Mathematical Model to Extrapolate the Population of Ghana: An Application of Newton's Divided Difference Formula

Richmond Opoku-Sarkodie, E. Acheampong, Patrick Enu, Amadu Bokoum, F. A. Gardiner, T. A. Manu
Methodist University College Ghana
patrickenu@gmail.com

Date Received: January 22, 2015; Date Revised: February 19, 2015

Abstract *- This work presents the derivation of a mathematical model for extrapolating the population of Ghana using Newton's divided difference formula. Newton's divided difference formula was used because of the uneven time intervals at which the population census was conducted. A simulation was done using mathematical software to estimate the parameters in the derived model because the exact solution was quite difficult. The model was able to predict the population of Ghana with a residual percentage error within 10%. We will therefore recommend the usage of this model to be used for countries in which the population censuses were conducted at uneven time intervals as was the case with Ghana.*

Keywords: *extrapolation, population, Newton's, divided difference formula, Ghana*

INTRODUCTION

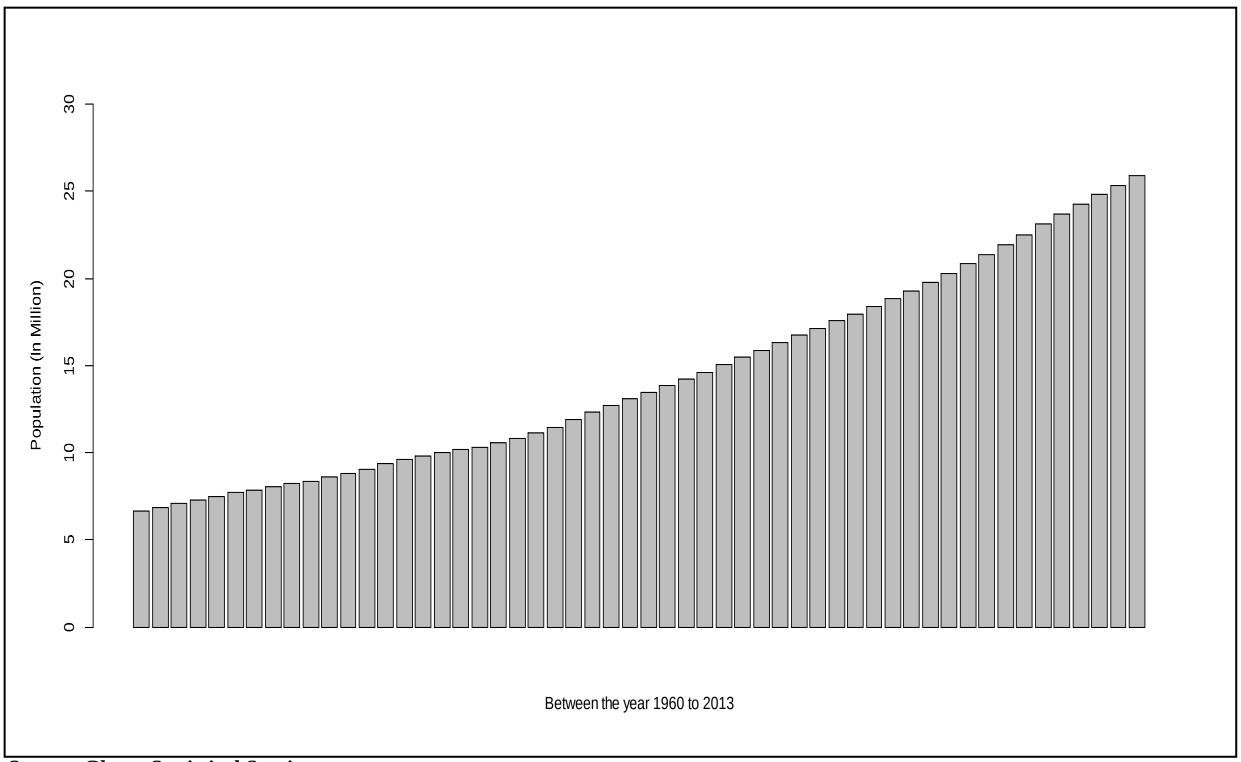
Ghana has had uneven population census over the past few years until 2010. This has made socio-economic and demographic planning very difficult for the country – Ghana, more especially activities in the services sector (electricity, water, housing etc). The difficulty in socio-economic planning and demographic planning are due to the incorrect predictability of the population size within the decennial periods based on exponential growth model. This study, therefore, seeks to develop a population model for Ghana using Newton's Divided Difference Formula which will be used to predict Ghana's population figures with a greater amount of accuracy.

One of the major problems facing the world today is excessive population growth in relation to the development of the natural resources of the world. The composition and growth of the population of any country have a direct bearing upon its socio-economic and demographic developments. The most important fact underlying the current population debate in developing countries and for that matter Ghana is that while population is increasing to an uncontrollable size, economic resources are very limited. It is therefore important to study population issues to be able to control population growth so that the desired socio-economic and demographic developments can be realized.

Most countries including Ghana conduct population census periodically to ascertain the population growth of their country at a particular point in time. This is to aid research, business marketing, and national economic and financial planning. The first census in Ghana was conducted in 1891 by the then British Administration. Since then population censuses were conducted every ten years in Ghana. Ghana has experienced population census in 1891, 1901, 1911, 1921 and 1931 when the second World War discontinued the series and as a result there was no population census conducted in the year 1941 (Ghana Statistical Service 2010 Population and Housing Census National Analytical Report, 2013). After the Second World War a population census was conducted in 1948, the last census organized by the then British Administration. These earlier population censuses were conducted in the same years as censuses in the United Kingdom.

After Ghana's independence in 1957, the country adopted the United Nation's recommendation to conduct population censuses in years ending with zero or closer to zero. So, the first post-independence population census was conducted in 1960 and the second in 1970. Between these periods the population of Ghana grew from 6.7 million in 1960 to 8.6 million in 1970. This gave an inter-censual growth rate of 2.4% per annum. There was no population census in

1980 due to political instability at that time. This broke the decennial census conducted in Ghana. Suddenly, a population census was conducted in 1984 and then 2000, breaking the expected decennial census taking in Ghana and presenting unusual intervals between the censuses- 14 years between 1970 and 1984 and 16 years between 1984 and 2000. Between 1970 and 1984, the population of Ghana increased by over 40% to 12.3 million, giving a growth rate of 2.6% between the two censuses. Ghana's population was estimated at around 15.3 million with an annual growth rate of 3.1% between 1984 and 1992. According to the population census of 2000, Ghana's population is 18.9 million, which gives a growth rate of about 3.2% per annum. The 2010 census has restored the decennial process of population census taking in Ghana. However, these uneven population census patterns experienced in the past years have affected the predictability of the population size within the decennial periods and beyond (Ofori et al., 2013). This is because the Ghana Statistical Service uses the exponential functions to predict the population of Ghana; however, we think that this is not a good model because of the uneven time intervals. The only way out is to use the Newton's Divided Difference Formula (which takes into account the uneven time interval at which the census was conducted) to predict the population growth of Ghana correctly, hence, this important research paper.



Source: Ghana Statistical Service

Fig. 1: Population of Ghana, 1921 – 2010

In addition, there are two main mathematical models of population dynamics of human population. These models are Malthusian growth model (exponential model) and logistic growth model. The exponential growth model was proposed by Reverend Thomas Robert Malthus (1798). His main argument was that population was growing exponentially while food supply was growing linearly. He believed that population would go on increasing more than food supply until it was checked by famine, disease, pestilence or war. The second type of mathematical models of population dynamics we can talk about is Logistics (Verhulst) Growth Model. This model was proposed by Belgian Scientist called Pierre Francois Verhulst in 1838. He believed that population growth depends not only on population size, but also on the effect of a "carrying capacity" that would limit growth

(Gbogbo, 2011). These models have been empirical used by some researchers.

Ofori et al. (2013) used mathematical models to predict the population of Ghana. They used the Exponential and Logistic growth models to model the population growth of Ghana using data from 1960 to 2011. The exponential model predicted a growth rate of 3.15% per annum and also predicted the population to be 114.8207 million in 2050. The Logistic model predicted a growth rate of 5.23% and predicted Ghana's population to be 341.2443 million in 2050.

The Ghana Statistical Service of Ghana, together with the work of Ofori et al. (2013) uses the exponential growth model to predict the population of Ghana. However, we think that this methodology is incorrect since Ghana has had uneven past population census figures for prediction. The reason is that the exponential growth function does not take into account the uniform time intervals of data being used. This may to a large extent affect the implementation of the exponential growth function to extrapolate the population of Ghana. As a result, the Newton's Divided Difference Formula will be applied in this study to predict the population of Ghana and we think that this will give a better result as compared to the exponential growth function.

MATERIALS AND METHODS

If a table does not have equally spaced values of x, we introduce divided difference. Let's assume that the values of x are $x_0, x_1, x_2, x_3, \ldots$ and that the function is $f(x)$, we define successive divided difference by:

$$f(x_0, x_1) = \Delta\mathbf{f} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$f(x_0, x_1, x_2) = \Delta^2 f = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

$$f(x_0, x_1, x_2, x_3) = \Delta^3 f = \frac{f(x_1, x_2, x_3) - f(x_0, x_1, x_2)}{x_3 - x_0}$$

These are known as divided difference of order 1, 2 and 3 respectively.

We can represent these differences in a table form as below:

Table 1. Table of divided difference

x	$f(x)$	Δf	$\Delta^2 f$	$\Delta^3 f$
x_0	$f(x_0)$			
		$f(x_0, x_1)$		
x_1	$f(x_1)$		$f(x_0, x_1, x_2)$	
		$f(x_1, x_2)$		$f(x_0, x_1, x_2, x_3)$
x_2	$f(x_2)$		$f(x_1, x_2, x_3)$	
		$f(x_2, x_3)$		
x_3	$f(x_3)$			

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com

Table 1.1 Table of Divided Difference using the census values from 1960 to 2000

Time(x)	x	**Population** $f(x)$	**First Divided Diference** Δ**f**	**Second Divided Difference** $\Delta^2 f$	**Third Divided Difference** $\Delta^3 f$
1960 $(= x_0)$	0	**6652516**			
			194461.7		
1970 $(= x_1)$	10	**8597133**		**73420.63**	
			928668		**50856.6**
1984 $(= x_2)$	14	**12311805**		**581986.6**	
			3256614.5		
2000 $(= x_3)$	16	**18825034**			

Source: Authors computation, 2014.

In modeling the population growth, the common yardstick is the population growth per unit population. This quantity is expressed mathematically as:

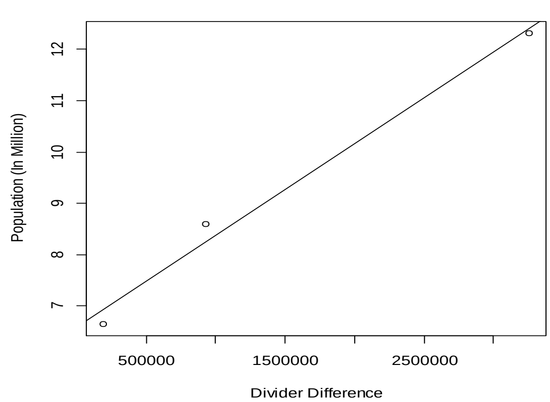
$$\left(\frac{1}{P}\right)\left(\frac{dP}{dt}\right).$$

Where P is the population and $\frac{dP}{dt}$ is the change in population with respects to time.

This $\left(\frac{1}{P}\right)\left(\frac{dP}{dt}\right)$ quantity approximate numerically using divided difference because the intervals in 1960, 1970, 1984 and 2000 are not evenly spaced.

The graph of the population against the divided difference using excel is shown in the figure 2 below.

Fig 2 Graph of population against divided difference



From the graph we can infer that the quantity $\left(\frac{1}{P}\right)\left(\frac{dP}{dt}\right) \approx qP + d$, where q and d are constants. This is so because of the linear relationship between the population and the divided difference. This assertion leads us to the equation

$$\left(\frac{1}{P}\right)\left(\frac{dP}{dt}\right) \approx qP + d \rightarrow \frac{dP}{dt} = qP^2 + Pd$$

Integrating both sides of the equation with respect to t, we obtained

$$\int \frac{dp}{qP^2 + Pd} = \int \frac{(qP^2 + Pd)}{qP^2 + Pd} dt$$

$$\int \frac{dp}{P(qP + d)} = \int dt$$

Now, expressing $\frac{1}{P(qP+d)}$ as a partial fraction

$$\frac{1}{P(qP + d)} = \frac{A}{P} + \frac{B}{Pq + d}, \text{A, B are constants}$$

$$\frac{1}{P(qP + d)} = \frac{A(Pq + d)}{P(Pq + d)} + \frac{BP}{P(Pq + d)}$$

$$\frac{1}{P(qP + d)} = \frac{A(Pq + d) + BP}{P(Pq + d)}$$

$$1 = APq + Ad + BP$$

$$1 = Ad + P(B + Aq)$$

Comparing coefficients gives

$$1 = Ad \text{ and } 0 = APq + BP$$

$$1 = Ad \rightarrow A = \frac{1}{d}$$

$$0 = APq + BP \rightarrow \frac{1}{d}Pq + BP = 0 \rightarrow B = -\frac{Pq}{dP} \rightarrow B = -\frac{q}{d}$$

$$\frac{1}{P(qP + d)} = \frac{A}{P} + \frac{B}{Pq + d} = \frac{1}{dP} - \frac{q}{d(Pq + d)}$$

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com

Hence $\int \frac{dP}{P(qP+d)}$ becomes

$$\int \frac{dP}{P(qP+d)} = \int (\frac{1}{dP} - \frac{q}{d(Pq+d)})dP = \int dt$$

$$\int (\frac{1}{dP} - \frac{q}{d(Pq+d)})dP = \int \left(\frac{1}{dP}\right) dP - \int (\frac{q}{d(Pq+d)})dP = \int dt$$

$$\frac{1}{\mathrm{d}} \ln P - \frac{I}{d} \ln(Pq+d) = t + c$$

$$\ln P - \ln(Pq+d) = dt + dc$$

$$\ln(\frac{P}{Pq+d}) = dt + dc$$

$$\frac{P}{Pq+d} = e^{dt+dc} = e^{dt} e^{dc}$$

$$\frac{P}{Pq+d} = Ae^{dt} \text{ Where } e^{dc} = A$$

$$P = Ae^{dt}Pq + dAe^{dt} \rightarrow P - Ae^{dt}Pq = dAe^{dt} \rightarrow P\left(1 - qAe^{dt}\right) = dAe^{dt}$$

$$P = \frac{dAe^{dt}}{\left(1 - qAe^{dt}\right)} \rightarrow P = \frac{dAe^{dt}/e^{dt}}{\frac{1}{e^{dt}} - \frac{qAe^{dt}}{e^{dt}}} = \frac{dA}{e^{-dt} - qA} = \frac{G}{e^{-dt} - T} \text{ Where } G = dA, T = qA$$

Hence

$$P = \frac{G}{e^{-dt} - T} \quad \text{Model 1}$$

Where d and, G and T are constants.

Fitting model 1 to population data

To fit model (1) to the Ghanaian population data, we need start values for the parameters. It is often important in nonlinear least-squares estimation to choose reasonable start values, and this generally requires some insight into the structure of the model. We know that G represents asymptotic population. The data in Figure 1 show that in 2010 the Ghanaian population stood at about 26 million and did not appear to be close to an asymptote; so as not to extrapolate too far beyond the data, let us set the start value of G to 30.

It is convenient to scale time so that $x_1 = 0$ in 1960, and so that the unit of time is 1 year. Then substituting $G = 30$ and $x_1 = 0$ into the model, using the value $y_1 = 6.652516$ from the data, and assuming that the error is 0, we have

$$6.652516 = \frac{30}{e^{-d*0} - T}.$$

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com

Solving for β_3 gives us a plausible start value for this parameter:

$$6.652516 - 6.652516 * T = 30,$$

$$T = \frac{6.652516 - 30}{6.652516} \approx -3.5095.$$

Finally, returning to the data, at time $x_2 = 1$ (i.e., at the second Census, in 1961), population was $y_2 = 6.866780$. Using this value, along with the previously determined start values for G and T, and again setting the error to 0, we have

$$6.866780 = \frac{30}{e^{-d} - 3.5095}.$$

Solving for d,

$$6.866780e^{-d} = 30 - 6.866780 * 3.5095,$$

$$e^{-d} = 0.8,$$

$$d = 0.15157.$$

The exact solution of the model derived above is quite difficult and so we use the R software to do a model simulation. The results of the simulation are presented in figure 3 below.

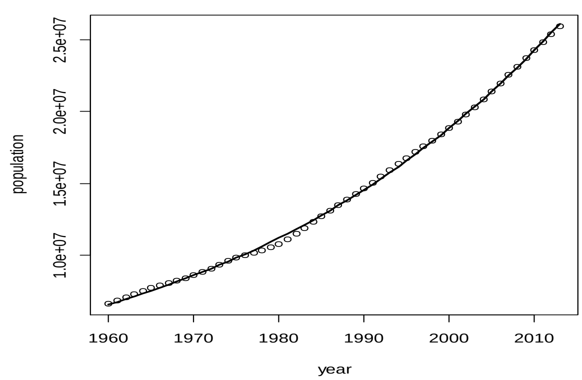


Fig. 3: simulation using mathematical software

The broken lines represents the actual data represented in Figure 3 above and the curve represents the results of the simulation using mathemtical software. Figure 4 also show the graph of the residuals.

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com

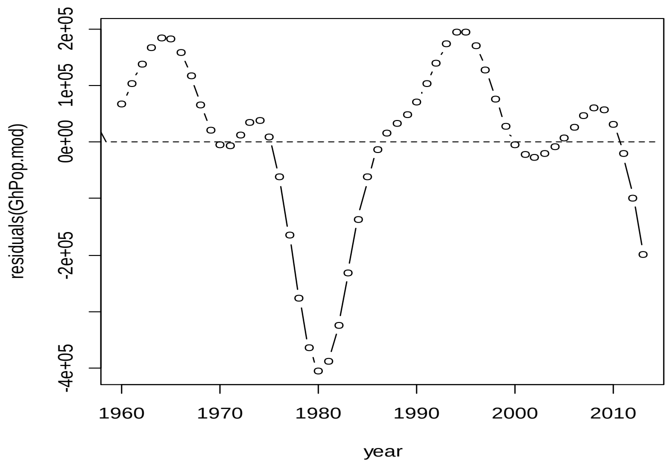


Fig 4: Graph of the Residuals

The estimation of the parameters in the model above has been summarized below.

Table 1.5: The Estimation of the Parameters in the Model

Parameters	Estimate	Standard Error	t-value	Pr(>|t|)
G	6.759e+06	3.738e+04	180.836	<2e-16 ***
d	2.754e-02	6.002e-04	45.879	<2e-16 ***
T	-2.657e-02	8.703e-03	-3.053	0.0036 **

Significance . codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 151800 on 51 degrees of freedom
Number of iterations to convergence: 12
Achieved convergence tolerance: 9.075e-07

The final model becomes

$$P(t) = \frac{G}{e^{-dt} - T} = \frac{6.759 \times 10^{6}}{e^{-2.754\times 10^{-02t}} + 2.657 \times 10^{-2}}$$

Predictions using the model

Year	Actual Population Values of Ghana	Predicted Values
1960	6652516	6584061
1961	6866780	6763019
1962	7085698	6946704
1963	7303664	7135234
1964	7513510	7328729
1965	7710761	7527312
1966	7891194	7731108
1967	8057629	7940243
1968	8221194	8154849
1969	8397508	8375058
1970	8597133	8601005
1971	8827429	8832830
1972	9083737	9070672
1973	9350286	9314676
1974	9604475	9564988
1975	9831636	9821756
1976	10023738	10085134
1977	10190202	10355275
1978	10354855	10632337

1979	10551189	10916481
1980	10802497	11207869
1981	11118132	11506667
1982	11488683	11813045
1983	11895742	12127172
1984	12311805	12449224
1985	12716887	12779378
1986	13104616	13117812
1987	13480997	13464709
1988	13853171	13820255
1989	14233006	14184636
1990	14628693	14558043
1991	15043053	14940669
1992	15471695	15332708
1993	15907265	15734358
1994	16339278	16145819
1995	16760926	16567293
1996	17169151	16998984
1997	17568461	17441098
1998	17968830	17893845
1999	18384302	18357433
2000	18825034	18832076
2001	19293392	19317987
2002	19786307	19815380
2003	20301686	20324473
2004	20835514	20845483
2005	21384034	21378628
2006	21947779	21924130
2007	22525659	22482208
2008	23110139	23053083
2009	23691533	23636976
2010	24262901	24234108
2011	24820706	24844702
2012	25366462	25468978
2013	25904598	26107155
2014		26759455
2015		27426094

Source: Authors Computation 2014

RESULTS AND DISCUSSION

While Figure 3 confirms that the model (1) mean function generally matches the data, the residual plot in Figure 4 suggests that there are systematic features that are missed, reflecting differences in growth rates, perhaps due to factors such as changes in immigration, migration, death rate and birth rate.

To see the effective of this model for predicting the population of Ghana, we looked at a few predictions as compared to the actual. In 1960, the population census estimated the population to be 6,652,516 and the model predicted a population of to be 6,584,061. In 1961 and 1962, the estimated populations were 6,866,780 and 7,085,698 . The population for 1961 and 1962 as predicted by the model are respectively 6,763,019 and 6,946,704. The deviation of the predicted values from the model are very close from the actual values are not so large.

The level of significance is very good having a three star *** indicating that the model is good in predicting the population of Ghana.

CONCLUSION AND RECOMMENDATION

Having a good model to estimate the population of a country is very important. Such models will help in better planning and also help in managing the scare resources for development. The Newton's divided difference formulation proved useful in deriving the model due to the inconsistent manner the population census was conducted over uneven time intervals in Ghana.

We will therefore recommend the usage of this model to be used for countries in which the population censuses were conducted at uneven time intervals as was the case with Ghana.

REFERENCES

Bureau of Statistics (1964). 1960 Population Census of Ghana, vol. XXX. Ghana Bureau of Statistics, Accra.

Engmann, E.V.T (1985). Population of Ghana: 1860 – 1960. Ghana Universities Press, Accra.

Ghana Statistical Service (2013). 2010 Population and Housing Census, National AnalyticalReport. Ghana Statistical Service, Accra.

Gbogbo, D. Y. (2011). Evaluation of Population Census and Population Growth in Ghana through Demographic analsysis. Kwame Nkrumah University of Science and Technology, Ghana

Ofori, T., Ephraim, L., and Nyarko, F. (2013). Mathematical Model of Ghana's Population Growth. International Journal of Modern Management Sciences, 2(2): 57 – 66.

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com