# Age Distribution Model For GCC Countries

Shafiqur Rahman, Sultan Qaboos University, Sultanate of Oman

**ABSTRACT**

*Efficient and reliable estimates of the proportions of population at different age levels are essential for making quality budget of any developing or developed nation. These estimates are obtained from the best-fitted age distribution model and can be used to find the number of school age children, number of pensioners etc. Past population census data of GCC countries are analyzed to find the best-fitted age distribution model applying chi-square goodness of fit test and model selection criteria and observed that the age distribution of most of the GCC countries is exponential. A comparative study of the age distributions of six GCC countries with some developed countries is also provided.*

**Keywords:**  Probability distribution, Chi-square goodness of fit test, and Model selection.

## 1.      INTRODUCTION

$\mathcal{K}$ ingdom of Bahrain (Bahrain), State of Kuwait (Kuwait), Sultanate of Oman (Oman), State of Qatar (Qatar), Kingdom of Saudi Arabia (KSA), and United Arab Emirates (UAE) are the members of Gulf Cooperative Council (GCC). Populations of GCC countries are growing day by day. Every year governments of GCC countries need to make their yearly budget. Quality budget for a particular year depends on the accurate estimates of the proportions of population figures at different age groups of that year. The actual population figures at different age levels can only be obtained through census or complete enumeration. A population census is the single most extensive, complicated and expensive statistical operation that a country undertakes. Considering all these, like most other countries in the world, population census in GCC countries were conducted roughly once in every ten years. Therefore, it is almost impossible to know the actual population figures and the proportion of people at different age levels for every year. As a result, efficient and reliable estimates of the exact population figures at different age levels are required for making a quality budget. Efficient estimate of the exact population figure at different age levels can be obtained from the best-fitted probability distribution model. Unemployment rate, proportion of school age children, proportion of pensioners, economic activities and marital status all vary by age. Social relationships within a community are affected by relative ages. Population composition, social and economic characteristics vary with age levels. Therefore it is important to know population figures for different age groups.

The objective of this study is to develop the most efficient and reliable probability distribution model for GCC countries that can be used to project the proportion of population at different age groups.

## 2.      MATERIALS AND METHODS

Past population census data of six GCC countries were collected and presented in both tabular and graphical forms. To fit a suitable probability distribution for ages of the population of six GCC countries we calculated the means and variances for all countries and observed that variances are greater than their respective means in all cases. Age is a continuous variable. Three continuous probability distributions named gamma, exponential and Rayleigh distributions are considered as their variances are greater than their respective means. A brief description of the three distributions is given below.

**Gamma distribution**:

The probability density function (pdf) of gamma distribution is $f_g(x) = \dfrac{e^{-x/\lambda}\, x^{\nu-1}}{\lambda^{\nu}\,\Gamma(\nu)}, x > 0.$

First raw moment or mean of the distribution is $E(X) = \int_0^\infty x f_g(x)\,dx = \nu\lambda$ .

Second raw moment of the distribution is $E(X^2) = \int_0^\infty x^2 f_g(x)\,dx = \nu(\nu+1)\lambda^2$ .

Variance of the distribution is $V(X) = \nu\lambda^2$ .

Evidently the variance is greater than mean for all $\lambda > 1$. In order to fit gamma distribution it is necessary to estimate the parameters $\nu$ and $\lambda$. Estimators of $\nu$ and $\lambda$ obtained by the method of moments are as follows:

$$\hat{\lambda} = \frac{s^2}{\overline{x}} \text{ and } \hat{\nu} = \frac{\overline{x}^2}{s^2},$$

where $\overline{x}$ and $s^2$ are the sample mean and sample variance respectively.

**Exponential distribution**: The pdf of exponential distribution is $f_e(x) = \dfrac{e^{-x/\mu}}{\mu}$ , $x > 0$.

First raw moment or mean of the distribution is $E(X) = \int_0^\infty x f_e(x)\,dx = \mu$ .

Second raw moment of the distribution is $E(X^2) = \int_0^\infty x^2 f_e(x)\,dx = 2\mu^2$ .

Variance of the distribution is $V(X) = \mu^2$ .

Evidently the variance is greater than mean for all $\mu > 1$. In order to fit exponential distribution it is necessary to estimate the parameter $\mu$. Estimator of $\mu$ obtained by the method of moments or maximum likelihood is $\hat{\mu} = \overline{x}$ .

**Rayleigh distribution**: The pdf of Rayleigh distribution is $f_r(x) = \dfrac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}$ , $x > 0$.

First raw moment or mean of the distribution is $E(X) = \int_0^\infty x f_r(x)\,dx = \theta\sqrt{\pi/2}$ .

Second raw moment of the distribution is $E(X^2) = \int_0^\infty x^2 f_e(x)\,dx = 2\theta^2$ .

Variance of the distribution is $V(X) = \left(\dfrac{4-\pi}{2}\right)\theta^2$ .

Evidently the variance is greater than mean for all $\theta > 2.92$. In order to fit Rayleigh distribution it is necessary to estimate the parameter $\theta$. Estimator of $\theta$ obtained by the method of moments is $\hat{\theta} = \overline{x}\sqrt{(2/\pi)}$ .

The above three distributions are fitted for each country and tested by Chi-squares goodness of fit tests. The test results suggest that the age distributions of KSA, UAE, and Oman follow exponential distribution. Eight model selection criteria named as Akaike's (1973) information criterion (*AIC*), Schwartz's (1978) Bayesian information criterion (*BIC*), Rahman and King's (1999) joint information criterion (JIC), Theil's (1961) $\overline{R}^2$ criterion, Craven and Wahba's (1979) generalized cross validation (*GCV*) criterion, Hannan and Quinn's (1979) criterion (*HQC*), Hocking's (1976) $S_p$ criterion, and Mallows (1964) $C_p$ criterion are also applied. These eight criteria expressed in the following penalized residual sum of squares form are obtained from Rahman and Nahar (2004).

$$AIC \approx E_j^2\, e^{\frac{2k_j}{n}}, \quad BIC \approx E_j^2\, n^{\frac{k_j}{n}}, \quad JIC \approx \frac{E_j^2 (n)^{k_j/n}}{\sqrt{(n-k_j)}}, \quad \overline{R}^2 \approx \frac{E_j^2}{n-k_j},$$

$$GCV = \frac{E_j^2}{\left(1-\dfrac{k_j}{n}\right)^2}, \quad HQC = E_j^2 (\ln n)^{2k_j/n}, \quad S_p = \frac{E_j^2}{(n-k_j)(n-k_j-1)}, \quad C_p = \frac{(n+k_j)E_j^2}{n-k_j},$$

where $E_j^2$ is the residual sum of squares, $k_j$ is the number of parameters and *n* is the total frequencies. The results obtained by applying the above information criteria are similar to that of Chi-squares goodness of fit tests.

## 3. RESULTS AND DISCUSSION

The past percentage population data of six GCC countries by age groups were presented in Table 1.

**Table 1:  Percentages of Population of Six GCC Countries at Different Age Groups**

| Age group | Bahrain | KSA | Kuwait | Oman | Qatar | UAE |
|---|---|---|---|---|---|---|
| 0 --4 | 10.26 | 14.52 | 10.93 | 16.89 | 8.18 | 13.00 |
| 5--9 | 10.23 | 13.17 | 9.38 | 13.62 | 8.77 | 12.52 |
| 10--14 | 9.49 | 11.36 | 9.04 | 10.63 | 9.33 | 12.63 |
| 15--19 | 7.89 | 9.44 | 9.01 | 9.71 | 8.21 | 12.95 |
| 20--24 | 7.42 | 9.59 | 9.69 | 8.38 | 7.24 | 12.59 |
| 25--29 | 8.06 | 11.10 | 9.50 | 6.19 | 7.12 | 9.77 |
| 30--34 | 9.17 | 10.36 | 9.28 | 6.67 | 7.44 | 6.30 |
| 35--39 | 9.79 | 7.28 | 8.77 | 7.36 | 8.99 | 5.04 |
| 40--44 | 9.54 | 4.03 | 7.75 | 6.50 | 10.69 | 3.76 |
| 45--49 | 7.29 | 2.49 | 5.92 | 4.90 | 9.31 | 3.19 |
| 50--54 | 3.99 | 1.74 | 4.16 | 3.23 | 6.31 | 2.45 |
| 55--59 | 2.31 | 1.42 | 2.53 | 2.14 | 3.80 | 1.72 |
| 60--64 | 1.70 | 1.17 | 1.69 | 1.43 | 2.31 | 1.30 |
| 65--69 | 1.22 | 0.88 | 1.12 | 0.95 | 1.25 | 1.04 |
| 70--74 | 0.83 | 0.64 | 0.66 | 0.63 | 0.58 | 0.77 |
| 75--79 | 0.48 | 0.40 | 0.33 | 0.41 | 0.30 | 0.82 |
| 80+ | 0.34 | 0.39 | 0.20 | 0.37 | 0.16 | 0.16 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 |

**Figure 1: Graph Showing Percentage Distribution by Age Group of GCC Countries**
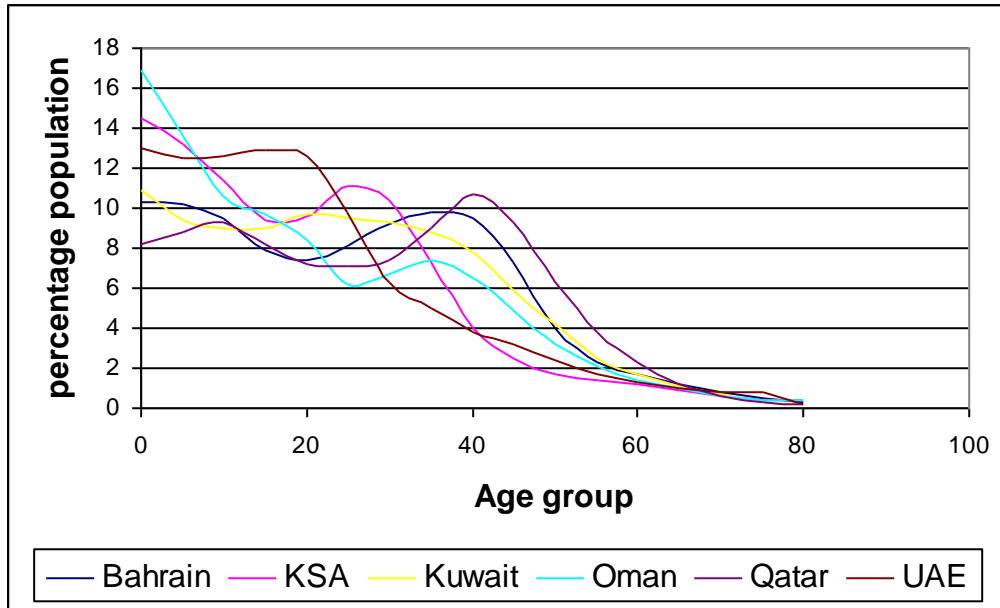


Figure 1 is the graphical representation of the data in Table 1. From the graph it is evident that there are differences between the proportions of population for different census years corresponding to any particular age groups.

In order to fit a suitable distribution first we have calculated the mean and variance of the observed distributions for each country and presented in Table 2.

**Table 2: Means and Variances of Observed Distributions for GCC Countries**

| Country | Mean | Variance |
|---------|------|----------|
| Bahrain | 27.84 | 341.55 |
| KSA | 22.49 | 286.85 |
| Kuwait | 26.86 | 318.55 |
| Oman | 23.28 | 348.76 |
| Qatar | 30.06 | 348.73 |
| UAE | 22.67 | 305.23 |

It is obvious that the variance is greater than the mean for all census years. We consider three continuous distributions gamma, exponential and Rayleigh, which have their variances greater than their respective means. We fit these three distributions for all census years and tested by the Chi-square goodness of fit test. The following test statistic is used

$$\chi^2 = \sum_{i=1}^{c}\left(O_i - E_i\right)^2 / E_i$$

where, $O_i$ and $E_i$ are the observed and expected percentages. Test results are presented in the following Tables 3.

**Tables 3: Chi-Square Test Results for GCC Countries**

| | Gamma | | Exponential | | Rayleigh | |
|---|---|---|---|---|---|---|
| **Country** | $\chi^2$ | **P-value** | $\chi^2$ | **P-value** | $\chi^2$ | **P-value** |
| Bahrain | 427.764 | 0 | 25.611 | 0.0012 | 45.125 | 0 |
| KSA | 202.480 | 0 | 13.186 | 0.1056 | 58.849 | 0 |
| Kuwait | 139.833 | 0 | 18.176 | 0.0199 | 43.773 | 0 |
| Oman | 430.137 | 0 | 7.693 | 0.4640 | 96.572 | 0 |
| Qatar | 136.028 | 0 | 38.548 | 0 | 34.856 | 0 |
| UAE | 212.692 | 0 | 7.518 | 0.4820 | 52.717 | 0 |

From the chi-square test we find that the population distributions of Kingdom of Saudi Arabia (KSA), Sultanate of Oman (Oman), and United Arab Emirates (UAE) may follow the exponential distribution. A comparative study of the age distributions of six GCC countries with some developed countries is given in Table 4.

**Tables 4: Comparative Study of the Age Distributions of Six GCC Countries**

| Country | Age (0-14) | Age(15-64) | Age (65+ ) | Median Age | Fertility rate | Growth rate | Literacy rate |
|---|---|---|---|---|---|---|---|
| Bahrain | 29.98 | 67.15 | 2.87 | 29.90 | 2.39 | 2.17 | 89.10 |
| KSA | 39.05 | 58.63 | 2.32 | 21.50 | 3.94 | 2.06 | 78.70 |
| Kuwait | 29.36 | 68.32 | 2.32 | 26.00 | 4.46 | 3.59 | 91.00 |
| Oman | 41.13 | 56.50 | 2.37 | 18.90 | 3.55 | 2.93 | 71.90 |
| Qatar | 26.28 | 71.42 | 2.30 | 30.80 | 2.94 | 2.28 | 88.60 |
| UAE | 23.28 | 75.93 | 0.79 | 27.90 | 1.74 | 1.60 | 77.90 |
| Australia | 19.60 | 67.30 | 13.10 | 37.30 | 1.78 | 1.20 | 99.00 |
| UK | 17.90 | 66.10 | 16.00 | 40.20 | 1.74 | 0.28 | 99.00 |
| Japan | 14.00 | 66.30 | 19.70 | 44.20 | 1.30 | 0.15 | 99.00 |
| Canada | 17.60 | 69.30 | 13.10 | 40.40 | 1.58 | 0.82 | 99.00 |
| USA | 20.80 | 66.90 | 12.30 | 36.70 | 2.00 | 1.30 | 99.00 |

**CONCLUSION**

There are no significant differences between the proportions of population for different GCC countries corresponding to any particular age group. The age distributions of KSA, Oman and UAE population follow exponential distribution. The median age of some GCC population is about half of that of other developed nations. Except UAE all other GCC countries have replacement level of fertility. GCC governments should give more importance on education sector as the literacy rate is low compared to developed nations. Proportions of male in all GCC countries are higher than that of female population where as in developed countries it is opposite. Ageing is not a problem for GCC countries but it is a big issue for most of developed countries. Young populations in all GCC countries are significantly higher than that of developed countries.

**AUTHOR INFORMATION**

**Dr. Shafiqur Rahman** was born and brought up in Bangladesh and then migrated to Australia. He has more than 28 years of teaching and research experience at six Universities named Chittagong University of Bangladesh, Dalhousie and Saint Mary's University of Canada, Monash University of Australia, University of Papua New Guinea and currently Sultan Qaboos University of Oman. He taught various Statistics courses at undergraduate and postgraduate levels. He has published 27 research papers in international referred journals. He is a member of several professional societies. Currently he is an Associate Editor of the Journal of Applied Probability and Statistics.

**REFERENCES**

1. Akaike, H. (1973), 'Information theory and an extension of the maximum likelihood principle', Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Csaki, Akademial Kiado, Budapest, 267-281.
2. Craven, P. and Wahba, G. (1979). 'Smoothing Noisy data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized   Cross Validation', *Numerische Mathematik*, 31, 377-403.
3. Hannan, E.J. and Quinn, B.G. (1979), 'The determination of the order of an auto-regression', *Journal of the Royal Statistical Society, Series B,* 41, 190-195.
4. Hocking, R.R. (1976), 'The analysis and selection of variables in linear regression',*Biometrics,* 32, 1-49.
5. Mallows, C.L. (1964), 'Choosing Variables in a Linear Regression: A graphical Aid',presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansas (May).
6. Rahman, M.S. and King, M.L.(1999), 'Improved Model Selection Criterion', Journal of *Communications in Statistics*, *Simulation and Computation*, 28(1), 51-71.
7. Rahman, M. S. and Nahar, S. (2004). "Generalized Model selection criterion." Far East Journal of Theoretical Statistics, 12(2), 117-147.
8. Schwartz, G. (1978). 'Estimating the dimension of a model', *The Annals of Statistics*, 6, 461-464.
9. Theil, H. (1961). *Economic Forecasts and Policy*, 2nd edition, North-Holland, Amsterdam.