Pak. J. Statist.
2015, 31(2), 219-228

PROJECTING THE POPULATION AGE DISTRIBUTION OF PAKISTAN USING MARKOV CHAIN

Muhammad Zakria[1] and **Faqir Muhammad**[2]
[1] Department of Statistics, Allama Iqbal Open University
Islamabad, Pakistan. Email: zakria@aiou.edu.pk
[2] Department of Management Sciences, Air University
Islamabad, Pakistan. Email: aioufsd@yahoo.com

ABSTRACT

Modeling the drastic change in population during this modern era has become the major interesting area among the demographers of the world, particularly of the developing countries. Usually, the exponential and time series models are exercised for overall population projection. These models seem lacking in projecting the population age distribution. The population age distribution projection has its own practical importance for several reasons in policy making. This article investigates the projection of population age distribution of Pakistan as well as the disparity index of the projected and census of populations. A Markov Chain (MC) model has been developed to project the population of different age groups and found to be parsimonious. The population age distribution is projected up to 2021 using 1972 and 1981 census populations to calculate the probabilities of survival. The Gini coefficients and Lorenz curve are also estimated to examine the disparity of projected and census populations. The findings demonstrated the mild level of disparity within the population census1972; 1981; 1998. The disparity measures of projected population age distribution for the year 2011 and 2021 are approximately same as that of 1998 population census. The results indicated that the population age distribution pattern would not be changed drastically if the social behavior of the society continues similar to the patterns reflected by the year 1998. However, the projection also anticipated about the decrease in 0-4 age group and increase in ageing population.

KEY WORDS

Population projection; Markov Chain; Gini coefficient; Lorenz Curve.

1. INTRODUCTION

Population may be defined as the total number of persons inhabiting in a country, city, any district or area. Although the awareness of overall population size is important but the age distribution of a population is the most important for a number of reasons. The age composition of a population has its own significance role on the socio economic development of the country. The proportions of children, working population and ageing population have much to do with the balance of national expenditures. It is a fact that all the policy matters directly or indirectly relate to the relative population proportions, even political influence may also be linked with it. Population studies, particularly, population

projection has become one of the most interesting research areas for the demographers all over the world. Currently, Pakistan is ranked 6th populous country in the world with population of 196.17 million (International Data Base [IDB], 2014) whereas in 1950, it was ranked 14th with population of 34 million (Statistics Division, 1972). It implies that the population of Pakistan has been increased more than five times during the last six and a half decade. No doubt, such massive increase in overall population size is an enormous pressure on the available national resources. Definitely, it complicates the policy decisions and population projects of Pakistan. Zakria and Muhammad (2009) projected the population of Pakistan using ARIMA models. Jan *et al.* (2007) projected the NWFP population of Pakistan using different mathematical models. Zakria, Muhammad and Salau-ud-din (2009) fitted the traditional and time series models on the population of Pakistan and projected the population for some years ahead. Keyfitz (1964) projected the Norwegian brown rats female population as well as for some hypothetical beetle's population using the matrix approach. The main requirement of this technique is the calculation of the survival probabilities of the group of objects. These probabilities were obtained from the female population of ages 0-45 during the years 1940-1955. Later on, the whole population was divided mainly into three groups i.e. 0-14, 15-29. 30-44 and projection were made for the next 60 years upto 2000. One million population of age 0-14 was used as an initial vector. Nichols *et al.* 1992) estimated the transition probabilities for stage-based population projection matrices using capture-recapture data. Fujiwara and Caswell (2002) used the multistage mark recapture method (MSMR) on the North Atlantic right whale (Eubalaena glacialis) data and transition probabilities were converted into matrix form using the MSMR method. Loh and George (2007) scrutinized the effect of net international migration on the Canadian population and its age sex distribution during the next 50 years using the base population of 2005. It was concluded that if net international migration continues, the size of the population may increase but the age sex distribution would not be increased as desired by Canadian Government. Moreover, it was also reported that if international migration continues, the percentage of the dependent population (<15 & 65+) is expected to be 44.3%, 61.2% and 68.7% for the years 2005, 2031 and 2056 respectively. Similarly, these percentages would be 68.0% &80.9% in 2031 and 2056 respectively without considering the net international migration. Klosterman (1990) discussed the long term and short term population forecasting. According to his point of view, approximately, twenty years population forecasting may be considered as long term and two to three years forecasting as short term. Klosterman's statement seems sensible due to the drastic instability in the growth rate of population as well as socio economic and demographic factors. Keeping in mind these considerations jointly with the significance of the population age distribution, in this work a MC Model to project the population age distribution of Pakistan is proposed. The population inequality overtime was also examined within the age distribution of projected and census population of Pakistan. Population of different age segments of Pakistan particularly childhood (0-4) and dependent population (0-15 and 64+), working group (15-64) and ageing population (64+) were also projected.

This article is organized as follows: in section 2 some basics of methodological background are given whereas, the following section 3 includes results and discussion and some conclusions are given in the 4th section.

2. MATERIAL AND METHODS

The most authentic and reliable empirical data of last three population census of Pakistan i.e. 1972, 1981 and 1998 have been taken from the Statistics Division (1972; 1984; 2001). One step transition probability matrix $\left(P_{ij}\right)$ with transition probabilities P_{ij} i.e. the probability of an object of moving from state i to state j. The P_{ij} has been discussed almost in all books of stochastic process. Ross (2008) also described this matrix in such a way, the probabilities are non-negative and the process must make a transition into some state, such that $P_{ij} \geq 0$, $i, j \geq 0$; $\sum_{j=0}^{n} P_{ij} = 1, i = 0,1,...,n$.

At present although different methods of overall population projection of any territory are available and practically acceptable but there is massive lacking regarding the projection of population age distribution. The main inspiration of this research article was to suggest a suitable mechanism or model to project the population age distribution of Pakistan. Keeping in view the fertility scenario, Ross's one step transition probability matrix was modified according to the real life population phenomena and developed a Markov Chain of order 9×9 (see matrix P_m). The one step transition probability matrix is considered in such a fashion that an object may go to the next state but can never be returned to the previous state. However, an object may remain in the same state (means object is passed away).It is a fact that the new born babies always compensate the passed away population as well as cause to maintain or to increase the total population. Although the census population in different age groups belongs to different birth cohort but there is no doubt that the new born babies are always positioned in first age group (0-9). The 9th column consists of the probability of dying of each state/age group whereas; the 9th row is named source/sink state. The first and the last entry of source/sink state are "r" and "$1-r$" respectively where "r" represents the new born babies equal to the relative proportion of the total population.

$$P_m = \begin{bmatrix} 0 & p_1 & 0 & \cdots & 0 & 0 & 1-p_1 \\ 0 & 0 & p_2 & \cdots & 0 & 0 & 1-p_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & p_{n-2} & 1-p_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ r & 0 & 0 & \cdots & 0 & 0 & 1-r \end{bmatrix}$$

where p_1 is the one step transition probability of moving of an object from State 1 (age group 0-9, babies) to State 2 (age group 10-19, adolescents) during the time period of ten years. Similarly p_2 is the transition probability of moving from State 2 (age group 10-19, adolescents) to State 3 (age group 20-29, adult) during the same time period of ten years and so on. The resulting Markov Chain is irreducible/regular and hence a stationary distribution (limiting distribution).In addition to the one step transition probability matrix (tpm), in proposed MC model, an initial row vector $X^{(0)}$ of the following type is also required.

$$X^{(0)} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \cdots\cdots & x_{n-1} & x_n \end{bmatrix}$$

The mechanism to project the population of different age groups for every next decade is described in the following mathematically relation.

$$\hat{X}^{(1981)} = X^{(1972)} \; P_m$$

$$\hat{X}^{(1991)} = \hat{X}^{(1981)} \; P_m$$

and so on

where $X^{(1972)}$ and P_m are the initial row vector containing the population of different age of 1972 population census (base year) and one step transition probability matrix respectively. This matrix has been developed from two consecutive population census (1972-1981) using the age interval of 10 years.

Disparity Measures

Population structure means the 'make up' or composition of a population. The majority of the social scientists have a keen interest in the population age distribution since the social relationships with a community are considerably affected by the relative numbers of individuals at each age. Almost all types of planning, particularly planning of community institutions and services, require the data on age composition (Siegel and Swanson, 2004). It is considered inevitable for planning and administration of the country from micro to macro level. No country can achieve its targets without the optimum knowledge of population age sex distribution. First of all Gini (1912) introduced the disparity measures for income and wealth inequality. These measures are known as Gini coefficients and Lorenz curve. The Gini coefficient theoretically ranges from 0 (complete equality) to 1(complete inequality). Cowell (2011) illustrated the use of inequality measures other than income and wealth areas but mostly within the Economics. Pan American Health Organization (2001) computed the biased and unbiased Gini coefficients and their confidences intervals to measure the inequality of different data sets e.g. GNP per capita, infant mortality rate, live births and infant deaths etc. Rodrigue *et al.* (2013) also discussed the traffic inequality at different terminals by computing the index of dissimilarity (ID); Gini coefficient (G) and Gini's means difference (GMD). Brown (1994) used the Gini-style indices to evaluate the spatial patterns of health practitioners and used Alberta data for theoretical considerations as an application. Keeping in view the importance of Gini coefficients and Lorenz curve, an attempt has been made to apply these measures in the field of population studies. The disparity measures have also been calculated to examine the inequality of census and projected population age distribution as well as the projected population of significant age segments like babies, adolescents, working, dependent and ageing population.

3. RESULTS AND DISCUSSION

The population of different age groups was projected by Markov Chain for the period of 1981-2021 and is given in Table 1. The overall projected population for the years 1981, 1991, 2001, 2011 and 2021 are 83.59 million, 107.52 million, 135.80 million, 176.18 million and 228.77 million respectively. The projected populations are almost

close to the population projections (Total Population by Country, 2014; Country Comparison Population, 2014; International Data Base [IDB], 2014; National Institute of Population Studies [NIPS] 2014; Finance Division, 2012-13; World Population Prospects [WPP], 2012).The projected population in different age groups from 0-9 age group to 70+ are approximately close to the actual population of 1981 population census of Pakistan (Statistics Division, 1984). It has been observed that the accuracy of projected population depends only on the reliability of probabilities of survival, the precision of average population growth rates and the initial row vector. The more accurate the data, the more trustworthy the projections would be.

Disparity index of projected population age distributions during 1981-2021 and last three population censuses were calculated and are given in Table 2. Disparity index of projected population age distributions indicated the medium level of concentration within the age distribution of projected population. Moreover, these coefficients do not specify the lower and higher side of more concentration. The Gini coefficient of 1998 population census is exactly the same as that of the projected age distribution of 2001. It indicates that if growth rates of 1998 population census as well as other associated social factors and cultural values continue; the population distribution pattern would not be changed drastically and vice versa.

The projected population percentages of significant age segments and population inequality measure were also computed (see Table 3). It has been examined that approximately 14.80%, 28.60%, 53.10% and 3.50% population were of ages 0-4, 5-14, 15-64 and 64+ respectively in 1998 population census. The corresponding projected population percentages for the year 2023 are 10.12%, 19.86%, 65.01% and 5.00% respectively. It indicates the decreasing trend of growth rate of population of Pakistan. According to the Total Population by Country (2014), Pakistan was ranked at 46^{th} position in the world with that of about 37.20% population of age 0-14 as compared to approximately 43% in 1998. While working age group (15-64) was reported to about 58.60% and ranked at 98^{th} position. It implies that there are 97 countries in the world that have healthier and bigger strata of work force, as compare to Pakistan. On the other hand, Pakistan was ranked at 94^{th} position according to the aged population percentage i.e. 4.20% whereas in this article, the projected aged population percentage will be about 5% during the year 2023. The results indicate the decrease in babies and adolescent population whereas the increase in working population (15-64) as well as ageing population. It is expected that in future there would be healthier and vigorous population in Pakistan. Ryan and Willits (2007) also reported the increase in ageing population of the United States. In 1900, the percentage of elderly population (75+) was 29% which is being expected about 56% in 2040. Our results are in agreement with this study. It indicates that in near future there would be an increase of ageing population in Pakistan. The inequality measure also indicates the larger disparity during 2013-23 as compared to the 1998 disparity index.

Lorenz curves of projected age distribution for the years 2011, 2021 were drawn and indicate the moderate level of concentration (see Figures 1-2) whereas; the comparison of population age distribution percentages during the years 1998-2023 are presented in Figure 3. Keeping in view the current scenario of population, the population of Pakistan will increase during the next decade with almost similar pattern. The population pyramids

for the years 2018 and 2023 are also compared with that of 1998 (see Figures 4-5). The decrease in younger population might be because of increased literacy rate while the increase in ageing population and health related facilities will be because of improved medication.

4. CONCLUSION

The proposed Markov Chain model was found to be parsimonious to project the population age distribution up to two to three decades. The degree of the parsimonious model depends on the accuracy of one step transition probability matrix and initial row vector. The main characteristic of the proposed MC model is its simplicity in which deaths and new born baby births are adjusted in one state (source/sink state) without violating the condition of row sum one of stochastic process. The projection made by MC model also remained consistent with that of the overall projected population of Pakistan by other organization (Total Population by Country, 2014; Country Comparison Population, 2014; IDB, 2014; NIPS, 2014; Finance Division, 2012-13; WPP, 2012).The Gini coefficient and Lorenz curve also provided incredibly important information regarding the inequality of projected as well as census populations at different age segments. The disparity index of the projected population demonstrated high level of concentration describing the increase in ageing population. The findings also indicated the decrease in new born babies (0-4) and adolescent population (5-14). On the other hand, the results pointed out the increase in working and ageing population. The disparity index of age distribution of last three census populations indicated the medium level of concentration.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous referees for a very careful reading on an earlier version of this paper that led to many improvements. The authors are also thankful to the Pakistan Bureau of Statistics (PBS), Islamabad, Pakistan for providing the population data to complete this research work.

REFERENCES

1. Brown, M. (1994). Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on the Alberta data. *Social Science and Medicine*, 38(9), 1243-1256.
2. Country Comparison Population (2014). Country Comparison Population. Retrieved from http://www.indexmundi.com/g/r.aspx?c=pk&v=24/
3. Cowell, F.A. (2011). *Measuring Inequality.*3rd ed. Oxford University Press Inc., New York.
4. Finance Division (2013). *Pakistan Economic Survey 2012-13.*, Economic Adviser's Wing, Islamabad, Pakistan.
5. Fujiwara, M. and Caswell, H. (2002). Estimating population projection matrices from multi stage mark-recapture data. *Ecology*, 83(12), 3257-3265.
6. Gini, C. (1912). *Variabilità e mutabilità.* Contributo allo Studio delle Distribuzionie dello Relazioni Statistiche. Bologna.

7. Government of Pakistan (1972). *Statistical Report of Pakistan.* Population Census Organization, Statistics Division, Islamabad, Pakistan.
8. Government of Pakistan (1984). *1981 Census report of Pakistan*. Population Census Organization, Statistics Division, Islamabad, Pakistan.
9. Government of Pakistan (2001). *1998 Census report of Pakistan*. Population Census Organization, Statistics Division, Islamabad, Pakistan.
10. International Data Base (2014). International Data Base (IDB), information gateway, US Census Bureau. Retrieved from http://www.census.gov/population/international/data/idb/.
11. Jan, B., Ishfaq, A. and Shuhrat, S. (2007). Selecting of mathematical model for projections of NWFP population. *The Journal of Humanities and Social Sciences*, XV(2), 69-78.
12. Keyfitz, N. (1964). The population projection as a matrix operator. *Demography*, 1(1), 56-73.
13. Klosterman, R.E. (1990). *Community analysis and planning techniques.* Lanham, MD: Rowman & Lettlefield.
14. Loh, S. and George, M.V. (2007). Projected population size and age structure for Canada and provinces: With and without international migration. *Canadian Studies in Population*, 34(2), 103-127.
15. National Institute of Population Studies. (2014). National Institute of Population Studies. Retrieved from http://www.nips.org.pk.
16. Nichols, J.D., Sauer, J.R., Pollock, K.H. and Hestbeck, J.B. (1992). Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, 73(1), 306-312.
17. Pan American Health Organization (2001). Measuring Health Inequalities: Gini coefficient and concentration index. *Epidemiological Bulletin of PAHO*, 22(1), 3-4.
18. Rodrigue, J.P., Comtois, C. and Slack, B. (2013). *The Geography of Transport Systems. 3rd ed.*, Routledge. New York.
19. Ross, M. (2008). *Introduction to Probability Models*. Academic press, USA.
20. Ryan, A. K. and Willits, F. K. (2007). Family ties, physical health, and psychological well-being. *Journal of Aging and Health*, 19(6), 907-920.
21. Siegel, J.S. and Swanson, D.A. (2004). *The Methods and Materials of Demography.* Elsevier Academic Press. California, USA.
22. Total Population by Country (2014). Total Population by Country. Retrieved from http://www.globalfirepower.com/total-population-by-country.asp.
23. *World Population Prospects* (2012). *World Population Prospects: The 2012 Revision*, Department of Economics & Social Affairs, Population Division, Population Estimates & Projection Section. United Nation. Retrieved from http://esa.un.org/wpp/unpp/p2k0data.asp.
24. Zakria, M. and Muhammad, F. (2009). Forecasting the population Pakistan using ARIMA models. *Pak. J. Agri. Sci.*, 46(3), 214-223.
25. Zakria, M., Muhammad, F. and Salah-ud-din (2009). Population projection of Pakistan using traditional and time series models. *The Journal of Humanities and Social Sciences*, XVII(2), 88-100.

Table 1
Projected Population by Markov Chain Model during1981-2021

Age Group	1981	1991	2001	2011	2021
00-09	24,923,736	32,555,598	41,842,713	54,320,283	70,472,116
10-19	18,565,770	23,675,057	30,924,562	39,746,393	51,598,836
20-29	11,706,620	16,504,969	21,047,126	27,491,936	35,334,543
30-39	8,814,787	11,205,577	15,798,557	20,146,309	26,315,281
40-49	6,941,179	8,393,440	10,669,951	15,043,386	19,183,315
50-59	4,937,582	6,358,120	7,688,391	9,773,675	13,779,741
60-69	2,968,989	4,192,007	5,398,044	6,527,444	8,297,850
70+	2,530,332	1,722,013	2,431,364	3,130,865	3,785,918
Total*	83,587,994	107,516,466	135,800,706	176,180,289	228,767,600

* includes the projected population of FATA, all entries are to the nearest integer.

Table 2
Disparity Index of Projected Population Age Distribution and Census Populations

Projection Year	Disparity Index		Census Year	Disparity Index	
	Gini Coefficient	95% Confidence Interval		Gini Coefficient	95% Confidence Interval
1981	0.4509	(0.3770, 0.5475)	**1972**	0.4504	(0.3895,0.5512)
1991	0.4732	(0.4064, 0.6170)	**1981**	0.4552	(0.4137, 0.5408)
2001	0.4720	(0.4011, 0.6293)	**1998**	0.4720	(0.3957, 0.5927)
2011	0.4725	(0.4055, 0.6166)			
2021	0.4715	(0.3931, 0.6195)			

Table 3
Projected Population (%) and Disparity Index

Projection Year	Projected Population (%)				Disparity Index	
	0-4	5-14	15-64	65+	Gini Coefficient	95% Confidence Interval
1998	14.80	28.60	53.10	3.50	0.5418	(0.4068,0.7799)
2013	11.67	21.56	62.50	4.24	0.6158	(0.5337,0.7745)
2018	10.93	20.79	63.76	4.52	0.6253	(0.5472,0.7662)
2023	10.12	19.86	65.01	5.00	0.6326	(0.5531,0.7500)

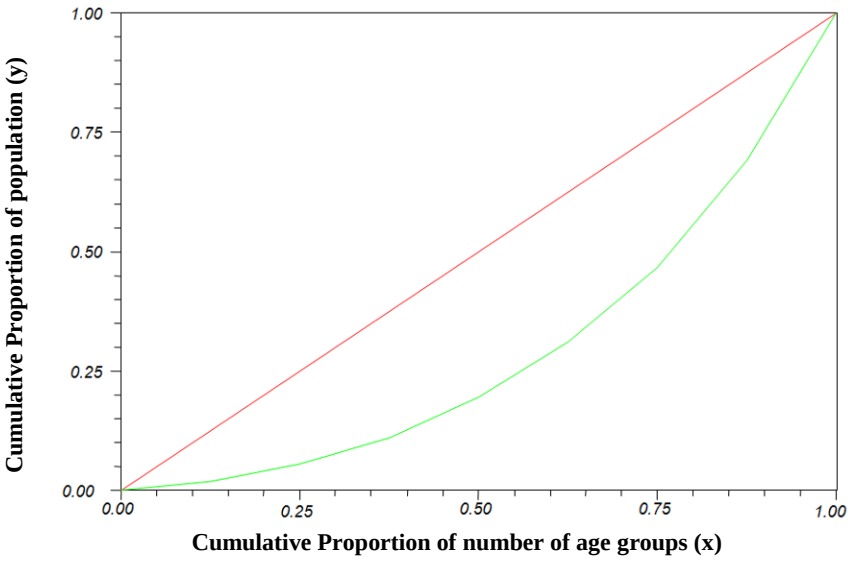

Fig. 1: Lorenz Curve of Projected Population 2011

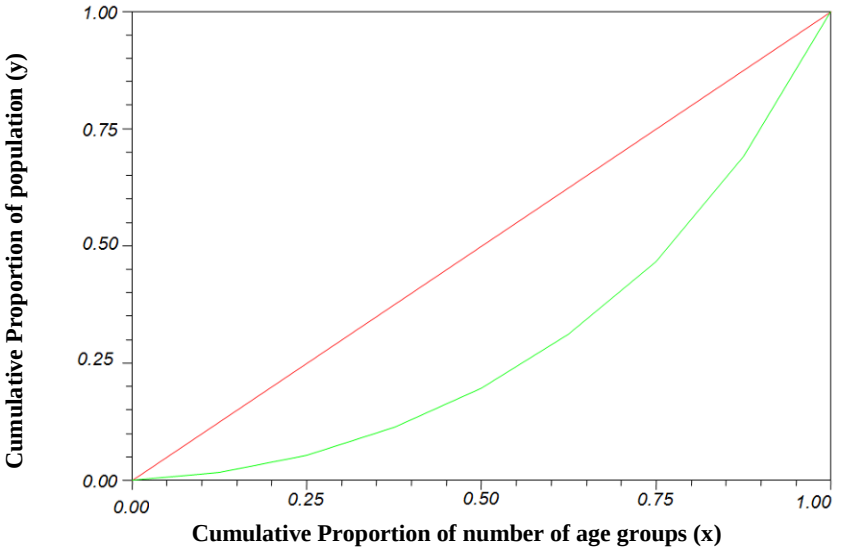

Fig. 2: Lorenz Curve of Projected Population 2021

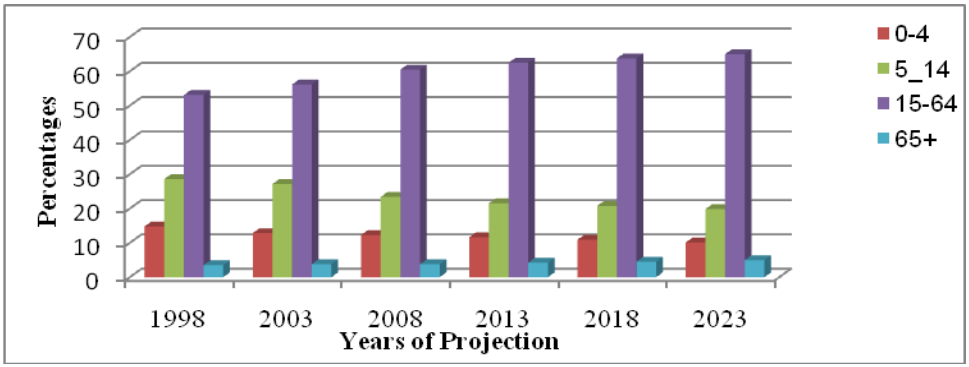

Fig. 3: Comparison of Age Distribution during1998-2023

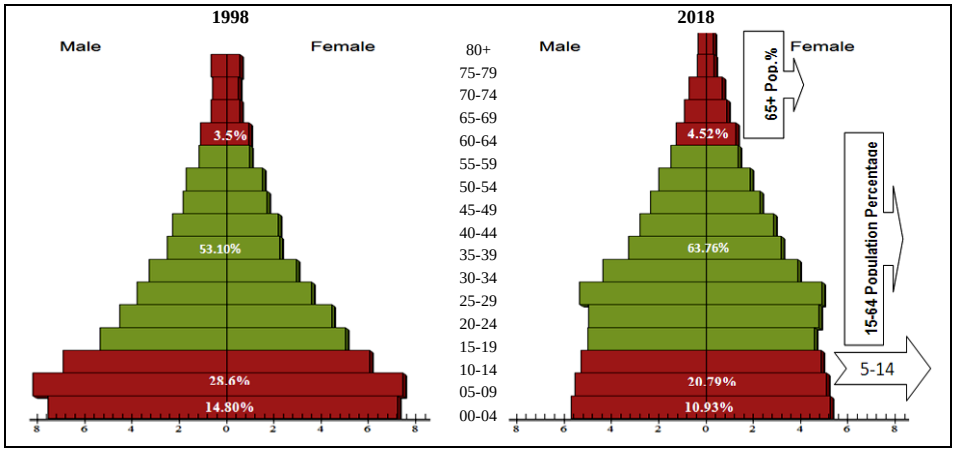

Fig. 4: Comparison of Population Pyramids of 1998 & 2018

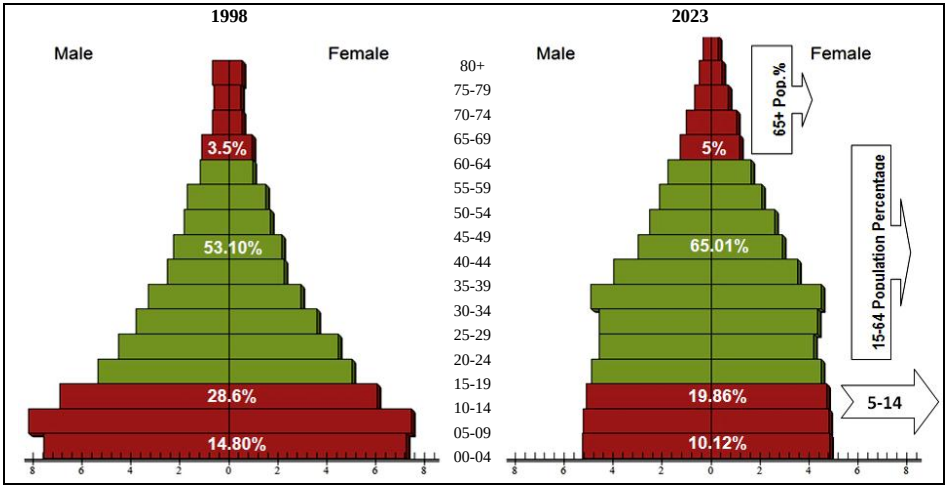

Fig. 5: Comparison of Population Pyramids of 1998 & 2023