Least Square Fits for Modelling Population Growth

Valentina Watanabe
21367661

October 2024

1 Introduction

This coding exercise aims to explore the importance of least square fits models in Python and their applications, in this case using the world population from years 1970 to 2016. Following the conversion of the exponentially rising function to a linear equation in terms of the natural log of the population, a least square fit was performed. To do so, two ranges of data from the logarithmic plot of the data where linear trends could be seen were chosen, 1750-1940 and 1950-2016. With the assistance of one of Pythons' scipy library, we carry out the linear regression to examine the dependency of the logarithm of the population on time. This allows the population growth rate for these periods to be evaluated by finding the required slope and constant values. These values are then used to rewrite and compare the exponentially rising function to the original values.

2 Plot of Population

The first part of the exercise involved performing a numerical analysis of the population density (in millions) from 1970 to 2016. When we plot the data given by Gapminder and the US Census, we find Figure 1. We can see the trend of rapidly increasing exponential growth, with a seemingly almost linear slope from 1970 onwards.

To illustrate the growing trend of the population with time more significantly, we illustrate a figure using the natural log values (Figure 2). The plot shows a shape akin to logistic growth, with a dispersed increase over the years and a sharp logarithmic increase in the last section.

3 Performing linear regression using Python

The formula that models the exponentially rising function with initial population n_0 at time t_0 and growth constant λ is given by,

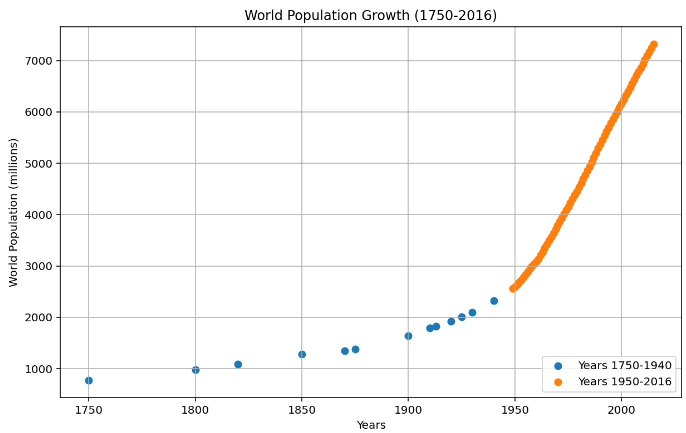

Figure 1: World population along the years

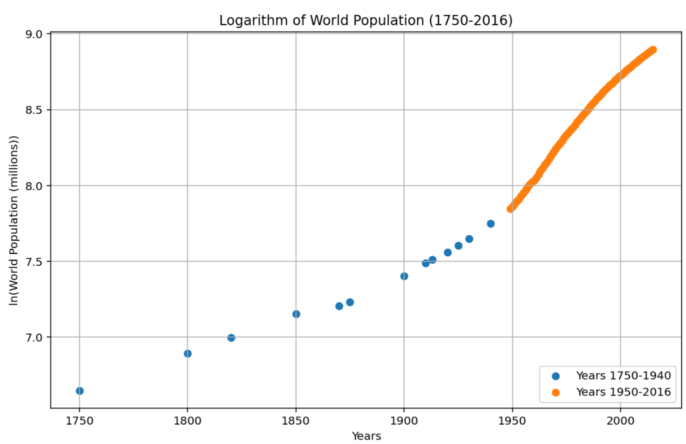

Figure 2: Logarithmic world population along years

$$n(t) = n_0 e^{\lambda(t-t_0)}. \tag{1}$$

Rewriting this in terms of $ln(n_0)$,

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

$$ln\ n(t) = ln(n_0) - \lambda t_0 + \lambda t. \tag{3}$$

Therefore, following linear equation ln $n(t)$ = a + bt, we can assign a = ln n_0 - λt_0 and b = λ.

To apply this to our code and fit a straight line to our log function as a function of time, we find two time ranges where the data can be described with straight lines. Looking at our logarithmic world population along the years plot (Figure 2), we can choose these ranges to be from 1750 to 1940 and 1950 to 2016.

To perform a non-linear least square fit the scipy library is used. The fitting is done by defining the function that outputs a + bx, and then through the parametrization.curve_fit function. Using this, we get the fit parameters. For the first range of data, we get values a_1 = -3.2865280159274923 and b_1 = 0.005647894190320731. For the second range of data, a_2 = -24.42258432287594 and b_2 = 0.016563810710220176.

Because of the given values, we get the first fit equation to be,

$$ln\ n(t) = -3.2865280159274923 + 0.005647894190320731t, \tag{4}$$

with $\lambda = 0.005647894190320731$ and $n_0 = 0.03738341896954663$.

Similarly,

$$ln\ n(t) = -24.42258432287594 + 0.016563810710220176t, \tag{5}$$

with $\lambda = 0.016470474586987263$ and $n_0 =$ 3.0233028695120477e-11.

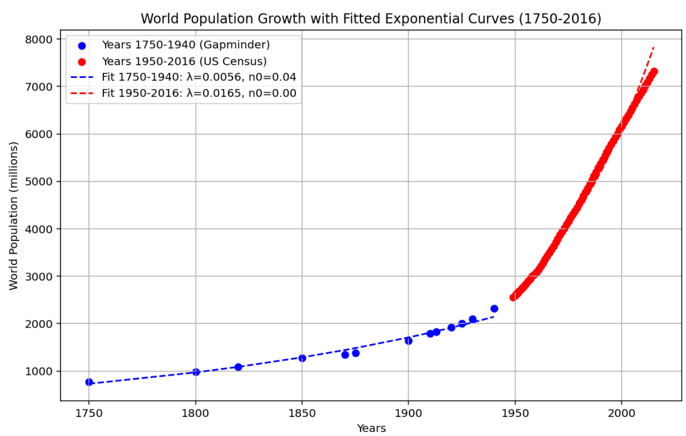

Figure 3: World Population Growth with Fitted Exponential Curves (1750-2016)

From the result of both growth rate values, we can observe its greater number after some time has passed. This can be also seen from the graph, where the second line follows a more accelerated population growth. This is also supported by the fact that in the calculated fitted equations, the slope is greater for the data extracted from more recent years. The current population growth is therefore increasing exponentially at an accelerated rate, which is bound to continue increasing. This could continue to rise unless logarithmic growth is achieved, where we would find the population to continue rising slowly until reaching a certain continuous plateau.

4 Conclusion

The report demonstrates the use of least square fitting in understanding population growth trends. By converting the exponential population model to a linear form and performing linear regression using Python, we identified two distinct growth phases—one from 1750 to 1940 and another from 1950 to 2016. The analysis revealed that the growth rate has significantly increased in recent decades, with the second period exhibiting a steeper slope, indicating an accelerated population growth rate. During the production of the code, technical errors regarding the year arrays caused misalignment with the data when writing the definitions for ranges and performing the fit. The command to load data which was suggested in the manual was attempted but did not work. For the sake of fixing these bugs, ChatGPT was used to write the time ranges correctly in the ydata parameter section of the parametrization.curve_fit definition.