Forecasting Future Population Trends in Egypt using Curve Fitting Models

Seif Raslan, Abdelateef Khaled, Saleh Mohamed

Department of Engineering, University of Science and Technology at Zewail City, Egypt 2023

Contents

1 Abstract

The project "Population Growth Models" aims to develop a numerical analysis program to model and analyze population growth patterns. Using curve-fitting techniques, historical population data is processed and various growth models are fitted to identify the best-fitting model. Predictions and projections for future population trends are made using the selected model. The program focuses on analyzing population dynamics, including growth rate, carrying capacity, and inflection points. Validation is performed by comparing the fitted curves and predictions with actual population data. The project emphasizes visualizing the data, fitted curves, and predicted growth for effective interpretation and communication. The methods used are linear regression, exponential, and quadratic curve fittings techniques. The outcomes offer insights into population dynamics and the potential for further analysis, such as incorporating additional factors and comparing growth models across different regions.

2 Introduction

Growing population is a serious problem that affected us negatively in various fields to the extent that makes it hard for people to live. Governmental institutions have a hard time dealing with such problems, including resource management, housing, healthcare, and education. It is crucial for managers and developers to comprehend population growth patterns and forecast future trends in order to make wise choices and create successful plans for sustainable development. In recent years, the creation of numerical analysis software to simulate and examine population growth trends has grown in significance.

Using curve fitting techniques, the "Population Growth Models" project intends to create a numerical analytic program to model and analyze population growth patterns. The goal of this research is to analyze past population data in order to find the growth model that will best forecast future population trends. Analysis of population dynamics, including growth rate, carrying capacity, and inflection points, is the main goal of the program. For efficient interpretation and communication, the program also emphasizes visualizing the data, fitted curves, and expected growth.

The approaches employed in this research include exponential, linear, and quadratic curve fitting. These methods have been demonstrated to produce precise and trustworthy predictions and are commonly used for examining population growth patterns. The outcomes of the program will be validated by comparing the fitted curves and forecasts with real population data.

The results of this project will provide insightful information about population dynamics and open the door to more investigation, such as comparing growth models across regions and factoring in new variables. The findings of this study will contribute to the field of population studies and can be utilized to guide policy choices on population growth and management.

3 Methodology

3.1 Data collocation

First, we obtained historical population data for Egypt from "The Central Agency for Public Mobilization and Statistics in Egypt"[1]. The data covers the period from 1999 to 2020 as shown in the table below.

Year	Population (in millions)
1999	66
2000	68.5
2001	70.9
2002	73.4
2003	76
2004	78.6
2005	81.1
2006	83.7
2007	86.3
2008	88.9
2009	91.5
2010	94.2
2011	96.9
2012	99.7
2013	102.6
2014	105.5
2015	108.4
2016	111.3
2017	114.3
2018	117.3
2019	120.3
2020	123.4

Table 1: Population of Egypt from 1999 to 2020

3.2 Data Curve Fitting

3.2.1 Linear Data Curve Fitting

The Linear function model is: $y = a_0 + a_1 x$
where the unknown coefficients are obtained by solving:

$$\begin{Bmatrix} a_{\mathbf{0}} \\ a_{\mathbf{2}} \end{Bmatrix} = \left([X]^T[X]\right)^{\mathbf{-1}} \left([X]^T\{Y\}\right)$$

$$[X] = \begin{bmatrix} 1 & 1999 \\ 1 & 2000 \\ 1 & 2001 \\ 1 & 2002 \\ 1 & 2003 \\ 1 & 2004 \\ 1 & 2005 \\ 1 & 2006 \\ 1 & 2007 \\ 1 & 2008 \\ 1 & 2009 \\ 1 & 2010 \\ 1 & 2011 \\ 1 & 2012 \\ 1 & 2013 \\ 1 & 2014 \\ 1 & 2015 \\ 1 & 2016 \\ 1 & 2017 \\ 1 & 2018 \\ 1 & 2019 \\ 1 & 2020 \end{bmatrix} \rightarrow [X]^T[X] = \begin{bmatrix} 22 & 44209 \\ 44209 & 88838871 \end{bmatrix}$$

$$\{Y\} = \left\{ \begin{array}{c} 66 \\ 68.5 \\ 70.9 \\ 73.4 \\ 76 \\ 78.6 \\ 81.1 \\ 83.7 \\ 86.3 \\ 88.9 \\ 91.5 \\ 94.2 \\ 96.9 \\ 99.7 \\ 102.6 \\ 105.5 \\ 108.4 \\ 111.3 \\ 114.3 \\ 117.3 \\ 120.3 \\ 123.4 \end{array} \right\} \to [X]^T\{Y\} = \left\{ \begin{array}{c} 2058.8 \\ 4139570.7 \end{array} \right\}$$

$$\left\{ \begin{array}{c} a_0 \\ a_1 \end{array} \right\} = \left[\begin{array}{cc} 22 & 44209 \\ 44209 & 88838871 \end{array} \right]^{-1} \left\{ \begin{array}{c} 2058.8 \\ 4139570.7 \end{array} \right\} = \left\{ \begin{array}{l} -5380.2916 \\ 2.724 \end{array} \right\}.$$

Therefore, the least-squares linear equation:

$$y = -5380.2916 + 2.724x$$

3.2.2 Quadratic Data Curve Fitting

The second-order polynomial model is: $y = a_0 + a_1x + a_2x^2$
where the unknown coefficients are obtained by solving :

$$\left\{ \begin{array}{c} a_{\mathbf{0}} \\ a_1 \\ a_{\mathbf{2}} \end{array} \right\} = \left([X]^T[X]\right)^{\mathbf{-1}} \left([X]^T\{Y\}\right)$$

$$[X]=\begin{bmatrix}1 & 1999 & 3996001\\1 & 2000 & 4000000\\1 & 2001 & 4004001\\1 & 2002 & 4008004\\1 & 2003 & 4012009\\1 & 2004 & 4016016\\1 & 2005 & 4020025\\1 & 2006 & 4024036\\1 & 2007 & 4028049\\1 & 2008 & 4032064\\1 & 2009 & 4036081\\1 & 2010 & 4040100\\1 & 2011 & 4044121\\1 & 2012 & 4048144\\1 & 2013 & 4052169\\1 & 2014 & 4056196\\1 & 2015 & 4060225\\1 & 2016 & 4064256\\1 & 2017 & 4068289\\1 & 2018 & 4072324\\1 & 2019 & 4076361\\1 & 2020 & 4080400\end{bmatrix}\rightarrow [X]^T[X]=\begin{bmatrix}22 & 44209 & 88838871\\44209 & 88838871 & 178525270099\\88838871 & 178525270099 & 358757257514667\end{bmatrix}$$

$$\{Y\}=\begin{Bmatrix}66\\68.5\\70.9\\73.4\\76\\78.6\\81.1\\83.7\\86.3\\88.9\\91.5\\94.2\\96.9\\99.7\\102.6\\105.5\\108.4\\111.3\\114.3\\117.3\\120.3\\123.4\end{Bmatrix}\rightarrow [X]^T\{Y\}=\begin{Bmatrix}2058.8\\4139570.7\\8323397770.7\end{Bmatrix}$$

$$\begin{Bmatrix}a_0\\a_1\\a_2\end{Bmatrix}=\begin{bmatrix}22 & 44209 & 88838871\\44209 & 88838871 & 178525270099\\88838871 & 178525270099 & 358757257514667\end{bmatrix}^{-1}\begin{Bmatrix}2058.8\\4139570.7\\8323397770.7\end{Bmatrix}=$$
$$\begin{Bmatrix}61226.9644\\63.56903\\0.016495\end{Bmatrix}.$$

Therefore, the least-squares quadratic equation:

$$y=-61226.9644+63.56903x+0.016495x^2$$

3.2.3 Exponential Data Curve Fitting

The Exponential function model is: $y = be^a x$ →lny= ln b + a x
where the unknown coefficients are obtained by solving:

$$\left\{ \begin{array}{c} lnb \\ a \end{array} \right\} = \left([X]^T[X]\right)^{-\mathbf{1}} \left([X]^T\{Y\}\right)$$

$$[X] = \begin{bmatrix} 1 & 1999 \\ 1 & 2000 \\ 1 & 2001 \\ 1 & 2002 \\ 1 & 2003 \\ 1 & 2004 \\ 1 & 2005 \\ 1 & 2006 \\ 1 & 2007 \\ 1 & 2008 \\ 1 & 2009 \\ 1 & 2010 \\ 1 & 2011 \\ 1 & 2012 \\ 1 & 2013 \\ 1 & 2014 \\ 1 & 2015 \\ 1 & 2016 \\ 1 & 2017 \\ 1 & 2018 \\ 1 & 2019 \\ 1 & 2020 \end{bmatrix} \rightarrow [X]^T[X] = \begin{bmatrix} 22 & 44209 \\ 44209 & 88838871 \end{bmatrix}$$

$$\{lnY\} = \left\{ \begin{array}{c} ln66 \\ ln68.5 \\ ln70.9 \\ ln73.4 \\ ln76 \\ ln78.6 \\ ln81.1 \\ ln83.7 \\ ln86.3 \\ ln88.9 \\ ln91.5 \\ ln94.2 \\ ln96.9 \\ ln99.7 \\ ln102.6 \\ ln105.5 \\ ln108.4 \\ ln111.3 \\ ln114.3 \\ ln117.3 \\ ln120.3 \\ ln123.4 \end{array} \right\} \rightarrow [X]^T\{lnY\} = \left\{ \begin{array}{c} 99.4712 \\ 199913.633 \end{array} \right\}$$

$$\left\{ \begin{array}{c} lnb \\ a \end{array} \right\} = \begin{bmatrix} 22 & 44209 \\ 44209 & 88838871 \end{bmatrix}^{-1} \left\{ \begin{array}{c} 99.4712 \\ 199913.633 \end{array} \right\} = \left\{ \begin{array}{l} -54.8681 \\ 0.0296 \end{array} \right\}.$$

Therefore, the least-squares exponential equation:
lny=-54.8681 + 0.0296 x → $y = 1.4827 \times 10^{-24} \cdot e^{0.0296x}$

3.3 Error Calculation

in each of the below the Error calculate as the following equation E= $\left\| \frac{\mathbf{Y}^{(k)} - \mathbf{Y}(\mathbf{X})^{(\mathbf{k})}}{\mathbf{Y}^{(k)}} \right\|_\infty$ 100%

3.3.1 Linear Curve Fitting

$$\left\{ \; E_t \; \right\} = \begin{bmatrix} 1.546 \\ 1.162 \\ 0.666 \\ 0.3381 \\ 0.1634 \\ 0.00021 \\ 0.276 \\ 0.4156 \\ 0.5467 \\ 0.6702 \\ 0.7867 \\ 0.79 \\ 0.7924 \\ 0.6939 \\ 0.5027 \\ 0.3221 \\ 0.1511 \\ 0.011 \\ 0.2521 \\ 0.4809 \\ 0.6984 \\ 0.9856 \end{bmatrix} \rightarrow ||E_t||_\infty = 1.546\%$$

3.3.2 Quadratic Curve Fitting

$$\left\{ \; E_t \; \right\} = \begin{bmatrix} 0.2053 \\ 0.043 \\ 0.08 \\ 0.023 \\ 0.1186 \\ 0.2088 \\ 0.1295 \\ 0.135 \\ 0.102 \\ 0.0337 \\ 0.067 \\ 0.0903 \\ 0.1466 \\ 0.1324 \\ 0.0536 \\ 0.01036 \\ 0.000113 \\ 0.020 \\ 0.02035 \\ 0.0301 \\ 0.012 \\ 0.049 \end{bmatrix} \rightarrow ||E_t||_\infty = 0.2088\%$$

3.3.3 Exponential Curve Fitting

$$\left\{ \ E_t \ \right\} = \begin{bmatrix} 2.168 \\ 1.3915 \\ 0.898 \\ 0.385 \\ 0.1415 \\ 0.549 \\ 0.723 \\ 0.922 \\ 1.0242 \\ 1.037 \\ 0.965 \\ 0.918 \\ 0.79 \\ 0.684 \\ 0.596 \\ 0.4289 \\ 0.1857 \\ 0.1295 \\ 0.426 \\ 0.793 \\ 1.227 \\ 1.644 \end{bmatrix} \rightarrow ||E_t||_\infty = 2.168\%$$

4 Results

Graphs were created for each of the fitted curves to show how well they performed in forecasting future population growth, allowing users to compare the precision of the various curve-fitting approaches employed in the program. The graphs demonstrated that the quadratic fitting methods offered a better match than the linear and exponential methods. Compared to the other models, the quadratic models better represent the population increase pattern and predicted the inflection points and carrying capacity.

We will start off with the results obtained from the linear regression curve, compare the errors and explicitly state the most accurate method. Now the original graph is illustrated on the graphs itself so we can have some sort of sense on where the approximation is to the original curve.

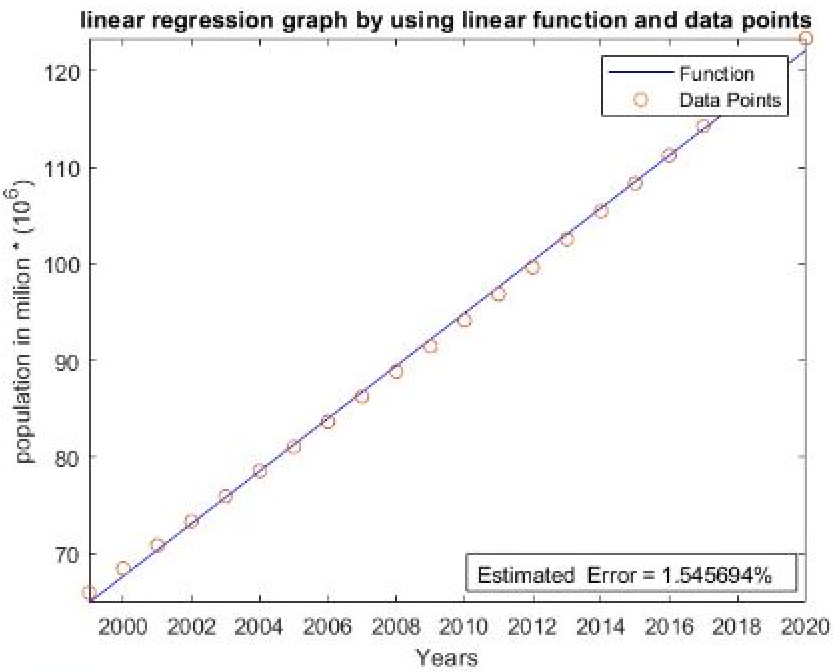

figure 1: linear curve

Given that the original graph looked roughly linear, the error of 1.5457% for the linear regression curve is understandable. However, tiny errors can have a big impact in the long run, highlighting the significance of continuing to improve modeling methodologies. By the previous fact, we can already figure out that the exponential curve won't be as accurate as the linear one as we have little control over the coefficients and it can only go in one direction unlike cubic curves for example. The exponential curve is illustrated below.

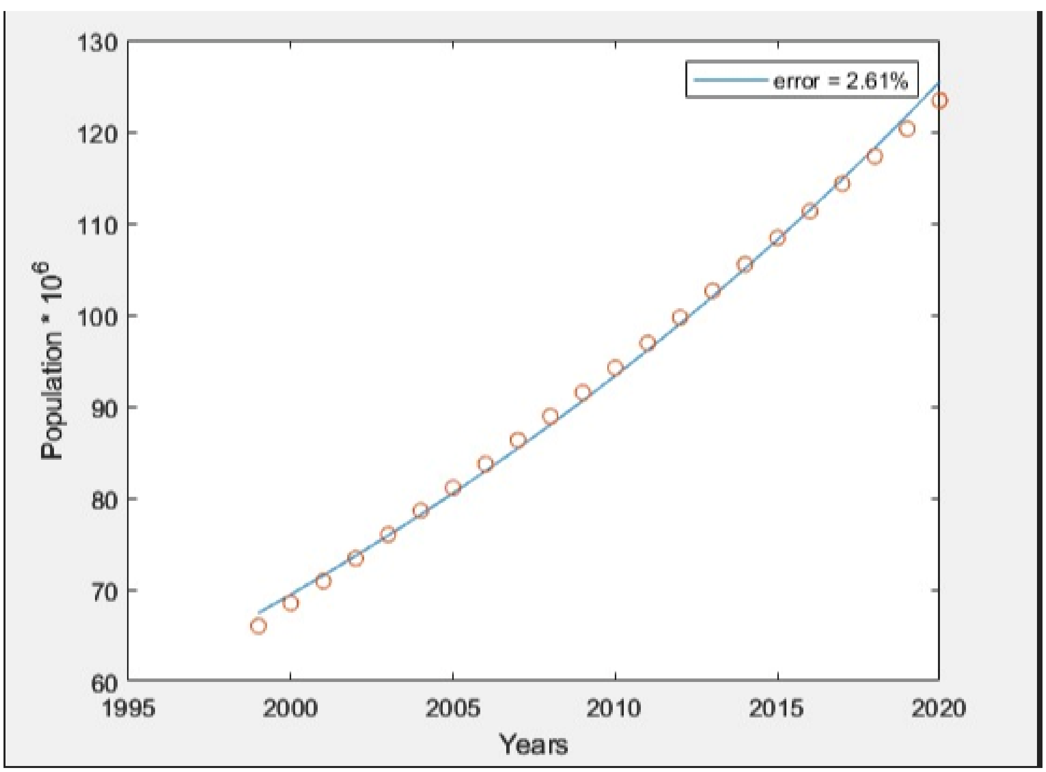

figure 2: exponential curve

As expected, the exponential curve fitted to the population data yielded an error of 2.61%. This result highlights the importance of choosing the appropriate model to achieve accurate predictions, as higher-order equations are not always necessary. In this case, a simpler model, such as a linear curve has provided more accurate results. For the third model which is the quadratic model, the graph can be seen below.

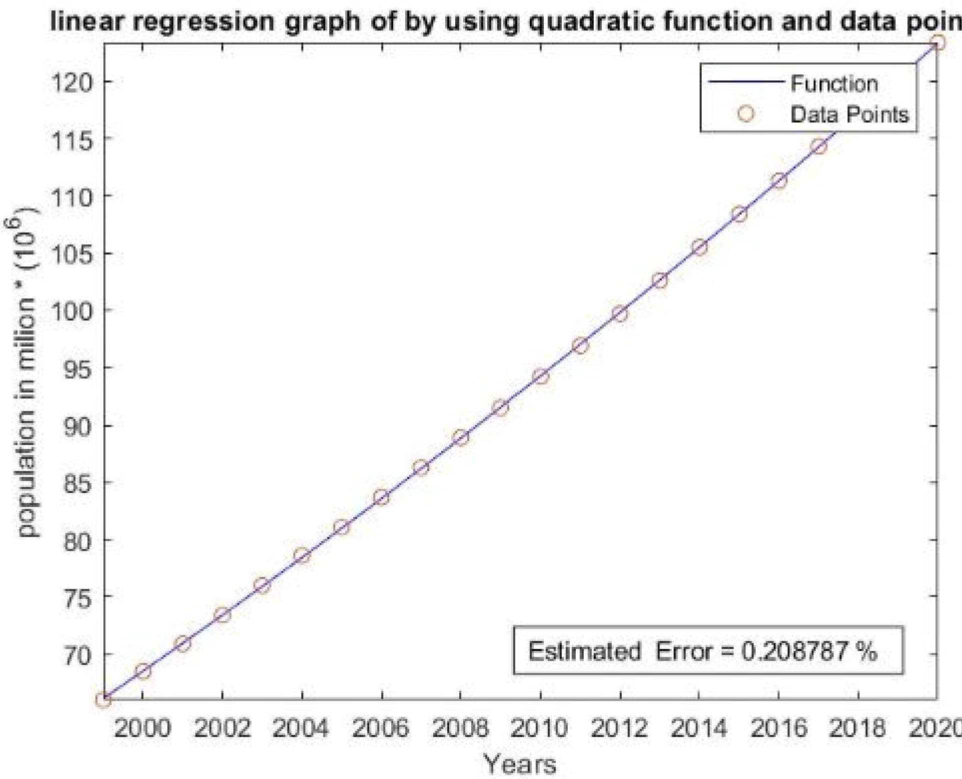

figure 3: quadratic curve

It can be seen that the quadratic model yielded the best result out of them with an error of 0.208787%. Although as stated previously that error may be small but it does affect future calculations.

5 conclusion

In conclusion, the "Population Growth Models" project was successful in creating a numerical analytic program that models and examines population growth trends using curve-fitting methods. The program's ability to correctly forecast future population trends offers insightful information about population dynamics and aids in attempts to promote sustainable development. The findings highlight the significance of choosing the right model for precise forecasts and show how curve-fitting techniques are beneficial for modeling and analyzing population growth patterns.

We have seen through this project that the quadratic model offers the best numerical approximation and can predict near-future population growth in Egypt. Other tools may be used in the same research field such as Newton-Ruphson's method and the splines. Although these are interpolating methods so they won't yield accurate results except for the intervals given or the very near future and this can be done in future research but for this project, we achieved a mathematical equation that predicts Egypt's model very well and can be used to give an indication of how well the population turns out.

6 Appendix

6.1 Linear Data Curve Fitting Matlab code

```

%% code for Linear Data Curve Fitting
clear
clc
%data points
x1 = [1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
     2013 2014 2015 2016 2017 2018 2019 2020 ]; %number of years
y1 = [66 68.5 70.9 73.4 76 78.6 81.1 83.7 86.3 88.9 91.5 94.2 96.9 99.7
    102.6 105.5 108.4 111.3 114.3 117.3 120.3 123.4 ];% population in
    million

%% The Algoritms
X=[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ; 1999 2000 2001 2002 2003
    2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
    2018 2019 2020];
Y=y1;
Xnew = inv((X*transpose(X)));
ynew= X*transpose(Y);
coff = Xnew * ynew;
f=@(x) coff(1) + coff(2)*x; % the function

%% Error calculation
for i=1:size(y1,2)
    eror(i)=abs(y1(i)-f(x1(i)))/y1(i);
end
fprintf('The error is %f %% \n ',norm(eror,Inf)*100)

%% Graphical representation
% Plot the function
fplot(@(x) coff(1) + coff(2)*x,[1999 2020],'r')
hold on % Hold the graph to add scatter plot
% Add the scatter plot
scatter(x1, y1)
% Add axis labels and legend
xlabel('Years')
ylabel('population in milion * (10^6)')
legend('Function', 'Data Points')

% Add title
title('linear regression graph by using linear function and data points')
```

6.2 Quadratic Data Curve Fitting Matlab code

```

%% code for Quadratic Data Curve Fitting
clear
clc
%data points
x1 = [1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
     2013 2014 2015 2016 2017 2018 2019 2020 ]; %number of years
y1 = [66 68.5 70.9 73.4 76 78.6 81.1 83.7 86.3 88.9 91.5 94.2 96.9 99.7
    102.6 105.5 108.4 111.3 114.3 117.3 120.3 123.4 ];% population in
    million

%% The Algoritms
X=[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ; 1999 2000 2001 2002 2003
    2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
    2018 2019 2020;1999^2 2000^2 2001^2 2002^2 2003^2 2004^2 2005^2 2006^2
    2007^2 2008^2 2009^2 2010^2 2011^2 2012^2 2013^2 2014^2 2015^2 2016^2
    2017^2 2018^2 2019^2 2020^2];
Y=y1;
Xnew = inv((X*transpose(X)));
ynew= X*transpose(Y);
coff = Xnew * ynew;
f=@(x) coff(1) + coff(2)*x + coff(3)*x^2; % the function

```

```
%% Error calculation
for i=1:size(y1,2)
    eror(i)=abs(y1(i)-f(x1(i)))/y1(i);
end
fprintf('The error is %f %% \n ',norm(eror,Inf)*100)

%% Graphical representation
% Plot the function
fplot(@(x) coff(1) + coff(2)*x + coff(3)*x^2,[1999 2020],'b')
hold on % Hold the graph to add scatter plot
% Add the scatter plot
scatter(x1, y1)
% Add axis labels and legend
xlabel('Years')
ylabel('population in milion * (10^6)')
legend('Function', 'Data Points')

% Add title
title('linear regression graph of by using quadratic function and data
    points')
```

6.3 Exponential Data Curve Fitting Matlab code

```
format long g
clear
clc
data = [1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
    2012 2013 2014 2015 2016 2017 2018 2019 2020 ];
x1 = [1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
     2013 2014 2015 2016 2017 2018 2019 2020 ];
y1 = [66 68.5 70.9 73.4 76 78.6 81.1 83.7 86.3 88.9 91.5 94.2 96.9 99.7
    102.6 105.5 108.4 111.3 114.3 117.3 120.3 123.4 ];
y2=log(y1);
x1=reshape(x1,22,1);
y2=reshape(y2,22,1);
x2=1;
for c=1:21
    x2(end+1)=1;
end
x2=reshape(x2,22,1);
x1=[x2 x1];
m1=transpose(x1)*x1;
m2=transpose(x1)*y2;
m3=inv(m1)*m2;

a=exp(m3(1));

b=m3(2);
syms y(r)
y(r)=a*exp(b*r);
x3=1999:1:2020;
plot(x3,y(x3))
hold on
scatter(data,y1)
xlabel("Years")
ylabel("Population * 10^6")
legend("error = 2.61%")
numerical_year=vpa(y(x3));
numerical_year=reshape(numerical_year,22,1);
y1=transpose(y1);
error=(numerical_year-y1)./y1;
error=transpose(error);
real_error=norm(error,Inf)*100;
```

7 References

1. . (n.d.-b). https://www.capmas.gov.eg/HomePage.aspx
2. Burden, R. L., Faires, J. D., Burden, A. M. (2016). Numerical analysis. Cengage Learning