Sains Malaysiana 40(12)(2011): 1449–1453

Methods of Expanding an Abridged Life Tables: Comparison between Two Methods

(Kaedah Mengembangkan Jadual Hayat Ringkas: Perbandingan Antara Dua Kaedah)

ROSE IRNAWATY IBRAHIM* & ZAILAN SIRI

ABSTRACT

A question of interest in the demografic and actuarial fields is the estimation of the complete sets of q_x-values when the data are given in age groups. This study presents two techniques of expanding an abridged life table to a complete ones. The two expansion techniques used in the study are King's Osculatory Interpolation and Heligman-Pollard model. This work evaluated and compared the performance of King's Osculatory Interpolation and Heligman-Pollard model. For that purpose, empirical data sets on abridged life tables for Malaysian populations for the period of 1991, 1995 and 1999 for both gender were collected. Then each of the expanding techniques considered was applied to these abridged data sets. The complete sets of q_x-values obtained from these two techniques were then used to produce the estimated abridged ones. These results were then compared with the actual values published in the abridged life tables.

Keywords: Abridged life table; complete life table; Heligman-Pollard model; King's osculatory interpolation

ABSTRAK

Satu soalan yang diminati dalam demografik dan bidang aktuari ialah penganggaran set-set yang lengkap bagi nilai-nilai q_x apabila data yang diberikan adalah dalam kumpulan umur. Kajian ini membentangkan dua teknik mengembangkan jadual hayat ringkas kepada jadual hayat penuh iaitu interpolasi Beroskulasi King dan model Heligman-Pollard. Kajian ini menyediakan beberapa penilaian dan perbandingan bagi interpolasi Beroskulasi King dan model Heligman-Pollard. Untuk tujuan tersebut, set-set data empirik bagi jadual hayat ringkas untuk populasi Malaysia bagi tempoh 1991, 1995 dan 1999 untuk kedua-dua jantina dikumpulkan. Kemudian setiap teknik pengembangan yang dipertimbangkan diguna ke atas set-set data ringkas ini. Set-set yang lengkap bagi nilai-nilai q_x yang diperolehi kemudiannya diguna untuk menjana nilai-nilai anggaran dalam jadual hayat ringkas. Keputusan ini kemudiannya dibandingkan dengan nilai-nilai dalam jadual hayat ringkas sebenar yang diterbitkan.

Kata kunci: Interpolasi beroskulasi King; jadual hayat penuh; jadual hayat ringkas; model Heligman-Pollard.

INTRODUCTION

A life table is a statistical device used by actuaries, demographers, public health employees and many others to present the mortality experience of a population. It is also referred to as the mortality table. A life table in which the values of the life table functions l_x, q_x, and others are given in single years of age or presented for every integral values of x', which is for every completed age is called a complete life table. Thus, a life table is considered complete when it contains mortality information for each year of age between 0 and x (in our case $x = 100$); otherwise the life table is abridged. However, the abridged life tables are more often used, in which each age is presented in groups, which are 0-1, 1-5, 5-10, 10-15 and so on. The use of abridged life tables has expanded because mortality data are usually available and sufficiently accurate in the form of rates for 5-year age groups and not for each individual age. The main reason for providing data in an abridged form is related to the phenomenon of age heaping caused by age misstatements in data registration. Another reason is of the incomplete and unstable documentation of vital statistics and therefore the quality of the data may not permit computation of a complete life table. Thus, it often happens that only an abridged life table is available when a complete life table is needed. Hence it is important to construct a complete life table given the abridged one. Therefore, the objectives of the study were to evaluate and compare the performance of King's Osculatory Interpolation and Heligman-Pollard model techniques in terms of expanding an abridged life table to a complete ones.

Several methods have been suggested in the literature to construct a complete life table given the abridged life table. One of the methods that have been extensively used is the interpolation formula as suggested in King (1914). However, an interpolation method that is still being used is the six-point Lagrangian interpolation formula by Elandt-Johnson (1980) which is applied to the $l(x)$ values in an abridged life table. This method provides good approximations for adult mortality but it is less accurate for the early childhood ages. However the application of Lagrangian formula is limited to the ages less than 75 years while literature proposes other complementary methods for the ages greater than 75 years old.

A recent attempt to represent mortality over the course of the entire life span using a single analytical expression has been made in Australia by Heligman and Pollard (1980). Several applications of Heligman and Pollard model on a wide variety of mortality experiences have been used, such as in the United State of America by Mode and Busby (1982) and in Sweden by Hartmann (1987). Hartmann concluded that Heligman and Pollard model is the best existing demographic model of mortality at all ages and is an efficient means of generating life tables model, for example for use of population projection. The Heligman and Pollard model was also discussed by Kostaki (1991). He concluded that this model provides quite a satisfactory representation of the age pattern of mortality. He also concluded that this model provided a new way to expand a life table through direct estimation of the complete set of probabilities of dying, q_x having the abridged ${}_nq_x$ values at the starting point. In this research we evoluated and compared the performance of King's Osculatory Interpolation and Haligman-Polland model.

METHODS

This study focused on King's Osculatory interpolation and Heligman and Pollard model. We start with an abridged life tables for Malaysian populations for both gender in period 1991, 1995 and 1999. We then apply each of the methods to estimate a complete life tables from the abridged ones. From these estimated complete life tables, we construct an estimate abridged life tables and then compare the result with the real abridged data.

KING'S OSCULATORY INTERPOLATION

As the data in the Abridged Life Table is only available in age group such as 5-9, 10-14, 15-19 and so on, hence some distribution or interpolation method was essential to interpolate it to the individual ages. In the study, King's Osculatory Interpolation method will be used. King's Osculatory Interpolation was introduced by George King in year 1914. This method is only applicable for intervals of five years of age such as 10-14, 15-19 and so on. However, this method requires the pivotal value and osculatory interpolation to be computed. The pivotal value formula is as follows:

$$U_o = 1/5(W_o - 1/25\Delta^2 W_{-1}), \tag{1}$$

where U_o is the pivotal value of ${}_nq_x$ for each age group, W_o is the value of ${}_nq_x$ for each age group, W_o is the value of ${}_nq_x$ for age group proceeding to the W_o and $\Delta^2 W_{-1}$ is the second difference of W_{-1}.

The required data for individuals ages can be calculated by using the osculatory interpolation formula as follows:

$$U_X = U_o + X\Delta U_o + \frac{X(X-1)\Delta^2 U_o}{2}, \tag{2}$$

where U_x is the value of ${}_nq_x$ ay age x; U_o is the pivotal value of ${}_nq_x$ for each age group; ΔU_o is the first difference of U_o; $\Delta^2 U_o$ is the second difference of U_o; X is 0.2, 0.4, 0.6, 0.8.

The next step is to obtain the fitted values of ${}_nq_x$ by using the Least Squares Method. This method was required to find the constant value of $\hat{a}$ and $\hat{c}$ from the osculatory interpolation results. The formula is as follows:

$$q'_x = \hat{a}q_x + \hat{c}, \tag{3}$$

where q'_x is the fitted value of ${}_nq_x$ and q_x is the value of ${}_nq_x$ that obtain from the Australian Life Table.

HELIGMAN AND POLLARD MODEL

This model is proposed by Heligman and Pollard (1980). The mathematical function of Heligman and Pollard model is given by:

$$\frac{q_x}{p_x} is A^{(x+B)^C} + D * \exp\left(-E\left(\ln\left(\frac{x}{F}\right)\right)^2\right) + GH^x, \tag{4}$$

where q_x is the probability that an individual who has reached age x+ will die before reaching age x+1; p_x = 1 – q_x; A, B, C, D, E, F, G, H are positive parameters to be estimated.

The model contains three terms, each representing a distinct component of mortality. The first term is a rapidly decreasing exponential function and reflects the fall in mortality at the infant and early childhood ages, which are at ages less then 10 years. This component of mortality has three parameters; A, which is nearly equal to q_1 and measure the level of morality; C, which measures the rates of mortality decline in childhood where a higher value of C implies a faster mortality decrease when the age increases; and B, which is an age displacement to account for infant mortality. It should be noted that when $B = 0$, q_0 = 0.5 regardless of any values A and C. The second term is a function similar to the lognormal density and reflects the middle life mortality. It reflects the accident mortality for males and accident plus maternal mortality for the females, often referred to in the demographic literature as the accident "hump". The accident term has three parameters, where F indicating location, E representing spread and D the severity. And finally, the last term is a Gompertz exponential function which reflects the rise in mortality for ages greater than 40 years old.

This study analyse the mortality rates for person aged 10 years and above. Thus, the first term of Heligman and Pollard model is negligible. Therefore, (4) can be simplified and becomes:

$$\frac{q_x}{p_x} = D * \exp\left(-E\left(\ln\left(\frac{x}{F}\right)\right)^2\right) + GH^x. \tag{5}$$

From this equation, it is clearly that we only need to estimate five parameters of Heligman and Pollard model, which are *D*, *E*, *F*, *G* and *H*. Since this problem involves nonlinear equations that are difficult to solve explicitly, the Matrix Laboratory Version 7.0 (MATLAB 7.0) software will be used in the study to estimate these parameters. A nonlinear least squares algorithm with the capability of approximating numerically all derivatives was used in order to estimate the parameters of (5). However, in the study the Levenberg-Marquardt iteration procedure will be used to get the estimated values of these parameters by using a damped Gauss-Newton iteration procedure. The Levenberg-Marquardt algorithm provides a numerical solution to the mathematical problem of minimizing a sum of squares of several, generally nonlinear functions that depend on a common set of parameters. Like other numeric minimization algorithms, the Levenberg-Marquardt algorithm is an iterative procedure. The Levenberg-Marquardt algorithm interpolates between the Gauss-Newton algorithm and the method of gradient descent. The Levenberg-Marquardt algorithm is more robust than the Gauss-Newton algorithm since in many cases it finds a solution even if it starts very far off the global minimum. However, the detail of the method can be referred to Ibrahim (2008).

RESULTS

ESTIMATED COMPLETE LIFE TABLES, $\hat{q}_x$

The data from the Malaysian Abridged Life Table which provides information on the population according to age and sex in the year 1991, 1995 and 1999 will be used to estimate the complete life tables. The data was obtained from the Department of Statistics, Malaysia. In particular, the Malaysian Abridged Life Tables provides information on the values of ${}_nq_x$, l_x, ${}_nd_x$, ${}_nL_x$, T_x, e_x and so on. However, only the values of ${}_nq_x$ from the Malaysian Abridged Life Table are required for this study.

ESTIMATED ABRIDGED LIFE TABLES (FIVE-YEAR INTERVAL), ${}_n\hat{q}_x$

In order to make comparisons between our results and actual abridged life tables, we have derived the abridged life tables (five-year age intervals) for for both gender, for the period 1991, 1995 and 1999. The process of obtaining an abridged life table from a complete life table is simpler than its construction directly from population and deaths data, given that a complete life table can be easily aggregated into 5- or 10- age groups. Throughout this section we shall assume that there are available certain population data from which it is desired to produce an abridged life table for individual ages at five-yearly intervals. We have also considered convenient to leave a first group containing the information referred to 0 years by his specific characteristics. Thus remaining intervals are [5, 10), [10, 14), and so on, and the last interval corresponded to ages 75 years and over. Note that the value for the last interval length is not important, since this is calculated as an open interval. Since we only analyse the mortality rates for person aged 10 years and above, the results are divided into five-year age categories are running from [10, 14), [15, 19), and so on, up to [75-79). We analyse until age 79 years because the actual data given only until age 79 years.

The procedures to produce the estimated abridged life tables are as follows: First, the probability of surviving inside each age interval has been obtained using the relation,

$$ {}_n\hat{p}_x = \prod_{t=0}^{n} \hat{p}_{x+t}, \tag{6} $$

where $\hat{p}_x$ is the the probability of surviving at age *x* from the estimated completed life table.

Next, we can estimate the probabilities of dying and the values of the cohort function directly as:

$$ {}_n\hat{q}_x = 1 - {}_n\hat{p}_x, \tag{7} $$

where $x = 10, 15, 20, \ldots, 75$.

STANDARD ERROR IN ESTIMATED ABRIDGED LIFE TABLES

The estimated abridged life tables resulted from the King's Osculatory Interpolation and the Heligman and Polard are compared with the actual abridged life tables, using the measure:

$$ se^Y = \left| {}_n^Y\hat{q}_x - {}_nq_x \right| \times 100, \tag{8} $$

where ${}_n^Y\hat{q}_x$ is the probability of death in age interval [x, x+n) estimated by method Y and ${}_nq_x$ is the real probability of death in age interval [x, x+n) from given abridged life tables.

Based on the minimum values of (8), the results in Table 1 and 2 show that the King's Osculatory Interpolation method is a better method compared to the Heligman and Pollard model for all ages of 10 to 74 years, for both man and woman, in year 1991. However, Table 1 and 2 show that the King's Osculatory Interpolation method is better for ages less than 60 years, while the Heligman and Pollard model is better for ages greater than 60 years for Malaysian population in year 1995 and 1999 for both man and woman. The reason for this is that the estimated values of the abridged life tables are closest to the actual values of the abridged life tables at that particular ages. Another result from Table 1 and 2 is that there are no considerable differences between the methods in estimating the standard error in estimated abridged life tables in both man and woman.

TABLE 1. Comparison of results between King's Osculatory Interpolation (king) and Heligman and Pollard model (hp) of estimating standard error in estimated abridged life tables for man in year 1991, 1995 and 1999

[x,x+n)	se^{king}-m91	se^{hp}-m91	se^{king}-m95	se^{hp}-m95	se^{king}-m99	se^{hp}-m99
10--14	0.02154	0.09529	0.07459	0.05796	0.06861	0.04509
15-19	0.00336	0.00331	0.02193	0.02954	0.01867	0.02733
20-24	0.01836	0.06362	0.01999	0.10235	0.01371	0.10567
25-29	0.01430	0.03517	0.01301	0.02731	0.01075	0.00363
30-34	0.00313	0.05695	0.00517	0.08483	0.00198	0.17813
35-39	0.01471	0.07198	0.02102	0.00649	0.00711	0.12042
40-44	0.03230	0.18845	0.01193	0.15769	0.00020	0.13915
45-49	0.02132	0.27879	0.03956	0.32400	0.05240	0.48209
50-54	0.02874	0.04625	0.00608	0.19401	0.00362	0.48568
55-59	0.20969	0.04390	0.11308	0.07910	0.03424	0.33590
60-64	0.02632	0.60809	0.01861	0.03284	0.33148	0.24251
65-69	0.26134	0.78836	0.67410	0.50127	0.91653	0.80360
70-74	0.32453	2.11545	3.72474	2.45563	3.00564	2.48170
75-79	2.97079	1.04780	4.46886	2.78211	6.63785	5.70228

Bold: Present the better results between King's Osculatory Interpolation and Heligman and Pollard model

TABLE 2. Comparison of results between King's Osculatory Interpolation (king) and Heligman and Pollard model (hp) of estimating standard error in estimated abridged life tables for woman in year 1991, 1995 and 1999

[x,x+n)	se^{king}-f91	se^{hp}-f91	se^{king}-f95	se^{hp}-f95	se^{king}-f99	se^{hp}-f99
10--14	0.00882	0.01872	0.01210	0.01799	0.01410	0.04222
15-19	0.00274	0.02705	0.00122	0.01755	0.00118	0.02349
20-24	0.00044	0.00114	0.00068	0.01695	0.00426	0.01946
25-29	0.00540	0.00272	0.00250	0.01368	0.00863	0.01059
30-34	0.01094	0.03061	0.00217	0.02330	0.00223	0.04904
35-39	0.00072	0.03825	0.01344	0.01064	0.00654	0.03444
40-44	0.02831	0.01721	0.01821	0.01495	0.01383	0.02212
45-49	0.00965	0.08810	0.02286	0.01626	0.02205	0.09230
50-54	0.04235	0.11639	0.01874	0.08114	0.04175	0.15234
55-59	0.01288	0.31036	0.14383	0.18827	0.02417	0.07547
60-64	0.31009	0.01213	0.21871	0.21550	0.44104	0.11625
65-69	1.47023	0.16396	0.65160	0.37134	1.27333	0.37475
70-74	0.70012	1.74700	2.59363	1.20767	1.28893	0.23808
75-79	1.34763	0.67763	3.05296	2.33879	1.88677	1.24119

Bold: Present the better results between King's Osculatory Interpolation and Heligman and Pollard model

CONCLUSION

The method proposed by King's Osculatory Interpolation provides the best estimates of abridged life tables for age 10 to 60 years, while the Heligman and Pollard model method provides the best estimates of abridged life tables from age 60 up to 79. Hence, the King's Osculatory Interpolation method is most useful for Malaysian population over the age range 10 to 60 years, while the method of Heligman and Pollard model is useful for Malaysian population over the age range 60 to 79 years, However, in order to analyse the mortality trends over the age range 10 to 79 years, we suggest using both methods to obtain the best estimate.

REFERENCES

Elandt–Johnson, R.C. & Johnson, N.L. 1980. *Survival Models and Data Analysis*. New York: Wiley Series.

Hartmann, M. 1987. Past and recent attempts to model mortality at all ages. *Journal of Official Statistics* 3: 19-36.

Heligman, L. & Pollard, J.H. 1980. The age pattern of mortality. *Journal of the Institute of Actuaries* 107 (Part 1): 49-80.
Ibrahim, R.I. 2008. Expanding an abridged life table using the Heligman-Pollard Model. *Matematika* 24: 1-10.
King, G. 1914. On a short method of constructing an Abridged Life Table. *Journal of the Institute of Actuaries* 48: 294-303.
Kostaki, A. 1991. The Heligman-Pollard formula as a tool for expanding an Abridged Life Table. *Journal of Official Statistics* 7: 311-323.
Kostaki, A. 2000. A relational technique for estimating the age-specific mortality pattern from grouped data. *Mathematical Population Studies* 9(1): 83-95.
London, D. 1987. *Survival Models and Their Estimation*. U.S: Actex Publications.
Mode, C. & Busby, R. 1982. An eight parameter model of human mortality-the single decrement case. *Buletin of Mathematical Biology* 44: 647-659.

Rose Irnawaty Ibrahim*
Faculty of Science and Technology
Universiti Sains Islam Malaysia
Bandar Baru Nilai
71800, Nilai, Negeri Sembilan
Malaysia

Zailan Siri
Institute of Mathematical Sciences
Faculty of Science
University of Malaya
50603 Kuala Lumpur
Malaysia

*Corresponding author; email: rose.irnawaty@usim.edu.my

Received: 29 September 2010
Accepted: 1 March 2011