Population Distribution Model For Oman

Shafiqur Rahman, Sultan Qaboos University, Sultanate of Oman

ABSTRACT

Past population census data of Sultanate of Oman are analyzed to find the best-fitted age distribution model applying chi-square goodness of fit test and model selection criteria. It is observed that the age distribution of the Omani population is exponential. The population figures for different age groups of Oman are estimated using exponential distribution. Age distribution of Omani population is compared with that of other Gulf countries and also with some developed nations. It is observed that, unlike other developed countries of the world, the age distribution of Omani population does not change significantly over the last two decades. It is also observed that the median age of the Omani population is about half of that of other developed nations. Ageing is not a problem for Oman or Gulf countries, but it is a big issue for most developed countries. Young populations in Oman are significantly higher than that of developed countries.

Keywords: Probability distribution; Chi-square goodness of fit test; and Model selection

INTRODUCTION

The population of the Sultanate of Oman is growing day by day. Every year the Omani government needs to make its yearly budget. Quality budget for a particular year depends on the accurate estimates of the population or proportions of population figures at different age groups for that year. The actual population figures at different age levels can only be obtained through a 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, Omani population censuses were conducted in 1993 and 2003; that is, once 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 estimates of the exact population figure at different age levels can be obtained from the best-fitted probability distribution model. Rahman et al (2005) proposed an efficient model for estimating the total yearly population of Papua, New Guinea (PNG) and Rahman and Nahar (2006) proposed an efficient age distribution model for PNG population. 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, as well as social and economic characteristics, vary with age levels. Therefore, it is important to know the population figures, or proportion of population figures, for different age groups.

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

MATERIALS AND METHODS

The first countrywide official census was conducted in 1993, then in 2003. The past two population census data were collected from the Ministry of National Economy of Oman. To fit a suitable probability distribution for ages of Omani population, we calculated the means and variances for two census years and observed that variances are greater than their respective means in both cases. Age is a continuous variable. Three continuous probability distributions - gamma, exponential and Rayleigh - are considered since their variances are greater than their respective means. Brief descriptions of the three distributions and the estimation of their parameters are given below.

Gamma Distribution (GD)

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 ν and λ. Estimators of ν and λ obtained by the method of moments are $\hat{\lambda} = \dfrac{s^2}{\overline{x}}$ and $\hat{\nu} = \dfrac{\overline{x}^2}{s^2}$ where $\overline{x}$ and s^2 are the sample mean and sample variance, respectively.

Exponential Distribution (ED)

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 $E(X) = \int_0^{\infty} x\, f_e(x)\, dx = \mu$.

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

Evidently, the variance is greater than mean for all $\mu > 1$.

Estimator of μ obtained by the method of moments or maximum likelihood is $\hat{\mu} = \overline{x}$.

Rayleigh distribution (RD)

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 $\theta > 2.92$.

Estimator of θ obtained by the method of moments is $\hat{\theta} = \overline{x}\sqrt{\left(2/\pi\right)}$.

The above three distributions are fitted for each census year and tested by Chi-squares goodness-of-fit tests. The test results suggest that the ages of Omani population follow the exponential distribution. We also apply model selection criteria to select the best one out of these three distributions. Model selection criteria also selects the exponential distribution for the ages of Omani population.

RESULTS AND DISCUSSION

The past population census data of Oman by age groups are presented in Table 1.

Table 1: Percentages of Omani Population at Different Age Groups

Age Group	1993	2003
0-4	17.24	12.10
5-9	18.12	13.33
10-14	16.24	15.17
15-19	11.82	14.33
20-24	7.45	12.03
25-29	5.45	8.32
30-34	4.39	5.21
35-39	4.09	4.10
40-44	3.16	3.42
45-49	2.85	2.84
50-54	2.74	2.40
55-59	1.65	1.75
60-64	1.83	1.80
65-69	0.86	1.03
70-74	0.96	0.95
75+	1.16	1.21
Total	100	100

Figure 1: Graph showing Percentage Distribution by Age Group of Oman

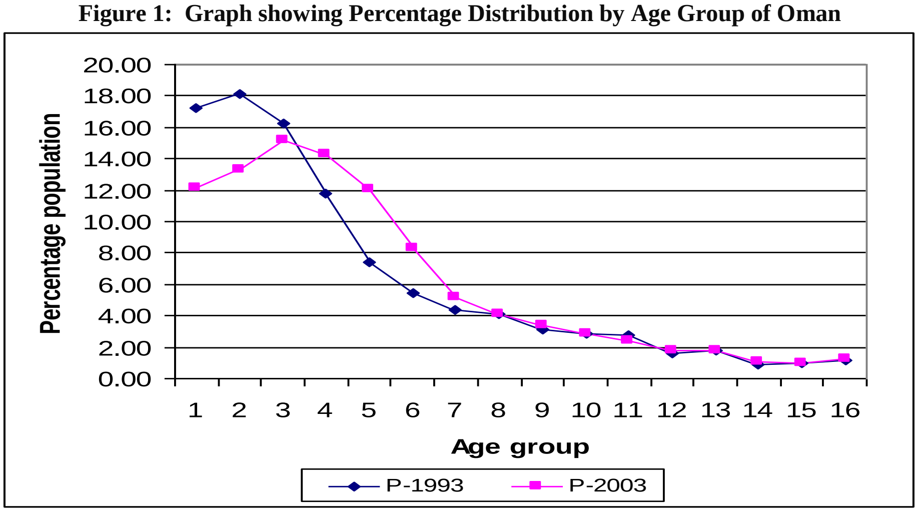

Figure 1 is the graphical representation of the data given in Table 1. From the graph, it is evident that there are some differences between the proportions of population for two census years corresponding to any particular age group. We have applied Wilcoxon's signed rank test to check whether these differences are statistically significant. The test results are given in Table 2.

Table 2: Results of Wilcoxon's Signed Rank Test

Age Group	Difference	Absolute Difference	Rank	Signed Rank
0-4	-5.14	5.14	16	-16
5-9	-4.79	4.79	15	-15
10-14	-1.07	1.07	11	-11
15-19	2.51	2.51	12	12
20-24	4.58	4.58	14	14
25-29	2.87	2.87	13	13
30-34	0.82	0.82	10	10
35-39	0.01	0.01	2	2
40-44	0.26	0.26	8	8
45-49	-0.01	0.01	2	-2
50-54	-0.34	0.34	9	-9
55-59	0.10	0.10	6	6
60-64	-0.03	0.03	4	-4
65-69	0.17	0.17	7	7
70-74	-0.01	0.01	2	-2
75+	0.05	0.05	5	5

In this test, we first find the differences between two percentages corresponding to all age groups and rank the absolute values of the differences. The equal differences are assigned the average ranking of their positions in the combined data set.

Test Statistic

T^+ = Sum of the ranks of positive differences = 77.

T^- = Sum of the ranks of negative differences = 59.

Decision Rule

We reject H_0 if T^+ or, $T^- \leq$ critical value. The critical values of T, obtained from Conover (1980) for n = 16 and significance levels α = 10% and 5%, are 43 and 36. As the values of T^+ or T^- are much higher than the critical value, we do not reject H_0 and conclude that there is no significant difference between the age distributions of the two census years.

In order to fit a suitable distribution we first calculated the mean and variance of the observed distributions for each census year and then estimated the parameters of all three distributions. The results are presented in Table 3.

Table 3: Estimation of Parameters of GD, ED and RD for Past Census Data

Census Year	Mean	Variance	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\nu}$	$\hat{\theta}$
1993	19.93	323.88	19.93	16.25	1.23	15.90
2003	21.96	301.82	21.96	13.74	1.60	17.52

We fit exponential, gamma and Rayleigh 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 Table 4.

Table 4: Summary Output of Census Data for Fitting GD, ED and RD

Census Year	Distribution	χ^2	P-value
1993	Gamma	130.721	0
1993	Exponential	1.528	0.981
1993	Rayleigh	46.018	0
2003	Gamma	58.467	0
2003	Exponential	6.974	0.539
2003	Rayleigh	23.875	0.001

From the chi-square test, we find that the population distributions of Oman by age group may follow exponential distribution.

We also apply 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) $\bar{R}^2$ 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 - to select the better model. These eight criteria, expressed in the following penalized residual sum of squares form, are obtained from Rahman and Nahar (2004).

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

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

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 criteria are given in Table 5.

Table 5: Penalized Residual SS for GD, ED and RD

	1993			2003		
	GD	**ED**	**RD**	**GD**	**ED**	**RD**
AIC	2048.93	60.39	530.95	1389.82	150.71	241.40
BIC	2338.29	64.52	567.21	1586.10	161.00	257.89
JIC	236.20	6.48	57.01	160.22	16.18	25.92
$\bar{R}^2$	21.76	0.62	5.47	14.76	1.55	2.49
GCV	2220.48	62.86	552.68	1506.18	156.87	251.28
HQC	2266.88	63.52	558.48	1537.66	158.52	253.92
Sp	0.22	0.01	0.06	0.15	0.02	0.03
Cp	2219.59	62.86	552.62	1505.58	156.86	251.26

Rahman and Nahar (2004) suggested that a model with minimum penalized residual sum of squares would be the best model. From Table 5, it is obvious that the penalised residual SS for the exponential distribution is lower than that of gamma and Rayleigh distributions. Therefore, we conclude that the ages of Omani people may follow the exponential distribution $f_e(x) = \frac{e^{-x/\mu}}{\mu}$, where μ is a parameter and x is the age level. Using the exponential distribution, we estimated the population for school age group "5 – 20" and that for age "above 50" and presented it in Table 6.

Table 6: Population for Specific Age Groups for the Year 2008

Year	Estimated Population	Population between 5-20	Population above 50
2008	2734347	1123816	262531

COMPARISON WITH OTHER DEVELOPED NATIONS

Population ageing is a major characteristic of most developed countries, like Australia, Canada, France, Greece, Italy, Japan, UK, and USA. Table 7 represents the data obtained from the latest issue of the World Fact Book.

Table 7: Population Age Structure of Oman and Some Nations (July 2009 Estimate)

	Aged 0-14 Years	Aged 15-64 Years	Aged 65 Years and Over	Median Age
Countries	**%**	**%**	**%**	**Years**
Oman	**42.7**	**54.5**	**2.8**	**18.8**
KSA	38.0	59.5	2.5	21.6
Kuwait	26.4	70.7	2.9	26.2
Bahrain	25.9	70.2	3.9	30.1
Australia	18.6	67.9	13.5	37.3
Canada	16.1	68.7	15.2	40.4
France	18.6	65.0	16.4	39.4
Greece	14.3	66.6	19.2	41.8
Italy	13.5	66.3	20.2	43.3
Japan	13.5	64.3	22.2	44.2
UK	16.7	67.1	16.2	40.2
USA	20.2	67.0	12.8	36.7

Evidently, in Greece, Italy, and Japan, the number of people aged 65 years and over already exceeds the number of children aged 0-14 years. Population ageing in developed countries is caused by sustained low fertility and increased life expectancy. However, ageing is not a problem in Oman or in other Gulf countries. It is also obvious from Table 7 that the median age of Omani population is about half of that of developed nations.

CONCLUSION

It is interesting to note that there are no significant differences between the proportions of population for the past two census years corresponding to any particular age group. That is unlike other developed countries of the world; the age distribution of Omani population did not change significantly over the last two decades. The median age of Omani population is about half of that of other developed nations. The age distribution of Omani population is exponential. Applying exponential distribution, population proportion or population for any age group can be estimated. Ageing is not a problem in Oman or in other Gulf countries. Young populations in Oman 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 holds a PhD in Statistics from Dalhousie University of Canada. His other qualifications include M.Phil. from Panjab University of

India, M.Sc. and B.Sc. Honours from Jahangirnagar University of Bangladesh and Graduate Certificate in Communication of Science and Technology from the University of Technology of PNG. He has more than 30 years of teaching 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 Sultan Qaboos University of Oman. He taught various Statistics and Mathematics 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. Conover, W.J. (1980), *Practical Nonparametric Statistics*, John Wiley and Sons, New York.
3. 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.
4. 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.
5. Hocking, R.R. (1976), 'The analysis and selection of variables in linear regression', *Biometrics*, 32, 1-49.
6. 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).
7. Rahman, M.S. and King, M.L.(1999), 'Improved Model Selection Criterion', Journal of *Communications in Statistics, Simulation and Computation*, 28(1), 51-71.
8. Rahman, M. S. and Nahar, S. (2004). "Generalized Model selection criterion." *Far East Journal of Theoretical Statistics*, 12(2), 117-147.
9. Rahman, M.S. and Nahar, S. (2006): Age distribution model for PNG population, *Journal of Applied Probability and Statistics*, 1(1), 39-48.
10. Rahman, M.S., Taminza, M. and Nahar, S. (2005): Population Projection Model for PNG, *Journal of Mathematics, Computing & Education*, 7(1), 30-40.
11. Schwartz, G. (1978). 'Estimating the dimension of a model', *The Annals of Statistics*, 6, 461-464.
12. Statistical year book 2008, Ministry of National Economy, Sultanate of Oman.
13. Theil, H. (1961). *Economic Forecasts and Policy*, 2nd edition, North-Holland, Amsterdam.
14. The World Fact Book (2009), https://www.cia.gov/library/publications/the-world-factbook/geos/mu.html.

NOTES