposed onto the plot in blue.

The data shows a convexity (increasing steepness) that is not well captured by the best-fit line; points in the middle of the plot are below the line, and points at the edges of the plot are above the line.

I believe that adding a quadratic function to the basis will allow me to capture this steepening rate of growth.



A Model Including a Quadratic Growth Term

For the second model, I'll add the basis function $f_2(x) = x^2$ along with the linear and constant functions. My revised coefficient matrix is,

$$\mathbf{A}^\intercal = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & 22 \\ 1 & 4 & 16 & 36 & 64 & 100 & 144 & 196 & 256 & 324 & 400 & 484 \end{bmatrix}.$$

Plugging this into the normal equations, we obtain the linear system,

$$\begin{bmatrix} 12 & 132 & 2024 \\ 132 & 2024 & 34848 \\ 2024 & 34848 & 639584 \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 1446.5 \\ 24425.6 \\ 445675.2 \end{bmatrix}.$$

Solving this linear system, we find that $r_0 \approx 6.221$, $r_1 \approx 0.061$, and $r_2 \approx 0.674$, giving us the approximation

$$y = 6.221 + 0.061x + 0.674x^2.$$

This quadratic function is plotted on the previous page in red. This curve fits the data much better than the line.

(Note: You do not need to graph your second function in order to receive full credit on the project. Here it is done for illustrative purposes.)

Analysis of Models

To quantitatively analyze our models, we compute their coefficients of determination. The average population value from our samples is $\overline{\mathbf{y}} \approx 120.5$. Plugging in to the coefficient of determination formula, we find that

$$R^2 = 0.928$$

for the best-fit line, and

$$R^2 = 0.9994$$

for the quadratic approximation. Since both of these coefficients are greater than 0.9, we may conclude that both of the models do a good job explaining the trends in the data (even the best-fit line lies fairly close to the sample data). However, the quadratic model does significantly better, achieving a near perfect (1) coefficients of determination. This is likely because our quadratic model allows us to capture the phenomenon of the ever-increasing growth rate of the population.

Citations

I utilized demographic data from the Wikipedia page:

- "Demographic history of the United States." Wikipedia, Wikimedia Foundation, 22 July 2021, https://en.wikipedia.org/wiki/Demographic_history_of_the_United_States.

This data was compiled from various reports by the US Census Bureau.