Journal of the Korean Data & Information Science Society
2018, **29**(2), 567–576
http://dx.doi.org/10.7465/jkdi.2018.29.2.567
한국데이터정보과학회지

A statistical approach on demographic dynamics and population projection in Tanzania

Peter A. Mwandri[1] · Kee-Won Lee[2] · Songyong Sim[3]

[123]Department of Statistics, Hallym University

Received 28 December 2017, revised 20 February 2018, accepted 22 February 2018

Abstract

Population projection provides profound information to policy makers of formulation of policies and decision making. It also provides information about evaluation of the progress towards achieving Millennium Development Goals in Tanzania. We use five different models to project Tanzanian population: linear model, exponential model, modified exponential model, logistic model and cohort component model. Among these models, logistic growth model provides the lowest projection followed by modified exponential model. On the other hand, the exponential growth model gives the highest projection followed by cohort component model. Cohort component model seems to be the best projection model among the above mentioned models since it incorporates birth, death and migration information in projection process.

Keywords: Concept of cohort, demography, fertility, population projection, survivorship function.

1. Introduction

The United Republic of Tanzania is located in the eastern part of Africa and is situated just south of equator at latitude and longitude 5.6944°S, 36.3223°E with the area of 947,303 square kilometres (Kimambo and Temu, 1969).

Demography is the statistical study of populations, especially human beings (John and Stephene, 1964). As a very general science, it can analyse any kind of dynamics living population, as one that changes over time or space due to birth, death and migration. Estimation of population mean is studied by Kim and Jung (2002) and Kim (1999, 2000) with different sampling methods. Demographics are quantifiable characteristics of a given population. One of the earliest demographic studies in the modern period was performed by Graunt (1676) which contains a primitive form of life table.

A population projection is an estimate of the population composition and distribution of a future population. These projections are influenced not only by the methods and assumptions

[1] Graduate students, Department of Statistics, Hallym University, Chuncheon 24252, Korea.
[2] Professor, Department of Statistics, Hallym University, Chuncheon 24252, Korea.
[3] Corresponding author: Professor, Department of Statistics, Hallym University, Chuncheon 24252, Korea. E-mail: sysim@hallym.ac.kr

used in their production, but also by the historical data series upon which they are based. Census counts and post-censal estimates typically serve as the empirical foundation for the population upon which projections are based, while vital statistics and immigration data serve as the empirical foundation for births, deaths, and immigration.

When conducting projection the purpose is not to predict the population number for the future years or decades, but to provide the size that population will have if certain assumption regarding mortality, fertility and migration levels and trend take place in the future (Mendonca, 2005).

2. Assumptions

For this study all the projection were conducted under the following assumptions:

- Population is considered as closed population and also stable population.
- The form of government, the political, economic and social organization and institutions of the Tanzania will remain substantially unchanged.
- No war, national wide devastation, epidemic disease, internal revolution or other disaster will occur.
- No natural calamities like earthquake, drought, flood, hurricane, tsunami and other related geologic processes resulting disasters.

Any of these events might have completely unpredictable effects on Tanzania population.

3. Population projection models

There are various models and techniques for population projections. Linear growth, exponential growth, modified exponential and logistic growths models are among the methods used for a population projection.

3.1. Linear model

When we assume that the growth of population is linear, which means that there is a constant amount of increase per unit of time, we can use a linear growth model.

Let P_t be the projected population in the year Y_t based on the launch year population P_l of year Y_l and base year population P_b of year Y_b. Then linear growth model use the following equation to project population P_t as follows;

$$P_t = P_l + B \cdot \Delta t, \tag{3.1}$$

where $\Delta t = Y_t - Y_l$ is the projection horizon and $B = (P_l - P_b)/(Y_l - Y_b)$ is the slope.

3.2. Exponential model

This growth model is mostly important for modelling biological population like bacteria and other micro-organisms. The number of micro-organisms in a culture will increase exponentially until an essential nutrient is exhausted (Nikolai *et al.*, 2014).

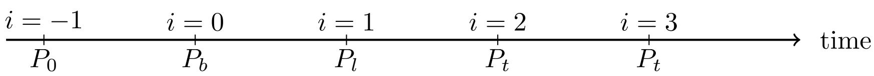
Let the target year population P_t at the year Y_t based launch year population P_l at the year Y_l and r be the population growth rate. The rate that the population changes as time passes can be written as $\frac{dP}{dt}$. This rate is equal to the population growth rate times launch year population (Malthus, 1798) so that $\frac{\mathrm{d}P}{\mathrm{d}t} = rP_l$. By integrating both the sides of the equation, we have $\int \frac{dP}{P} = \int rdt$, and solving the result of integration for P_t, we have

$$P_t = P_l e^{rt}. \tag{3.2}$$

3.3. Modified exponential model

Characteristics of linear growth model or exponential growth model is that the population level continues to grow to an indefinite extent. But Modified exponential growth model is based on the assumption that the population level have finite limit.

This model projects populations at the beginning of time interval i, where $i = -1$ for the interval (before the base year, the base year), $i = 0$ for the interval (base year, launch year), and $i = 1$ for the interval (launch year, first target year) and so on. All the time time intervals are of the same length.



For this model, we use: earlier population census previous the base year P_0, base year population P_b and launch year population P_l. Then, let $B = \frac{P_l - P_b}{P_b - P_0}$, $A = \frac{P_l - P_b}{B - 1}$ and $C = P_b - A$.

Then the projected population of target year t for the i-th time interval i is

$$P_t = C + AB^i. \tag{3.3}$$

3.4. Logistic model

Logistic growth of a population size occurs when resources are limited, thereby setting a maximum number of population an environment can support. It is designed to yield an S-shaped pattern representing an initial period of slow growth rates, followed by a period of increasing growth rates, and finally a period of declining growth rates that approach zero as a population approaches its upper limit (Shryock and Siegel, 1973).

Let P_{sat} be the maximum limit of the population, P_t be the projected population in the year Y_t based on the launch year population P_l of year Y_l, base year population P_b of year Y_b and the earlier population census previous the base year P_0. Then logistic growth model use the following equation to project population P_t;

$$P_t = \frac{P_{sat}}{1 + e^{a+b\Delta}}, \tag{3.4}$$

where $P_{sat} = \dfrac{2P_0P_bP_l - P_b^2(P_0 - P_l)}{P_0P_l - P_b^2}$, $a = \ln\left(\dfrac{P_{sat} - P_l}{P_l}\right)$ and $b = \dfrac{1}{n}\ln\left(\dfrac{P_0(P_{sat} - P_b)}{P_b(P_{sat} - P_0)}\right)$.

3.5. Cohort component model

In this model the future population is a function of present population and three demographic components: fertility, mortality and migration. This method was first introduced by Cannan (1895), after that used by Bowley (1924) and later rediscovered independently by Whelpton (1928). Under the cohort component model births, deaths and migration are projected independently for each age group in the population.

The survivorship function is the chance of an individual surviving from birth to age x. This model consisting of segmenting the population into different subgroups by age and sex.

Let P_x be the population of a cohort at age x and L_x be the number of persons in a cohort surviving from birth to age x, ${}_{10}P_x$ be the population of cohort at age interval x to $x + 10$ and ${}_{10}L_x$ be the number of persons lived by a cohort in the age interval x to $x + 10$.

For each projection interval, the method basically consist of three steps (Preston and Hauveline, 2001):

1. We start with projecting the population in each subgroup at the beginning of the time interval in order to estimate the population still alive at the beginning of the next interval based the survivorship function.

2. Then we compute the number of births for each subgroup over the time interval, add them across groups and compute the number of births who survive to the beginning of the next interval.

3. After that we add immigrants and subtract emigrants in each subgroups during the interval.

When using cohort component method we take into account that, births are produced by women only. Therefore the number of births are estimated by applying fertility rates to women only, and male's population are assumed to be irrelevant to rate of childbearing. We will produce the projection for 10 years age groups and 10 years projection intervals, assuming that for all 10 years interval the fertility and mortality rate will remain unchanged. We will denote the number of females or males aged x to $x+10$ at the beginning and at the end of projection intervals as ${}_{10}P_x(t)$ and ${}_{10}P_x(t + 10)$. Survival ratio will be denoted as S_x which is calculated from life table, $S_x = \dfrac{{}_{10}L_x}{{}_{10}L_{x-10}}$, the proportion of the persons aged $x - 10$ to x who will be alive 10 years later in a stationary population subject to the appropriate life table.

The projected population still alive 10 years later is done by applying survival ratio to each age group as follows; ${}_{10}P_x(t + 10) ={}_{10} P_x(t) \cdot S_x$.

For open-ended age group (80+) we need to combine the survivor from two previous age group; ${}_{\infty}P_x(t + 10) ={}_{\infty} P_x(t) \cdot S_x +{}_{10} P_{x-10}(t) \cdot S_{x-10}$.

We need to project the number of births during the projection period using age-specific fertility rate ${}_{10}F_x$; Birth $={}_{10} F_x \cdot 10 \cdot {}_{10} P_x(t) ={}_{10} F_x \cdot 10 \frac{{}_{10}P_x(t) +{}_{10} P_{x-10}(t) \cdot S_x}{2}$.

The total number of births during the period is then obtained by summing births across age group of the mother; $B[t, t+10] = \sum_{x=10-19}^{40-49} {}_{10}F_x \cdot 10 \frac{{}_{10}P_x(t) +{}_{10} P_{x-10}(t) \cdot S_x}{2}$.

The number of female births at the age 0-9, are then obtained by applying the ratio of male to female birth SRB. In our case for Tanzania SRB is 1.03 (United Republic of Tanzania, 2013).

$$B^F[t, t+10] = \frac{1}{1+\mathrm{SRB}} \sum_{x=10-19}^{40-49} {}_{10}F_x \cdot 10 \frac{{}_{10}P_x(t) +{}_{10} P_{x-10}(t) \cdot S_x}{2}. \tag{3.5}$$

In the same way the number of male births at the age 0-10 can be calculated as:

$$B^M[t, t+10] = \frac{\mathrm{SRB}}{1+\mathrm{SRB}} \sum_{x=10-19}^{40-49} {}_{10}F_x \cdot 10 \frac{{}_{10}P_x(t) +{}_{10} P_{x-10}(t) \cdot S_x}{2}. \tag{3.6}$$

3.6. Population Growth rate r

Is the rate at which the number of individuals in the population increase in a given time period articulated as a fraction of initial population. Growth rate increased by an increase in fecundity and survival rate.

Given the projected population P_t and present population P_l, population growth rate R can be found by the following equation:

$$R = \frac{P_t - P_l}{P_l} \times 100. \tag{3.7}$$

Also we can calculate average annual population growth rate r of the population. This can be obtained by using compound method by considering P_t, number of years n and P_l, then r can be obtained using the following equation:

$$r = \left[\frac{P_t}{P_l}\right]^{1/n} - 1. \tag{3.8}$$

4. Population projection

We project Tanzania total population, population by region and population by gender for the year 2022, 2032, 2042 and 2052 by using the models we discussed in section 3, also we will compare the projection using different models. The population of any given country grows or decline through the interaction of three demographic factors: fertility, mortality and migration (Shryock and Siegel, 1973).

4.1. Total population

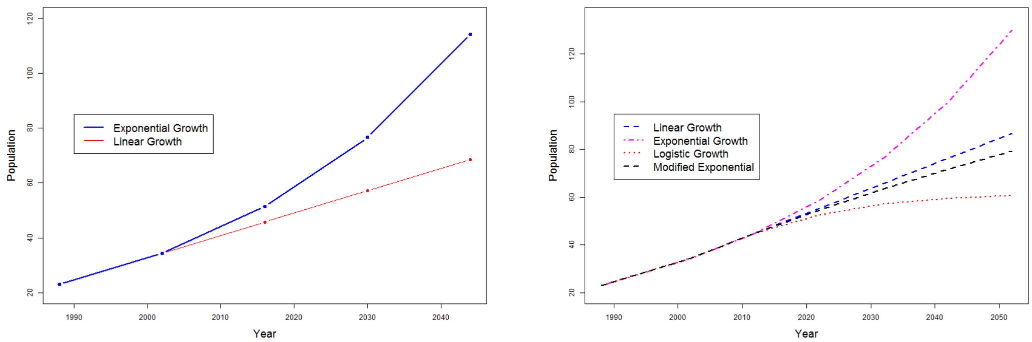
We used four projection models discussed earlier in subsection 3.2, subsection 3.3 and subsection 3.4 to project total population of Tanzania based on census data for the year 2002 and 2012. The projected populations are summarized in Table 4.1.

Table 4.1 Total population projection for the launch year population 44,928,923

Year	2022	2032	2042	2052
Linear Method	55,414,243	65,899,563	76,384,883	86,870,203
Exponential	58,606,183	76,447,075	99,719,090	130,075,571
Logistic	52,523,597	57,108,906	59,579,447	60,829,370
Modified	54,617,382	63,56,9541	71,841,356	79,484,531

The four model we used are linear growth model of equation (3.1), exponential growth model of equation (3.2), logistic growth model of equation (3.4) and modified exponential growth model of equation (3.3).

The Tanzania population is projected to reach 86.87 million in 2052 using linear growth model (Eq. (3.1)), an increase of 41.9 million for a projection horizon of 40 years. This is equivalent to 1.7 percent average annual rate of growth. High projections are observed when using exponential growth model (Eq. (3.2)) in which we project the population growth to 130.08 million for the period 2012–2052. This is equivalent to 2.7 percent average annual rate of growth which seems to be very high. The medium growth rate can also observed when using modified exponential growth model in which the population will grow to 79.48 millions in 2052, equivalent to 1.4 percentage average annual growth rate. The lowest population projections is obtained when using logistic growth model in which the average annual growth rate will be only 0.8 percent.



(a) Population projection using 1988 as base year (b) Population projection using 2002 as base year

Figure 4.1 Tanzania total population projection

In general, we observe the logistic model provides the lower population projections, the modified exponential model provides the second lowest, the linear model gives the second highest population projections while the exponential model provides the highest projections of population.

Also, according to logistic growth model, year 2042 Tanzania population is projected to grow more slowly than it was in many decades before, in which the average annual growth rate is projected to be slow as 0.2 percent for the year 2042–2052.

4.2. Population projection by regions

We observe the logistic model provides the lowest population projections, the modified exponential model provides the second lowest, the linear model gives the second highest population projections while the exponential model provides the highest projections of population across all regions as well as the total population in Section 4.1.

Table 4.2 Tanzania population projection by regions

Region	Linear	Exponential	Modified exp.	Logistic
Dar es Salaam	6,070,776	7,361,478	7,129,069	6,732,409
Kagera	3,068,461	3,302,701	3,320,192	3,218,569
Mbeya	3,278,906	3,457,941	3,300,378	3,212,614
Mwanza	3,389,057	3,607,760	3,115,290	2,908,923
Tabora	2,791,377	2,962,145	2,684,091	2,547,948
Morogoro	2,596,660	2,698,058	2,508,698	2,435,074
Tanga	2,404,162	2,494,251	2,434,217	2,405,977
Dodoma	2,409,319	2,485,339	2,332,319	2,279,215
Kigoma	2,294,669	2,352,190	2,102,180	2,046,548
Mara	2,050,385	2,136,925	1,989,851	1,928,852
Arusha	2,037,148	2,146,048	1,920,495	1,826,257
Kilimanjaro	1,825,510	1,862,088	1,786,223	1,764,834
Simiyu	1,806,287	1,849,592	1,687,534	1,648,278
Manyara	1,774,309	1,905,075	1,718,901	1,615,255
Mtwara	1,375,051	1,389,010	1,316,451	1,304,814
Rukwa	1,257,922	1,355,403	1,183,290	1,095,627
Lindi	917,692	923,062	882,626	877,987

4.3. Population projection by gender

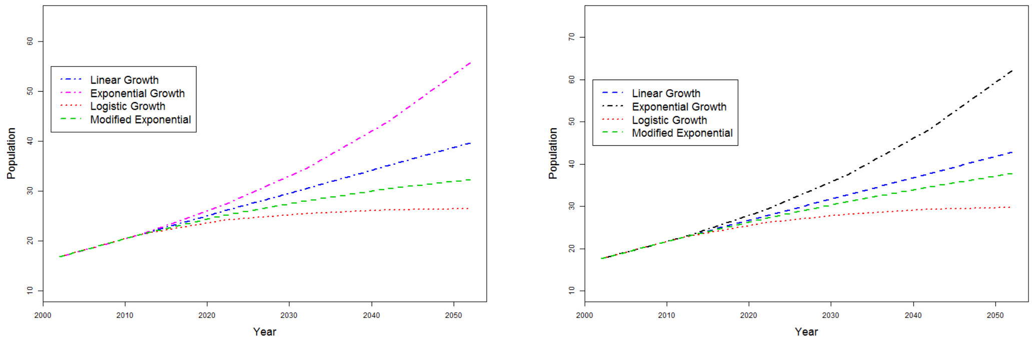
The lauch year female population in Tanzania is larger than male population across regions. It happens the same in projections, the projected female population is larger than male projected population.

In Table 4.3, for the year 2022 the highest population projection estimated to be 27,172,282, the medium population projection estimated to be 25,939,555 and the lowest population projection estimated to be 24,142,356 in case of male. On the other hand female population, the highest population projection estimated to be 29,165,286, the medium population projection estimated to be 27,716,594 and the lowest population projection estimated to be 26,156,914.

Table 4.3 Male and female population projection

Year	Male				Female			
	Linear	Exponen.	Logistic	Modified	Linear	Exponen.	Logistic	Modified
2022	25,939,555	27,172,282	24,142,356	25,060,093	27,716,594	29,165,286	26,156,914	27,139,486
2032	30,494,402	34,526,210	25,531,941	28,025,824	32,768,020	37,529,568	28,177,515	31,102,630
2042	35,049,249	43,870,411	26,167,617	30,418,924	37,819,446	48,292,632	29,228,494	34,612,998
2052	39,604,096	55,743,534	26,445,505	32,349,957	42,870,872	62,142,424	29,744,908	37,722,320

For the projection horizon of 40 years from 2012 to 2052, the highest average annual growth rate for male was 2.4% per year with exponential model and the lowest average annual growth rate is 0.5% with logistic model. On the other hand, the highest average annual growth rate for female was 2.6% with expoential model and lowest is recorded to be 0.7% per year for female with logistic model.



(a) Male population projection (b) Female population projection

Figure 4.2 Population projections by gender

In general all models estimate female population outnumbering male population, this scenario can be due to current female population is higher than male population.

We projected male and female births based on equations (3.6) and (3.5), respectively. When using cohort component method, the projection for male and female are done separately and the number of male and female births are then projected by applying sex ratio at birth (SRB). Table 4.4 summarizes Tanzania 2022 population projection for male and Table 4.5 shows 2022 population projection for female for 10 year age intervals. Also, in comparison to other previous projection models cohort component method appears to supply higher population projection nearly to exponential model in Table 4.3.

Table 4.4 Tanzania 2022 male population estimationby cohort component method

Age	${}_{10}P_x$	S_x	${}_{10}F_x$	$B[t, t+10]$	Migration	Projection
0-9	6,880,295	0.82110			1,140,010	8,339,579
10-19	5,035,730	0.97084	0.072	3,524,910	$-1,185,754$	4,463,720
20-29	3,206,320	0.95243	0.211	6,453,541	650,538	5,539,457
30-39	2,484,054	0.91085	0.179	4,069,005	68,838	3,122,639
40-49	1,618,311	0.86762	0.067	946,032	92,365	2,354,990
50-59	977,885	0.82390			8,003	1,412,088
60-69	609,095	0.54443			$-17,024$	788,657
70-79	365,962	0.42573			13,713	345,323
80+	207,056	0.49522			66,732	325,073
Sum						26,691,526

Table 4.5 Tanzania 2022 female population estimation by cohort component method

Age	${}_{10}P_x$	S_x	${}_{10}F_x$	$B[t,t+10]$	Migration	Projection
0-9	6,870,757	0.8477			1,720,931	9,783,495
10-19	5,165,433	0.9767	0.072	3637575	$-1,148,196$	4,676,455
20-29	3,909,921	0.9524	0.211	7913523	$-24,862$	5,020,319
30-39	2,741,846	0.9141	0.179	4507320	215,714	3,960,375
40-49	1,716,238	0.8649	0.067	1000185	$-1,135$	2,505,221
50-59	970,377	0.8455			47,446	1,531,904
60-69	635,621	0.5678			3,494	823,983
70-79	393,912	0.4308			24,421	385,378
80+	261,063	0.4993			40,967	341,028
Sum						29,028,158

5. Conclusion

We compare all five models that we have studied projections for male, female and total population in Table 5.1. The sizes of projected population are of the same order regardless of gender or total: logistic model projects the lowest and expoential model the highest.

Table 5.1 The size of projections by different models

Male	Logistic < Modified < Linear < Cohort < Exponential
Female	Logistic < Modified < Linear < Cohort < Exponential
Total	Logistic < Modified < Linear < Cohort < Exponential

The reason why logistic growth model projects the lowest is: it is a function of the amount of available resources in the environment for population to grow so that population growth rate gets smaller as population size approaches to a maximum imposed by limited amount of resources in the environment.

As the population at the intermediate time point is above the line of two end points, modified exponential model projects lower than the linear model. With the same reason, modfied expoential model projects lower than exponential model as it is modified by intermediate time pont.

Linear model assume that population will grow by the same growth rate in the future as in the past. This assumption is theoretically true but practically is not true.

Exponential growth model considers the population growth rate stays the same regardless of population size, which makes the projected population grow faster and faster as it gets larger. This model is the best for studying bacterial growth in the laboratory.

While logistic, modified, linear and exponential models use only the populations of previous two or three time points, cohort component model incorporates birth, death and migration information as well as the population sizes. So we think projection based on cohort component method can be a reference when we compare above models.

The projection results show that Tanzania population is growing very fast. This phenomenon can be due to various factors including improvements in health services/facilities, increases in life span, migration and high fertility rate. The total fertility rate for the year 2012 was 4.8 children per woman, which makes the population to grow very fast close to exponential growth model. In comparison with EU, the total fertility rate for EU is recorded to be 1.58 live birth per woman (Statistical office EU), this slows the population growth rate down and sometimes the population remains constant.

References

Bowley, A. (1924). Births and population in Great Britain. *The Economic Journal*, **334**, 188-192.

Cannan, E. (1895). The probability of a cessation of the growth of population in England and Wales during the next century. *The Economic Journal*, **5**, 506-515.

Graunt, J. (1676). *Natural and political observation made upon the bills of mortality*, Oxford, London.

John, D. and Stephene, A. (1964). *Demography*, Springer Science + Business Media, New York.

Kim, H. J. (1999). A study on estimating population mean by use of interpolation and extrapolation with balanced systematic sampling. *Journal of the Korean Data & Information Science Society*, **10**, 91-102.

Kim, H. J. (2000). A new estimator of population mean based on centered balanced systematic sampling. *Journal of the Korean Data & Information Science Society*, **11**, 91-101.

Kim, H. J. and Jung, S. J. (2002). On the effectiveness of centering, interpolation and extrapolation in estimating the mean of a population with linear trend. *Journal of the Korean Data & Information Science Society*, **13**, 365-379.

Kimambo and Temu (1969). *The history of Tanzania*, East African Pub. House, Tanzania.

Malthus, T. (1798). *An essay on the Principal of Population*, Oxford, London.

Mendonca, M. (2005). *Population projection*, National Statistics Directorate, Timor-Leste.

Nikolai, B., Bogdan, A., Schwab, D., Edoardo, A., van Oudenaarden, M. and Alexander (2014). Constant growth rate can be supported by decreasing energy flux and increasing aerobic glycolysis. *Cell Reports*, **7**, 705-714.

Preston, S. and Hauveline, P. (2001). *Demographic measuring and modelling population process*, Oxford, London.

Shryock, H. and Siegel, J. (1973). *The methods and materials of demography*, U.S Census Bureau, Washington DC.

Statistical office EU. `http://ec.europa.eu/eurostat/statistics-explained/index.php/statistics`

United Republic of Tanzania (2013). *Population distribution by age and sex*, National Bureau of Statistics, Tanzania.

Whelpton, P. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, **34**, 253-270.