Estimating the Population of the U.S.A

Samuel Tran, Andrew Lastinger, Jian Kang
MATH 442 - 930: April 30, 2020

Census Data of the United States (1790 - 1990)

Year	Population
1790	3,900,000
1800	5,300,000
1810	7,200,000
1820	9,600,000
1830	12,900,000
1840	17,100,000
1850	23,100,000
1860	31,400,000
1870	38,600,000
1880	50,200,000
1890	62,900,000

Year	Population
1900	76,000,000
1910	92,000,000
1920	105,700,000
1930	122,800,000
1940	131,700,000
1950	150,700,000
1960	179,000,000
1970	205,000,000
1980	226,500,000
1990	248,700,000

Table 1.1 US census data from 1790 to 1990.

Method I: Natural Logarithm

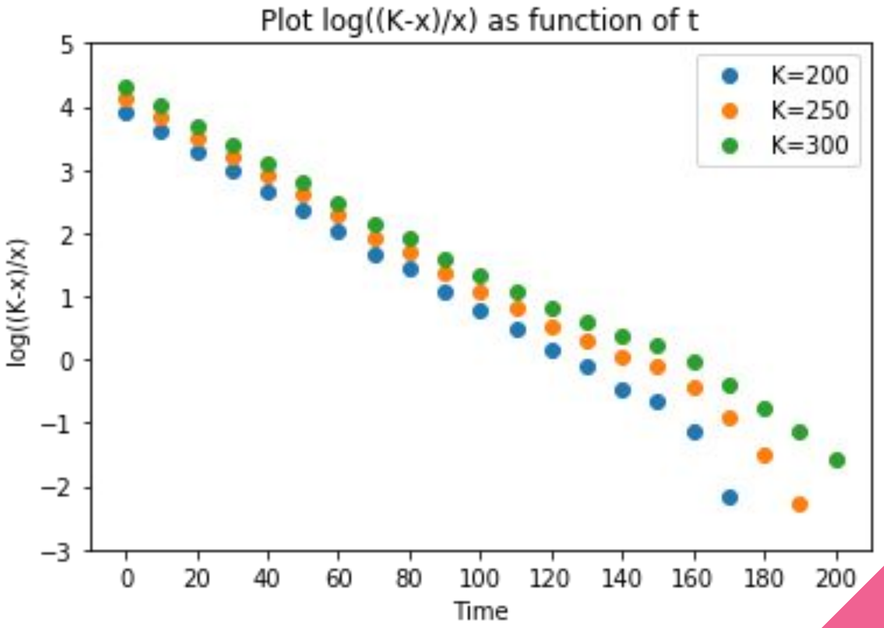
$$x(t) = \frac{K}{1 + \left(\frac{K - x_0}{x_0}\right) e^{-rt}}$$

$$\Rightarrow \quad \frac{K - x}{x} = \frac{K - x_0}{x_0} e^{-rt}$$

$$\Rightarrow \quad \log \frac{K - x}{x} = \log \frac{K - x_0}{x_0} - rt$$

Plotting the Data

Using the data, we plot $log\frac{K-x}{x}$ as a function of t,



Method of Least Squares

Use method of least squares to find the best straight line: $y(t) = a + bx$

Estimate the intercept a = $log \frac{K - x_0}{x_0}$ and the slope b = *-r*

Our goal is to minimize b = *-r*,

$$\sum_{i=1}^{n} (a + bt_i - y_i)^2$$

Solving for b,

$$\sum_{i=1}^{n} 2t_i(bt_i + a - y_i)$$

$$\rightarrow 2b\sum_{i=1}^{n} t_i^2 + 2a\sum_{i=1}^{n} t_i - 2\sum_{i=1}^{n} y_i t_i$$

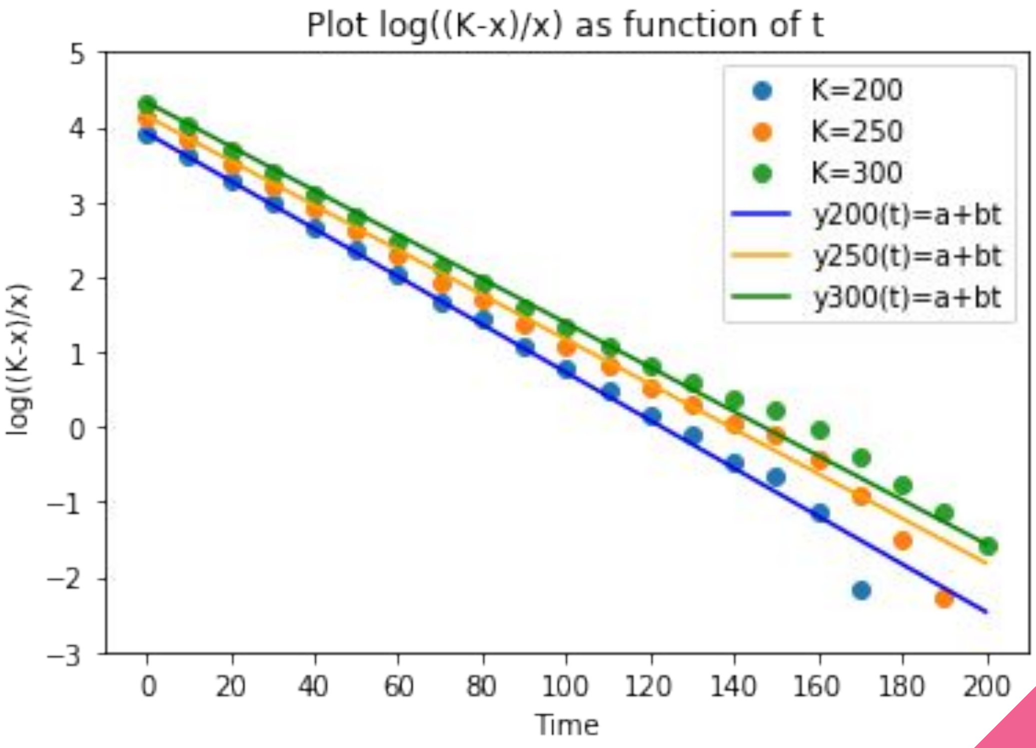
$$\rightarrow b = \frac{\sum_{i=1}^{n} y_i t_i - a\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} t_i^2}$$

Method of Least Squares

For x_0 = 3.9 (million),

Carrying Capacity K	**Line of Best Fit** y(t) = a + bx
K = 200	$y(t) = log(\frac{200 - 3.9}{3.9}) - 0.0319t$
K = 250	$y(t) = log(\frac{250 - 3.9}{3.9}) - 0.0298t$
K = 300	$y(t) = log(\frac{300 - 3.9}{3.9}) - 0.0294t$

Plot for the Best Fit Line



Prediction for Year 2000 Census

$$x = \frac{Kx_0}{x_0 + (K - x_0)e^{-rt}}$$

At year 2000, t = 210 years later from 1790:

Carrying Capacity K	**Population Size** x(t = 210)
K = 200	188,332,000
K = 250	223,046,000
K = 300	259,036,000

Method II: Estimate x’

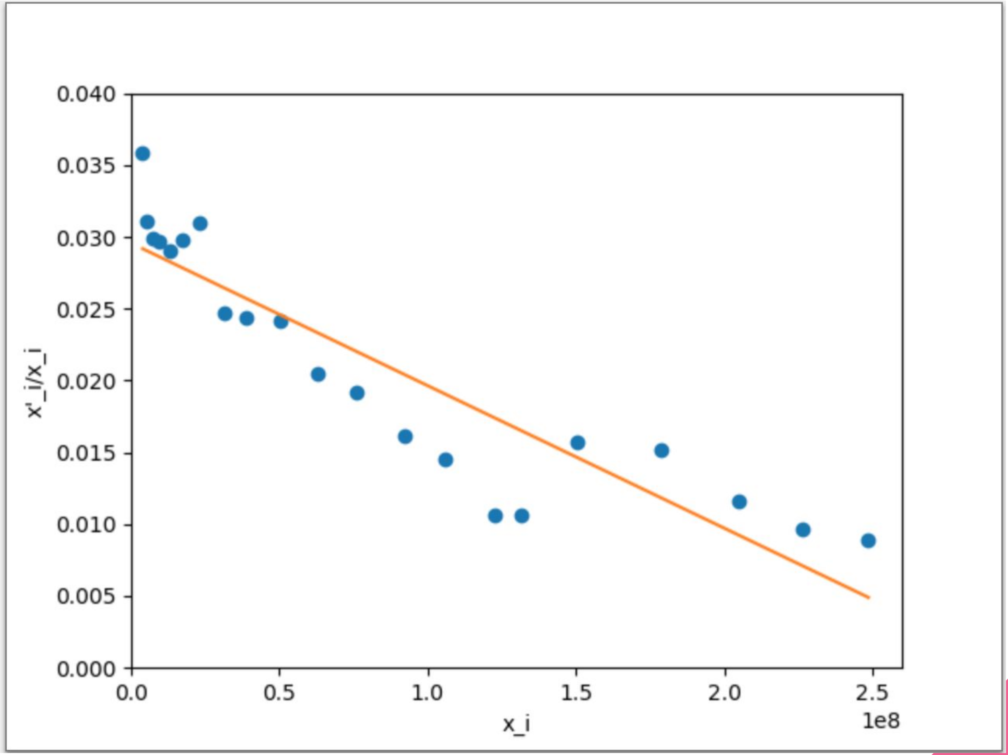
Logistic model: $$\frac{x'}{x} = r(1 - \frac{x}{K})$$

If we plot x’/x as a function of x, we will obtain a straight line with x-intercept K and slope -r/K, so that we could use the Census Data of the United States to estimate x’. We will use $\frac{x_{i+1} - x_{i-1}}{2h}$ to approximation x’ , The estimate will be calculated using the function below :

$$\begin{cases} \frac{x_{i+1}-x_i}{10} & i=0 \\ \frac{x_{i+1}-x_{i-1}}{20} & 1 \leq x \leq 19 \\ \frac{x_i-x_{i-1}}{10} & i=20 \end{cases}$$

Note: the reason we did this adapted version for i= 20 is because there is no available Xi+1 for this data point, so we adjusted the model to one before.
Now,we plot x'i/ xi against xi using our newly estimated values of xi.

Plot for x’/x against x



Method of least square

$$\frac{x'}{x} = r(1 - \frac{x}{K}) \quad = \quad -\frac{r}{K}x + r \qquad\qquad a = \frac{-r}{K} \qquad\qquad r = b$$

Using the method of least square, we could calculate a and b using the formula below:

$$a = \frac{\sum_{i=1}^{N} x_i y_i - \frac{1}{N}(\sum_{j=1}^{N} y_j)(\sum_{i=1}^{N} x_i)}{\sum_{i=1}^{N} (x_i)^2 - \frac{1}{N}(\sum_{j=1}^{N} x_j)^2}$$

$$b = \frac{1}{N}(\sum_{i=1}^{N} y_i - a \sum_{i=1}^{N} x_i)$$

Use method of least square to estimate the population of 2000

After plug the Census Data of the United States from 1790-1990 into formula, we get: $a = -9.91502657525e - 11$ $b = 0.0295583147615 = r$

And since: $a = \frac{-r}{K}$ then : $K = \frac{-r}{a} = 298116344.289$

So now for the logistic model: $x(t) = \frac{Kx_0}{x_0 + (K - x_0)e^{-rt}}$

We already have the K , r and t, so we could estimate the population of 2000, which 2000 is the 210 years after 1790, so we get:

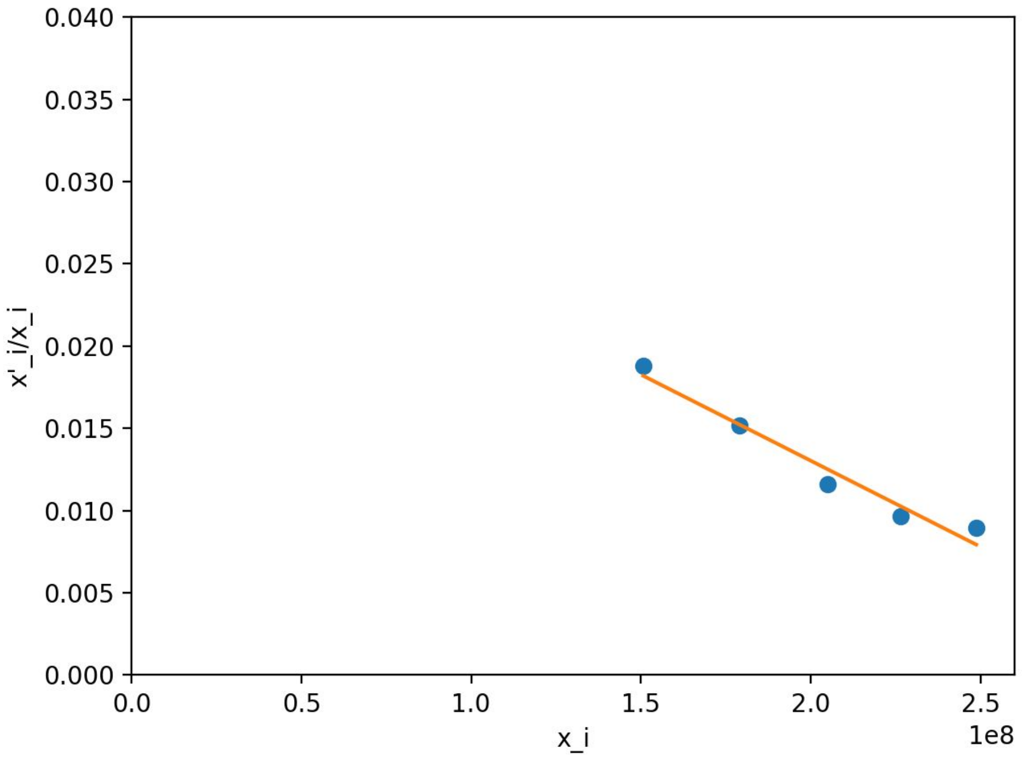
$x(210) = 258782659.69$

1950 and On

- Next we were tasked to estimate the population in the year 2000 using only the census data from 1950 and on
- Our first step was selecting which models to use for estimation
- This involved using different models for different data points, ultimately we decided on this:

$$
\begin{cases}
\frac{x_{i+1}-x_i}{10} & i=0 \\
\frac{x_{i+1}-x_{i-1}}{20} & 1 \leq x \leq 3 \\
\frac{x_i-x_{i-1}}{10} & i=4
\end{cases}
$$

Plot for x’/x against x



Estimating r and K

- Next, we rearranged the logistic equation to get a and b in terms of r and K
- Using the “Help with Least Squares” page provided, we then plug in all of our known data to get a and b
- Once we had a and b, we simply plugged them into these two equations and solved for r and K

$$a = -\frac{r}{K}$$

$$b = r$$

Estimating the population in 2000

- From here, all we had to do was plug in all our calculated values, and set our time value t = 50 years.
- Ultimately we ended up getting 267,922,246.16 as our predicted population
- Obviously we would need round as population cannot be a decimal value
- This was higher than the prediction using all of the data

Potential Reasons for Carrying Capacity Jump

- Smallpox goes away in 1949 meaning more less deaths for small children, increasing the population
- US officially malaria free in 1948, meaning less deaths and therefore increased population
- CDC was formed in 1946, began to work to increase our understanding of disease and how it spreads and how we can control it
- In the 1940s, ~20 percent of Americans moved states or countries, meaning we can better utilize all the resources of our country. Also lowers spread of disease as population became less concentrated

Conclusion

- Overall, we learned a lot from this project
- Real world application really helps understanding the material, compared to random homework problems
- Initially struggled on some of the python work, but with the help of the TA's we pushed through and came out much stronger in this area
- Grew as a team over time, as in the beginning we didn't know each other well and were hesitant to ask questions