PYU33C01 – Assignment 2:
Least Square Fits for Modelling Population Growth

Matthew Conway
Student Number: 22337426

October 15, 2025

1 Introduction

This report presents a numerical analysis of world population growth from 1750 to 2016 using two datasets: Gapminder data (1750–1940) and US Census Bureau data (1950–2016)[1]. The primary objective is to determine whether population growth follows an exponential model and to quantify the growth rate through linear regression analysis of logarithmically transformed data.

1.1 Theoretical Background

If a population grows exponentially, it follows the functional form:

$$n(t) = n_0 \exp[\lambda(t - t_0)] \tag{1}$$

where $n(t)$ is the population at time t, $n_0 = n(t_0)$ is the population at reference time t_0, and λ is the growth constant (per year).

Taking the natural logarithm of both sides yields:

$$\ln n(t) = \ln n_0 + \lambda(t - t_0) \tag{2}$$

By defining $a = \ln n_0 - \lambda t_0$ and $b = \lambda$, this can be rewritten as:

$$\ln n(t) = a + bt \tag{3}$$

This linear relationship between $\ln n(t)$ and t provides a straightforward method to test for exponential growth: if the logarithm of population varies linearly with time, the underlying growth is exponential. The slope b directly gives the growth constant λ, while n_0 can be recovered from the intercept a using:

$$n_0 = \exp(a + \lambda t_0) \tag{4}$$

2 Methodology

Both datasets had population values in millions. The data was imported into the python script using NumPy's `loadtxt` function with appropriate parameters to skip header rows.

The analysis proceeded as follows:

-Calculate the natural logarithm of population values for each dataset
-Plot $\ln n(t)$ versus time to visually locate periods of near linear behaviour
-Select two time ranges demonstrating linear trend:

- Range 1: 1870--1940 (Gapminder data)
- Range 2: 1950--1990 (Census Bureau data)

-Perform least-squares linear regression using scipy.optimize.curve_fit
-Extract fit parameters a and b from each linear fit
-Calculate growth constant λ and initial population parameter n_0 using Equations 3 and 4
-Generate exponential curves using fitted parameters for comparison with original data

The linear regression was achieved using the following function:

```
def linear_func(t, a, b):
    return a + b * t
```

Fitting was completed using:

```
fitparams = optimization.curve_fit(linear_func, year_range,
                                   ln_pop_range)[0]
```

3 Results

3.1 Visual Analysis

Figure 1 shows the graphed population data from both sources. The graph displays monotonic growth throughout the period, with a notable acceleration visible after 1950.

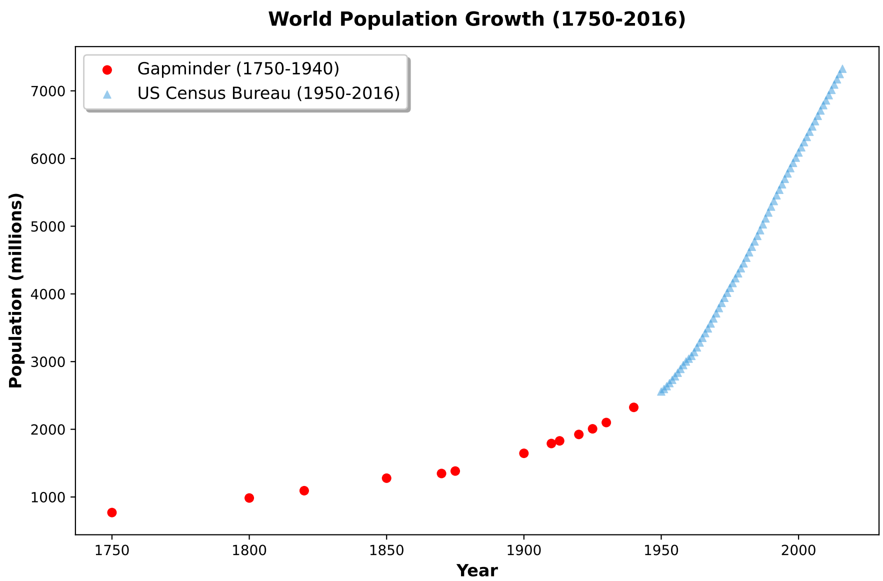

Figure 1: World population growth from 1750 to 2016 using Gapminder (1750–1940) and US Census Bureau (1950–2016) datasets. Population is measured in millions.

Figure 2 displays the natural logarithm of population versus time. The segments of in this graph which appear closest to linear were selected for further analysis under linear regression.

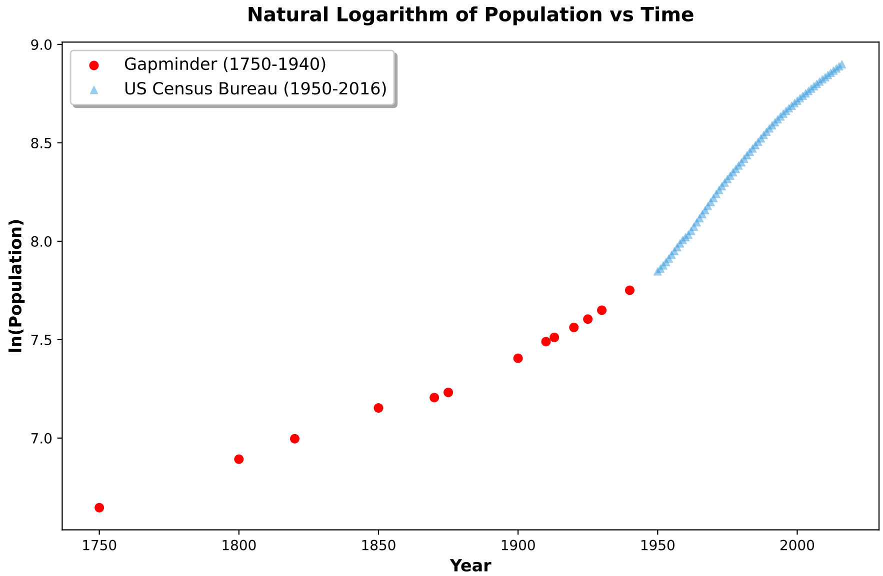

Figure 2: Natural logarithm of population as a function of time. Linear segments suggest periods of exponential growth.

3.2 Linear Regression

Fit 1: Period 1870–1940

For the Gapminder dataset in the range 1870–1940:

$$a_1 = -7.326788 \tag{5}$$
$$b_1 = 0.007616 \text{ year}^{-1} \tag{6}$$

Using $t_0 = 1870$ years, the exponential model parameters are:

$$\lambda_1 = 0.007616 \text{ year}^{-1} \tag{7}$$
$$n_0 = 1241.88 \text{ million} \tag{8}$$

This corresponds to a population growth rate of approximately 0.76% per year during this period.

3.2.1 Fit 2: Period 1950–1990

For the Census Bureau dataset in the range 1950–1990:

$$a_2 = -28.559517 \tag{9}$$
$$b_2 = 0.018583 \text{ year}^{-1} \tag{10}$$

Using $t_0 = 1950$ years, the exponential model parameters are:

$$\lambda_2 = 0.018583 \text{ year}^{-1} \tag{11}$$
$$n_0 = 2518.27 \text{ million} \tag{12}$$

This represents a significantly higher growth rate of approximately 1.86% per year, more than double the rate observed in the earlier period.

3.3 Model Evaluation

Figure 3 shows the original population data, along with the exponential curves created using the fitting parameters discussed in the previous section. This fit shows close agreement with data, and confirms that an exponential model works well for modelling world population growth in this selected time period.

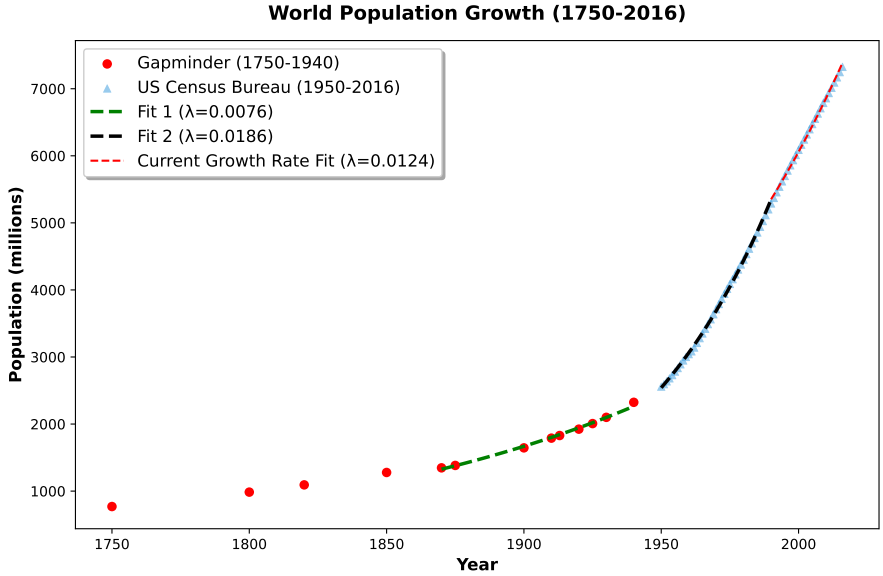

Figure 3: World population data with fitted exponential models. Fit 1 ($\lambda = 0.0076$ year^{-1}) applies to 1870–1940; Fit 2 ($\lambda = 0.0186$ year^{-1}) applies to 1950–1990.

Figure 4 shows the same analysis in logarithmic space, here the change in growth rates is more visible, and appears to occur around the middle of the 20th century. There also appears to have been another change in growth rates around the years 1880 and 1990, with the former change being an increase, and the latter a decrease in growth rates.

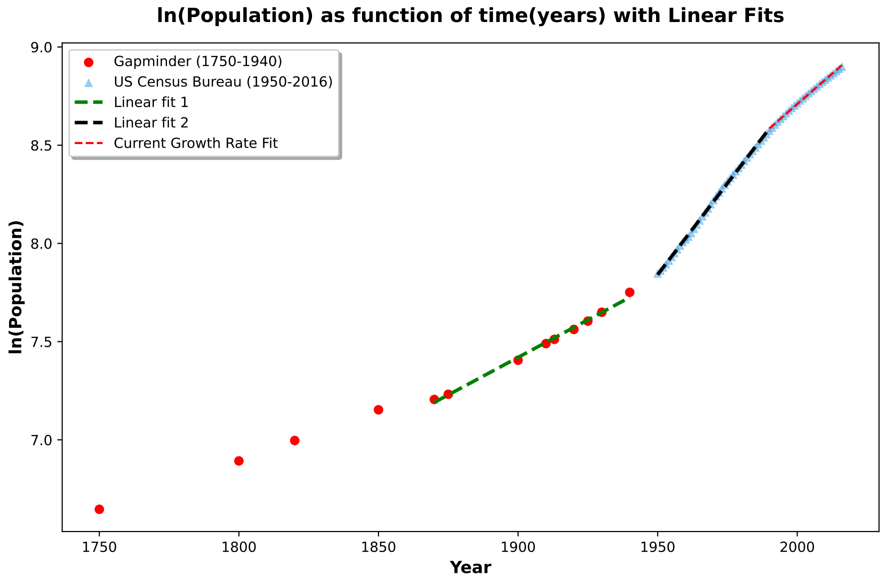

Figure 4: Natural logarithm of population with linear regression fits. The linear relationships confirm exponential growth in the selected time ranges.

4 Discussion

Interpretation of Results

The analysis demonstrates that the growth rates have changed at multiple points in time, one may reflect on historical records of the time in order to assign certain events in the world which were causally related to these changes in growth rate, however this was beyond the scope of this report. The change in growth rates seen around the 1950's(more than doubled from $\lambda_1 = 0.0076$ to $\lambda_2 = 0.0186$) was of particular interest for this reason, I suspect that the end of the second world war may have been related to this change. However, I do not believe my report provides any direct evidence to this statement.

1. **1870–1940**: Growth rate of $\lambda_1 = 0.0076$ year^{-1} (0.76% per year)
2. **1950–1990**: Growth rate of $\lambda_2 = 0.0186$ year^{-1} (1.86% per year)

Current Population Evolution

This analysis has shown that the growth rate is dynamic, and subject to change dramatically in a relatively short period of time. Due to this, it is dangerous to extrapolate based off of the linear regression performed on a period of time that ended 35 years ago(**1950–1990**), especially considering the length of time of said period was only 40 years.

However, after running a third regression, on the period of time starting in **1990**, and ending in **2016**, I found the current approximate growth rate to be $\lambda = 0.012396$ per year, this is roughly 60% of the growth rate seen in the later half of the previous century. Clearly the global population is growing at a slower rate, potentially due to a range of factors from increasing costs of energy to higher costs of childcare.

5 Conclusion

This report outlines the utility of analysing population growth in the log space. Linear regression was applied to global population data over the course of nearly three centuries. This analysis found multiple periods of sustained near exponential growth in the population punctuated by short periods of rapidly changing growth rates. This was most clearly seen in the middle of the 20th century, where the growth rate approximately doubled over roughly a 50 year period, going from $\lambda_1 = 0.0076$ to $\lambda_2 = 0.0186$ respectively.

5.1 AI Usage Declaration

Claude AI was used to enhance the visualization aesthetics of the plots, including color schemes, marker styles, gridlines, and overall figure presentation. Claude was also used to provide initial guidance on the structure of the linear regression implementation using `scipy.optimize.curve_fit`.

References

[1] Wikipedia contributors, "Estimates of historical world population," *Wikipedia, The Free Encyclopedia*, `https://en.wikipedia.org/wiki/Estimates_of_historical_world_population` (accessed October 15, 2025).