A Comparative Study of Mathematical and Statistical Models for Population Projection of Nigeria

Kerry C. C, Subeno Tessy, Dr J.N. Ezeora, Okafor Joseph Iweanandu Department of Mathematics and Statistics, Delta State Polytechnic, Ogwashi-Uku Department of Mathematics and computer Science Ebonyi State University, Abakaliki

> PH: 08037743555, Email: kerryedu1@gmail.com Email: jerryezeora@yahoo.com PH: 08032275807, Email: josephiokafor@yahoo.com

Abstract: For a country to develop socially and economically, it needs to plan its activities not only for the present, but also for the future. Population projection provides future estimates of population sizes needed by planners. In Nigeria a number of factors have contributed negatively to having an accurate and reliable census data. Successive Nigerian Governments have embarked on Population Census in which huge sums of money was expended. Population projection can be made by examining the influence of demographic components, such as birth, death and migration of population as it affects population from time to time so as to determine future population. In this work we compare the exponential growth model and a statistical model. Using performance measures we determine which is more accurate in forecasting the future population of Nigeria projected to 2036.

---- -

INTRODUCTION

In the last census exercise conducted by the National Population Commission in 2006, Nigeria's population was estimated to be at nearly one hundred and forty million. This is indeed an alarming figure for a country with distressed economy. The consequences of overpopulation is well known, with characteristics of Socio-economic problems such as unemployment or UNDEREMPLOYMENT, LOW LEVEL OF PER CAPITAL INCOME, LOW STANDARD OF LIVING, POVERTY, HUGE EXTERNAL debt burden and many more social vices become manifested.

An important area where data collection is essential is human population. Its data collection procedure is usually by direct enumeration of every individual in the territory at approximately the same time. Collecting data on human population involves a complex series of related activities. There is no doubt that accurate population statistics is vital to sound development planning and economic management. Apart from information on the stock of the country's population, it is essential to know the rate at which the population is changing, structurally and in the aggregate.

Factors affecting population growth and decline

Health Provision

_ _ _ _ _ _ _ _

Provision of good health services and facilities can clearly be a major cause of high birth rate. Adequate and affordable health services enhance the chances of safe delivery, reduce miscarriages, still births and also reduce infant mortality.

According to Jamison (2003), improvement of life expectancy especially in developing countries like Nigeria and the sharp drop in child mortality can be attributed to the huge investment in the health sector by most governments, especially in the developing countries. Further causes of high birth rates are high birth fertility and low mortality. According to Fapohunda (1979), the size of any population is correlated with the level of fertility and infant mortality. It is positively correlated in the first case and negatively correlated with the later. She argues that the Nigerian population is growing at a high rate of 2.5% per year due to a high fertility rate and a decreasing mortality rate, particularly among infants. Fapohunda further argued that since the population is growing rapidly due to a high fertility rate and a declining mortality rate, an increasing proportion of the population is under 15years of age and is for the most economically inactive. They are basically dependent on the economically active group.

High Fertility

It is discovered Fapohunda (1979), that as long as Nigerians' fertility remains high, the prospect is that the proportion of the population who are children will remain high and finally she maintained that the human resources of a country are similar to the capital resources in that not only can the stock increase through rapid birth rates, but investments can also improve its production capacities. Also urban unemployment exists in Nigeria because the growth of new job seekers is increasing at a faster pace than the growth of employment opportunities. The genesis of this social problem is traced from high birth, low infant mortality and consequently high turnout of graduates of diversified disciplines who could not be accommodated in the employment market. Becker (1975), added to this, by conceptualizing economic growth in terms of human capital as a strategic input. Thingani (1995), conceived economic growth as a generalized process of capital accumulation in which human capital has a complimentary limits with other forms of capital. From this definition, it becomes clear that the role of human effort in economic development cannot be overlooked. It therefore implies that an increase in population matched with necessary educational and training facilities is capable of providing a nation with skilled labour force whose contributions to economic advancement cannot be over emphasized.

Mortality

Death on the other hand, is caused basically by several diseases which are environmental oriented. Poor environmental sanitations as evident in poor sewage disposal system, inadequate supply of

clean drinkable water are major causes of some known diseases like Typhoid, Malaria, Cholera and several others. Malaria is one of the commonest diseases that claims many lives in tropical countries like Nigeria. According to Hanson (2004), malaria strikes an estimated 100 million or more people in the developing countries each year, killing one to two million of them. It is one of the biggest killers of children in sub-Sahara Africa. The economic cost of malaria is also stressed by Anne Mills. According to Ann (2008), the impact of the disease on the economy and economic development efforts as seen in reduced productivity caused by workers illness and its financial burden on individuals and government is evident.

Due to its widespread nature, precise estimates of the mortality and morbidity of malaria are often hard to come by.

Migration

Migration in any population is as old as humankind itself. However, migration over the years has resulted in a whole new set of socio-economic, political and demographic circumstances. As of 2010, the total number of internal migrants in Nigeria was estimated at 11,257 (NPC, 2010). There has been an age-long belief that migration is both sex and age selective. Though, the study undertaken by the National Population Commission (NPC) was not meant to prove or disprove this myth, the result was all the same striking, as there were more female than male migrants; 51.5 percent compared to 48.5 percent with variations from state to state.

Statement of the problem

In spite of the unenviable history of census in Nigeria, there has always been a mounting concern that the country must have accurate, reliable and acceptable demographics data.

Presently, the national population commission is the constituted body to provide this data. Moreover, the body is faced with the problem of organization and co-ordination, though, it has not relented much in its efforts to achieve the goals as far as this study is concerned. We will analyze the Nigerian population with regards to population projection. Two models are employed for this purpose, namely the Logistic differential equation model and a nonlinear Regression model to estimate population projection up to 2036.

The main models

In this section we present the various models involved in population modeling and prediction in this project. They are namely;

- 1. The Malthus's Model
- 2. The Logistic Model
- 3. The Exponential Regression model

Intrinsic Rate of Growth

The rate at which a population increases in size if there are no density-dependent forces regulating the population is known as the *intrinsic rate of increase*.

$$\frac{1}{N}\frac{dN}{dt} = r$$

Where (dN/dt) is the rate of increase of the population and *N* is the population size, r is the intrinsic rate of increase. This is therefore the theoretical maximum rate of increase of a population per individual. The concept is commonly used in insect population biology to determine how environmental factors affect the rate at which pest populations increase. See also exponential population growth and logistic population growth.

Some Common Mathematical Models

In this section we examine some common mathematical models which can be used to study population dynamics.

Exponential population growth

Exponential growth describes unregulated reproduction. It is very unusual to see this in nature. In the last 100 years, human population growth has appeared to be exponential. In the long run, however, it is not. Paul Ehrlich and Thomas Malthus believed that human population growth would lead to overpopulation and starvation due to scarcity of resources. They believed that human populations were going to grow at rate in which they exceed the ability at which humans can find food. In the future, humans would be unable to feed large populations. The biological assumptions of exponential growth are that the per capita growth rate is constant. Growth is not limited by resource scarcity or predation.

Simple discrete time exponential model

Consider the discrete time exponential model given by

(2)

 λ is the discrete time per capita growth rate. At $\lambda = 1$, we get a linear and a discrete time per capita growth rate of zero. At $\lambda < 1$, we get a decrease in per capita growth rate. At $\lambda > 1$, we get an increase in per capita growth rate. At $\lambda = 0$, we get extinction of the species.

Continuous time version of exponential growth (Malthus's Model)

In 1798 the Englishman Thomas R. Malthus proposed a mathematical model of population growth.

$$\begin{cases} \frac{dN}{dt} = \lambda N, \quad \lambda > 0\\ N(t_0) = N_0 > 0 \end{cases}$$

Where λ in equation 3(a) expresses population growth rate and it reflects a strong impact on how fast the population will grow. 3(a) indicates dN/dt increases to infinity as *t* increases, i.e. the model population increases to infinity as time goes to infinity. 3(b) expresses the population size when $t = t_0$. Malthus's model voices such principles: dN/dt

(1) Food is necessary for human existence;

(2) Human population tends to grow faster than the power in the earth to produce

subsistence.

(3) The effects of these two unequal powers must be kept equal.

(4) Since humans tend not to limit their population size voluntarily, population reduction

 $N_{t+1} = \lambda N_t$

tends to be accomplished through the "positive" checks of famine, disease, poverty and war.

The Solution of Malthus's Model

We give the mathematical resolution of equations 3(a)-3(b) here. Based on equation 3(a), separating variables, we have that;

$$\frac{1}{N}dN = \lambda dt$$

$$\Rightarrow \int_{t_0}^t \frac{1}{N}dN = \int_{t_0}^t \lambda dt$$
(4)

Thus

$$N = (N_0 e^{-\lambda t_0}) e^{\lambda t} = N_0 e^{\lambda (t-t_0)}$$
(5)

where $c = N_0 e^{-\lambda t_0}$ is some constant, for some fixed λ .

The resolution (5) of equations 3(a)-3(b) clearly indicates N is an exponential function of e. Its physical meaning expresses that population sizes grow exponentially as time t.

The Logistic Model

Malthus's model describes unconstrained growth, i.e. a model in which the population increases in size without bound. However, most populations are constrained by limitations on resources, even in the short run, and none is unconstrained forever. The following figure depicts three possible scenarios for growth of the population. The first curve expresses super-exponential growth and approaching a vertical asymptote, the second curve follows an exponential growth pattern, and the third curve is constrained so that the population is always less than some number K. When the population is *small* related to K, the patterns are virtually identical, in particular, the constraint doesn't make much difference. But as N becomes a significant fraction of K, the curves begin to diverge, and, in the constrained case, as N gets close to K, the growth *rate* drops to 0 (Here $K = \lambda$).

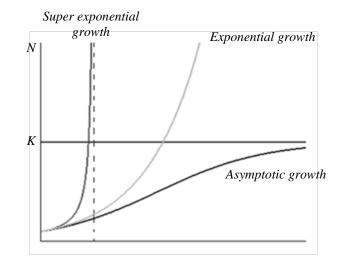


Fig.1 Population growth

as a function of time

In 1840 a Belgian mathematician Verhulst modified Malthus's Model. He believed population growth not only depends on the population size but also on how far this size is from its upper limit. He proposed a new model, called the Logistic model which is written,

$$\frac{dN}{dt} = \lambda N \left(1 - \frac{N}{K} \right)$$
⁽⁶⁾

Remark

All known variants of the logistic equation describe either a single-step evolution, called the *S*-curve, or an oscillatory behaviour around a constant level. However, the development of many complex systems consists not just of a single step, but by multistep growth phases, where a period of fast growth is followed by a lasting period of stagnation or saturation, which is itself followed by another fast growth regime, and so on. To capture the previously described phenomenology and much more, the constant *K* in equation (3.4) is replaced with a non-constant carrying capacity, which obviously increases the complexity of the model's behaviour. This is subject for future

International Journal of Scientific & Engineering Research, Volume 8, Issue 2, February-2017 ISSN 2229-5518

research. For the moment we are interested in obtaining a solution to equation (6).

Solution of the Logistic model

This is a first order ODE which can be solved by the method of separation of variables as follows. Separating variables we have

$$\frac{dN}{\left(K-N\right)N} = \frac{\lambda}{K}dt$$

Integrating both sides

$$\int \frac{dN}{\left(K-N\right)N} = \frac{\lambda}{K} \int dt + c$$

Where *C* is an arbitrary constant of integration to be determined from the initial condition

 $N(0) = N_0.$

Thus;

$$\int \frac{dN}{\left(K-N\right)N} = \frac{\lambda}{K}t + c$$

(7)

The LHS of the above equation can be integrated by first applying partial fractions to get

$$\int \frac{1}{\left(K-N\right)N} dN = \frac{1}{K} \int \left(\frac{1}{K-N} + \frac{1}{N}\right) dN$$
$$= \frac{1}{K} \ln \left(\frac{N}{K-N}\right)$$

Hence equation (7) becomes;

$$\frac{1}{K} \ln\left(\frac{N}{K-N}\right) = \frac{\lambda}{K}t + c$$
$$\Rightarrow \qquad \ln\left(\frac{N}{K-N}\right) = \lambda t + cK$$

Removing the logarithm, we get

$$\frac{N}{K-N} = e^{\lambda t + cK} = e^{\lambda t} e^{cK}$$

But e^{cK} is a constant equal to *A* say. Hence

$$\frac{N}{K-N} = Ae^{\lambda t}$$

Now applying the initial condition that $N(0) = N_0$.

$$\therefore \qquad \frac{N_0}{K - N_0} = A$$

$$\Rightarrow \qquad \frac{N}{K - N} = \frac{N_0}{K - N_0} e^{\lambda t}$$

Making N the subject of the above equation gives,

$$N = \frac{N_0 K}{N_0 + (K - N_0)e^{-\lambda t}}$$

(8)
We observe that as
$$t \to \infty$$
 , $e^{-\lambda t} \to 0$. Therefore

$$N \to \frac{N_0 K}{N_0} = K$$

This implies that after a very long period of time the population reaches a saturation level equal to *K*, the carrying capacity.

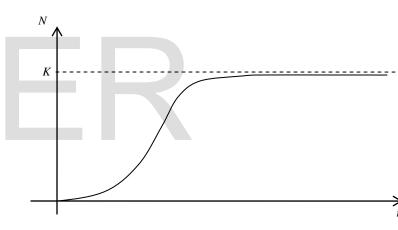


Fig. 2 Sketch of solution to the logistic population model.

Remark

From the sketch, we can see that when the population starts to grow, it does go through an exponential growth phase, but as it gets closer to the carrying capacity, the growth slows down and it reaches a stable level. This slowing down to a carrying capacity is perhaps the result of war, pestilence, and starvation as more and more people contend for the resources that are now at their upper bound. There are many examples in nature that show that when the environment is stable the maximum number of individuals in a population fluctuates near the carrying capacity of the environment. However, if the environment becomes unstable, the population size can have dramatic changes.

Results from the models

It is observed that Malthus's model works fairly well for a short period, that is the predicted population sizes are quite similar to those of the actual sizes. It starts to fall apart after that. However, this model assumes that the relative growth rate is constant. In fact, even if we ignore natural disasters, wars, and changes in social behavior, the growth rate would change as the population increased due to crowding, disease, and lack of natural resources. The model predicts that the population would grow without bound, however this cannot possibly happen indefinitely. One of the biggest failures in the Malthusian theory was that Malthus failed to foresee the immense technological innovation that was to occur, which increased crop yields and discovered resources. Malthus interprets new his mathematical conclusions in terms of the real world and compares the real world to the model. Well, ideally, this is supposed to be the case.

The logistic growth equation on the other hand is a useful model for demonstrating the effects of density-dependent mechanisms. However, the logistic growth model for real populations is limited because the dynamics of populations are complex and because it is difficult to come up with the real value for *K* in a given habitat. In addition, K is not a fixed number over time; it is always changing depending on many conditions. It is often limited by the current level of technology, which is subject to change. More generally, species can sometimes alter and expand their niche. If the carrying capacity of a system changes during a period of logistic growth, a second period of logistic growth with a different carrying capacity can superimpose on the first growth phase. For example, cars first replaced the population of horses but then took on a further growth trajectory of their own.

Modeling with Regression

Regression method is one of the most widely used statistical techniques. As a general definition, regression is a statistical technique for finding the best-fitting function for a given set of data. This function could be a straight line, polynomial or any arbitrary continuous function. Correlation analysis which is closely related to regression analysis is a statistical methodology used to assess the strength of relationship between predictor variable and response variable.

In this work we make use of the exponential regression model for population projection.

Exponential Regression

In our analysis we will make approximate some of our data with the exponential function $Y = \beta_0 e^{\beta_1 t}$, where β_0, β_1 are constants to be determined from the regression analysis. For the above exponential function we only need to transform it to a linear equations follows. Given

$$Y = \beta_0 e^{\beta_1 t}$$

Taking natural logarithms on both sides to get; $\ln Y = \ln \beta_0 + \ln e^{\beta_1 t}$

$$\Rightarrow \qquad \ln Y = \ln \beta_0 + \beta_1 t$$

Set $y = \ln Y$ and $\beta'_0 = \ln \beta_0$, the equation then
transforms to:

$$y = \beta_0' + \beta_1 t$$

Thus we can obtain the unknown parameters by performing a linear regression.

(9)

Derivation of the Linear Regression Model Given $y_i = \beta_0 + \beta_1 t_i + \varepsilon_i$, the square of the errors ε_i^2 is given by

$$\varepsilon_i^2 = \left(y_i - \beta_0 - \beta_1 t_i\right)^2$$

The sum of squares of the errors (SSE) is then

$$SSE = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 t_i \right)^2$$

It is the SSE we wish to minimize with respect to the unknown parameters β_0 and β_1 .

Hence we must compute;

$$\frac{\partial SSE}{\partial \beta_0} = 0, \qquad \frac{\partial SSE}{\partial \beta_1} = 0.$$

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 t_i \right) = 0$$
(10)

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 t_i \right) t_i = 0$$
(11)

Rearranging equation (10) we get

$$-2\sum_{i=1}^{n} (y_{i} - \beta_{0} - \beta_{1}t_{i}) = 0 \implies \sum_{i=1}^{n} \beta_{0} + \sum_{i=1}^{n} \beta_{1}t_{i} = \sum_{i=1}^{n} y_{i} \quad \beta_{1} = \frac{(\sum y_{i})(\sum t_{i}^{2}) - (\sum t_{i})(\sum t_{i}y_{i})}{n(\sum t_{i}^{2}) - (\sum t_{i})^{2}}$$

i.e.

$$n\beta_0 + \beta_1 \left(\sum_{i=1}^n t_i\right) = \left(\sum_{i=1}^n y_i\right)$$
(12)

Year	Population (× 10 ⁵)	
1973	54226	
1983	71948.3	

Similarly, rearrangin g equation (3.2) we get

$$-2\sum_{i=1}^{n} (y_{i} - \beta_{0} - \beta_{1}t_{i})t_{i} = 0$$

$$\Rightarrow \beta_0 \left(\sum_{i=1}^n t_i \right) + \beta_1 \left(\sum_{i=1}^n t_i^2 \right) = \left(\sum_{i=1}^n t_i y_i \right)$$
(13)

The normal equations are thus;

$$n\beta_0 + \beta_1 \left(\sum_{i=1}^n t_i\right) = \left(\sum_{i=1}^n y_i\right)$$
$$\beta_0 \left(\sum_{i=1}^n t_i\right) + \beta_1 \left(\sum_{i=1}^n t_i^2\right) = \left(\sum_{i=1}^n t_i y_i\right)$$

These can be put in matrix form as

$$\begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

The unknown coefficients can then be computed as;

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

From which we obtain the normal equations

$$\beta_0 = \frac{n \sum t_i y_i - \left(\sum t_i\right) \left(\sum y_i\right)}{n \left(\sum t_i^2\right) - \left(\sum t_i\right)^2},$$

(15)

Population without control

By using formula (5), we produce a chart of virtual population without control.

Table 1

Using this information (Appendix A) with t = 0 corresponding to the year 1973, we have $N_0=54226$. We can estimate $\,\lambda$ using the fact that N = 71948.3 when t = 10. By using $N=N_{0}e^{\lambda\left(t-t_{0}
ight)}$, we have;

$$71948.3 = 54226e^{10\lambda} \Longrightarrow \lambda = \frac{1}{10} \ln\left(\frac{71948.3}{54226}\right) \approx 0.028279$$
(16)

This gives us the general solution

$$N(t) = 54226e^{0.0283t}$$
(17)

We can now compute the population at later years and compare it with the actual data. The result is presented in the following chart.

International Journal of Scientific & Engineering Research, Volume 8, Issue 2, February-2017 ISSN 2229-5518

Year	POPA (× 10 ⁵)								
1973	55782.5	1982	71963.6	1991	92838.4	2000	119768.5	2009	154510.3
1974	57383.71	1983	74029.3	1992	95503.3	2001	123206.3	2010	158945.4
1975	59030.9	1984	76154.2	1993	98244.6	2002	126742.9	2011	163507.8
1976	60725.3	1985	78340.2	1994	101064.7	2003	130380.9	2012	168201.1
1977	62468.4	1986	80588.9	1995	103965.6	2004	134123.4	2013	173029.2
1978	64261.5	1987	82902.1	1996	106949.9	2005	137973.3	2014	177995.9
1979	66106.1	1988	85281.6	1997	110019.8	2006	141933.8	2015	183105.1
1980	68003.6	1989	87729.7	1998	113177.9	2007	146007.9		
1981	69955.6	1990	90247.9	1999	116426.5	2008	150198.9		

COMPARISON OF MODEL OUPUT

YEAR	MALTHUS MODEL	REGRESSION MODEL	TRUE POPULATION VALUE x10 ⁵
	x10 ⁵	x10 ⁵	
1973	54226	54899.72	55782.5
1983	71948.3	336052.8	74029.3
1994	74031.4	98243.55	101064.7
2004	133722.5	130378.9	134123
2014	170560.7	169930.8	177995.9

Based on the predicted population value from our research work, Regression model is considered to be the best model. Although in 1983, the model performed poorly due to some factors that is beyond our control.

Year	POPA (× 10 ⁵)								
1973	54226	1982	69865.2	1991	89263	2000	117171	2009	148278.3
1974	55865	1983	71948.3	1992	92057	2001	120481.3	2010	152488.8
1975	57905	1984	74031.4	1993	94934	2002	123791.6	2011	156818.8
1976	59143	1985	76114.5	1994	97900	2003	127101.9	2012	161271.8
1977	60782	1986	78197.6	1995	100959	2004	130412.2	2013	165851.2
1978	62421	1987	80280.7	1996	104095	2005	133722.5	2014	170560.7
1979	64060	1988	82363.8	1997	107286	2006	137032.8	2015	175403.9
1980	65699	1989	84446.9	1998	110532	2007	140203		
1981	67782.1	1990	86530	1999	113829	2008	144184.1		

POPULATION ESTIMATE (ACTUAL)

Population growth rate

Year	Growth								
	Rate (%)								
1973	0	1982	2.4033	1991	1.4396	2000	0.7678	2009	0
1974	0.2519	1983	2.7233	1992	1.7455	2001	1.0249	2010	0.2443
1975	0.5038	1984	3.0433	1993	2.0604	2002	1.2821	2011	0.2581
1976	0.7558	1985	0	1994	2.3851	2003	1.5392	2012	0.6543
1977	1.0075	1986	0.2456	1995	2.7201	2004	1.7963	2013	0.7821
1978	1.2592	1987	0.4561	1996	3.0634	2005	2.0534	2014	0.987
1979	1.5117	1988	0.6842	1997	0	2006	2.3106	2015	1.188
1980	1.7633	1989	0.9132	1998	0.2521	2007	2.4102		
1981	2.0833	1990	1.1403	1999	0.5082	2008	2.5216		

[1993, Census Department, Bureau of Census (for Nigeria) International Data Base]

References

- [1] Kpendekpo G. M. K. (1982). Essentials of Demographic Analysis for Africa. New
- Hampshire:Heinemann Educational Books Inc.

[2] Olabode, O.. & Akinyokun, O. C. (2002). Computer Aided System for Demography in

Nigeria, Master Thesis, Federal University of Technology, Akure.

[3] United Nations. (2001). The State of the Demographic Transition In Africa: Executive

Summary, Economic Commission for Africa, Joint ECA/OAU/ADB Secretariat in

Collaboration withUNFPA. FSSDD/ICPD/FC.4/01/3, Yaounde, Cameroon.

[4] Takahashi, S., Kaneko, R., Ishikawa, A., Ikenoue, M., & Mita, F. (1999). Population

Projections for Japan: Methods, Assumptions and Results. *Review of Population and*

Social Policy, 8, 75-115.

[5] Heisel, D. F. (1969). Measuring Current Population Changes. *Population of Tropical*

Africa (3rd ed., pp. 34-49). London: Lowe & Bryone (Printer) Ltd.

[6] Henry, M. A. (1978). *Population of Tropical Africa* (3rd ed.). London: Lowe & Bryone

(Printer) Ltd.

[7] Kpendekpo G. M. K. (1987). Essentials of Demographic Analysis for Africa. New

Hampshire: Heinemann Educational Books Inc.