U.P.B. Sci. Bull., Series A, Vol. 69, No. 1, 2007 ISSN 1223-7027

LIFE TABLES FOR ROMANIA SURVIVAL FUNCTION.LIFE EXPECTANCY FOR ROMANIAN PEOPLE

Elena Corina CIPU[1], Carmen DRAGNEA[2]

S-a obținut un algoritm de completare a tabelului de supravietuire redus folosind datele de mortalitate din anul 2002. Pe baza tabelului de supravietuire prin metode grafice si analitice s-a gasit o aproximare a functiei de supravietuire.

A calculation procedure of abridged life table was obtained using the mortality data from year 2002. Using graphical and analytical methods, based on the life table, it was obtained an estimation of the survival function.

Keywords: life tables, survival function.
Mathematics Subject Classification 2000: 62N01, 62N02

1. Introduction

The most important life table applications are risk evaluation in the insurance industry, pensions, social planning, long term care and life expectancy estimation. In the late 16th century, local parishes were the first population which recorded mortality data. All the tables from this time didn't take into account the risks, so the mortality rate didn't have any information about the mortality intensity.

The concept of "life table" is introduced in "Natural and political observation upon bills of mortality" by John Graunt, in 1662, who made the first detailed investigation upon the mortality of Londoners. Johann de Witt (1625 – 1672) has introduced a new method of mortality rates usage through the study of life annuities and statistics. Dr. Edmond Halley constructed, about 1693, the first and the most complete life table. He had difficulties in finding records of the age of death, using at last a record for five years of Breslau births and deaths, with the age at death stated. He was the first who took into account that the population was dynamic using corrections for compensating the birth excess over the deaths. In

[1] Lecturer, Dept. of Mathematics III, University "Politehnica" of Bucharest, ROMANIA
[2] Assist, Dept. of Mathematics III, University "Politehnica" of Bucharest, ROMANIA

1765 Dr. Richard Price constructed the first mortality table which could be used as base for the premium calculations.

Through time multiple mortality tables were made. The most general demography research was made using English Life Expectancy tables. These tables used conditional probabilities and they showed the fact that ensured persons had a life expectancy greater than the rest of the population.

2. Position of the problem

The study of survival function is essential both for knowing the evolution and the structure of the population depending on its main indicators and also for demographic researches.

The need to analyze the survival function recently has raised as a consequence of the process of getting older that some countries are passing through, like for instance, our society. This gradual effect but constant has direct consequences on the healthcare, learning and pension planning.

The demographers and biologists are especially interested of the structure study and evolution of a population. For this many methods were developed to present statistics about populations. One of the methods is the survivorship table. Here the births and deaths effects on the number of population are very clear illustrated and studied.

The life tables are used to describe and understand the dynamic of a species. The gathered information are important in the fauna study (for instance the reintroduction of a specie in a new habitat), in agriculture (for instance the removal of undesired plants from a plant culture) and human health (for instance the development of an epidemics).

After the length of the elementary age intervals, the survival tables are of 2 types: c***omplete:*** the tables are created for 1 year periods; a***bridged:*** on age classes, usually from 5 to 5 years (0-4, 5-9, 10-14...), used when age, year or dates from the table are incomplete.

To be able to put it in practice, the mortality tables are created based on the mortality rate independence at age X regarding the total number of people from that population.

3. Biometrical functions. The survival function

The intensity of the mortality is measured by certain numeric coefficients which are called biometric functions. For the construction of the survival tables we use the following set of parametric functions: l_x - the number of survivors from the arbitrary number l_0 corresponding to age class x, ${}_n d_x$ - the number of deaths in the interval $[x, x+n)$, ${}_n q_x$ - the probability of dying in the interval $[x, x+n)$, ${}_n L_x$ - the number of person - years that survived , with age in the interval $[x, x+n)$, T_x - the number of person - years that survived after age x, e_x - life expectancy, ${}_n \mu_x$ - the force of mortality, ${}_n p_x$ - the probability of survival in the interval $[x, x+n)$.

The functions ${}_n q_x$, l_x and ${}_n d_x$ are measures of frequency (or intensity) because they show the events frequency (death or survival), and T_x, ${}_n L_x$, e_x are measures of the lengths expressed in person – years. The survival function l_x which describes the number of survivors at age x satisfies the relation $\frac{dl_x}{dx} = -l_x \mu_x$, where μ_x is the force of mortality and is a non negative function. This differential equation has the solution: $l_x = l_0 e^{-\int_0^x \mu_s ds}$, l_0 being the value of the function at the moment considered as reference.

The mortality tables contain values of l_x for age classes**,** for the intermediary ones it is used a method of interpolation. For this reason l_0 is chosen such that the interpolation is as accurate as possible. From practical reasons the mortality tables use a maximum age ω = 86, 100, 105 or 110 for x greater than ω, l_x is 0. The probability that a person at age x to live another **t** years is denoted with ${}_t p_x$ and the probability that a person dies in this period is denoted with ${}_t q_x = 1 - {}_t p_x$. We get:

$$ {}_t p_x = l_{x+t} / l_x = e^{-\int_0^t \mu_{x+s} ds}, \quad {}_t q_x = (l_x - l_{x+t}) / l_x = \int_0^t {}_s p_x \mu_{x+s} ds $$

Through convention when the survival period is from x to $x+1$, the left index is eliminated such that $p_x = {}_1 p_x$ and $q_x = {}_1 q_x$.

4. Calculation methods. Completion of abridged tables

A survival table is a rectangular matrix formed from biometric functions (columns) and age classes (rows). We shall use the following notations: N – the total number of individuals of a population, X – age class for interval $[x, x+n)$, n – the length of an elementary age class, D – the number of deceased with the age in the interval $[x, x+n)$, ω **-** the upper age limit.

For the construction of the table we use each class with the lengths of 5 years except the class 0-1, 1-4 and 100+ class (the interval of the last class is open at the upper bound). Taking into account these considerations we build the survival table in the following way:

Table1.

The scheme of the survival table

X	D	lx	dx	Qx	px	Lx	Tx	ex
0--1								
1--4								
5--9								
10--14								
....								

In D column we introduced the census observed values from the year 2002, showing the number of deceased persons with age within age class X.

For completing the survival table (Table .1) we use the following algorithm for row calculus:

1: l_0 is initialized with the value 100.000;

2: The values of d_x are computed by multiplying the column D with $100000 / N$. It is given by the following formula $_n d_x = 1000000 D_x / N$. For verification, we sum the "D" column and obtain 99999;

3: The number of people alive at the age x, l_x is obtained by rounding the difference $l_x -_n d_x$: $l_{x+n} = ROUND(l_x -_n d_x)$;

4: For the rest of the biometric functions, we calculate the corresponding values like this: $_n q_x = \frac{_n d_x}{l_x}$; $_n p_x = 1 -_n q_x$; $_n L_x = n \cdot \frac{l_x + l_{x+n}}{2}$; $T_i = \sum_{j=i}^{x} {_n L_j}$; $e_x = \frac{T_x}{l_x}$.

We constructed the survival table for Romania based on the row calculus algorithm corresponding to the Romanian population (Table 2).

The source of the used data.

The data used in this paper were given by the National Institute of Statistics, using "The population census made at 1 July 2002 and demographic statistics bulletin completed by the registry and the health departments from each district with data regarding births and deaths have been the sources of the data regarding the population and the demographic structure. The stable population of a place at 1st of July includes the population which resides in that place." At the 2002 census, the stable population has been determined according to CEE/UNO recommendation for population and living places censuses.

We used the numerical values for "Deaths by age group and sex" and "Infant deaths by age group".

5. The construction of the life table for Romanian population (total/ men/women) at 1st July 2002

For construction of the survival table we will be using the algorithm presented in the 3rd paragraph.

We initialize the variables according to the given data from the census, such that: N = 269666 – the total number of deceased people from the population, X – the age class for the interval $[x, x+n)$, n =5 – the length of an elementary age interval, D – the number of deceased, with age included in the interval $[x, x+n)$, ω =100 – upper age limit. The initial data were distributed by age classes like this: 0-4, 5-9, 10-15 ... 95-100, 100+.

Due to the high infantile mortality rate, we divided the first age class in the intervals 0-1 and 1-4. This division was possible because of the separate statistics supplied by NSI about infantile mortality.

In the construction of the survival tables for the dying probability we have the following constraints: ${}_{\infty}q_{\omega} = 1$, if the last age interval is open, $q_x \leq 1$ if the last age interval is closed. We also observe that due to the using of the abridged tables, the last column represents the modified life expectancy.

Table 2:

Life table for Romanian population in the year 2002.

class X	D	lx	dx	qx	px	Lx	Tx	ex
0—1	3648	100000	1352.785	0.013528	0.9865	495940	7178478	71.78
1—4	731	98647	271.0761	0.002748	0.9973	492152.5	6682538	67.74
5—9	436	98376	161.6815	0.001644	0.9984	490745	6190385	62.93
10—14	787	98214	291.8425	0.002971	0.9970	489580	5699640	58.03
15—19	820	97922	304.0799	0.003105	0.9969	487652.5	5210060	53.21
20—24	1291	97618	478.7404	0.004904	0.9951	485410	4722408	48.38
25—29	1598	97139	592.5849	0.0061	0.9939	481722.5	4236998	43.62
30-34	2687	96546	996.4178	0.010321	0.9897	477517.5	3755275	38.90
35-39	2936	95550	1088.754	0.011395	0.9886	469532.5	3277758	34.30
40-44	5926	94461	2197.533	0.023264	0.9767	457122.5	2808225	29.73
45-49	10450	92263	3875.164	0.042001	0.9580	439025	2351103	25.48
50—54	13595	88388	5041.422	0.057037	0.9430	416302.5	1912078	21.63
55—59	14060	83347	5213.857	0.062556	0.9374	383070	1495775	17.95
60—64	22254	78133	8252.431	0.10562	0.8944	340920	1112705	14.24
65-69	31404	69881	11645.52	0.166648	0.8334	283147.5	771785	11.04
70-74	40064	58235	14856.9	0.25512	0.7449	211107.5	488637.5	8.39
75-79	46302	43378	17170.13	0.395826	0.6042	144197.5	277530	6.40
80-84	32110	26208	11907.32	0.454339	0.5457	79107.5	133332.5	5.09
85-89	23909	14301	8866.153	0.619967	0.3800	38005	54225	3.79
90-94	12227	5435	4534.127	0.834246	0.1658	13777.5	16220	2.98
95-99	2224	901	824.7239	0.915343	0.0847	2252.5	2442.5	2.71

Using Sprague interpolation formula we can obtain the annual survival probability (that is used with annual rate of interest on primes of insurance) on the side with survival function.

Sprague Interpolation

Suppose we interpolate between two values v_0 and v_1 by a polynomial function p if there are known six values v_{-2}, v_{-1}, v_0, v_1, v_2, v_3. In order to determine this polynomial there are imposed the following conditions: **a.** find two quadratic polynomials *f* and *g* such that $f(i)=v_i$, $i \in \{-2,-1,0,1,2\}$ and $g(i)=v_i$, $i \in \{-1,0,1,2,3\}$), **b.** find a polynomial of degree five by conditions: $p(i)=v_i$, $i \in \overline{0,1}$, $p'(0)=f'(0)$, $p''(0)=f''(0)$, $p'(1)=g'(1)$, $p'''(1)=g''(1)$.

Let be Δ the forward difference operator: $\Delta u_x = u_{x+1} - u_x$. By Newton's interpolation formula f and g are: $f_x = (1+\Delta)^{x+2} u_{-2}$, $g_x = (1+\Delta)^{x+1} u_{-1}$. With $\delta^x = \Delta^x u_{-2}$ and $\delta_1^x = \Delta^x u_{-1}$, $\delta_1^x = \Delta^x (1+\Delta) u_{-2} = \delta^x + \delta^{x+1}$ we have:

$$f_x = 1 + (x+2)\delta + \frac{(x+2)(x+1)\delta^2}{2} + \frac{(x+2)(x+1)x\delta^3}{6} + \frac{(x+2)(x+1)x(x-1)\delta^4}{24}$$

$$g_x = 1 + (x+1)\delta_1 + \frac{(x+1)x\delta_1^2}{2} + \frac{(x+1)x(x-1)\delta_1^3}{6} + \frac{(x+1)x(x-1)(x-2)\delta_1^4}{24}. \quad (4.1)$$

The first and second derivative of f in $x=0$ are:

$$f_x'(0) = \delta + 3\delta^2/2 + \delta^3/3 - \delta^4/12 = F_1, \; f_x''(0) = \delta^2 + \delta^3 - \frac{1}{12}\delta^4 = F_2, \quad (4.2)$$

and the two derivatives of g in $x=1$ are:

$$g_x'(1) = \delta_1 + 3\delta_1^2/2 + \delta_1^3/3 - \delta_1^4/12 = \delta + 5\delta^2/2 + 11\delta^3/6 + \delta^4/4 - \delta^5/12 = G_1,$$

$$g_x''(1) = \delta_1^2 + \delta_1^3 - \delta_1^4/12 = \delta^2 + 2\delta^3 + 11\delta^4/12 - \delta/12^5 = G_2. \quad (4.3)$$

If we search for the function $p(x) = a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + +a_1x + a_0$ then for the coefficients $(a_i)_{i\in\overline{0,5}}$ appear the conditions:

$$p(0) = a_0 = u_0, p(1) = a_5 + a_4 + a_3 + a_2 + a_1 + a_0 = u_1,$$
$$p'(0) = a_1 = F_1, \; p'(1) = 5a_5 + 4a_4 + 3a_3 + 2a_2 + a_1 = G_1, \quad (4.4)$$
$$p''(0) = 2a_2 = F_2, \; p''(1) = 20a_5 + 12a_4 + 6a_3 + 2a_2 = G_2,$$

Solving the system (4.4) we find the coefficients $(a_i)_{i\in\overline{0,5}}$ therefore the polynomial p. We rewrite the system of equations (5.4) as:

$a_0 = u_0, a_1 = F_1, a_2 = F_2/2$, $a_5 + a_4 + a_3 = u_1 - u_0 - F_1 - F_2/2 = a$,

$$5a_5 + 4a_4 + 3a_3 = G_1 - F_2 - F_1 = b\,,\ 20a_5 + 12a_4 + 6a_3 = G_2 - F_2 = c\,, \quad (4.5)$$

with the solution: $a_0 = u_0$, $a_1 = F_1$, $a_2 = F_2/2$, $a_3 = a^2[6a - c/2 - 3b]$,

$$a_4 = -a^2[15a + c/2 - 7b]\,,\ a_5 = a^2[10a + c/2 - 4b] \quad (4.6)$$

where: $a = 3u_2/24 - u_1/3 + u_0/4 - u_{-2}/24$, $b = -u_3/12 + 5u_2/6 - 2u_1 + 11u_0/6 - 7u_{-1}/12$, $c = -u_3/12 + 17u_2/12 - 23u_1/6 + 23u_0/6 - 17u_{-1}/12 + u_{-2}/12$.

Annual data determination for the survival function

Once the survival probability p_x determined we can find the survival function l_x for $l_{x+t} = l_x \cdot {}_t p_x$, $t \in [1,4]$ (if we use the left string values of survival probabilities and ${}_{t_1+t_2} p_x = {}_{t_1+t_2} p_{x+t_1} \cdot {}_{t_1} p_x$, $\forall t_1, t_2 > 0$, $x > 0$), respectively $l_x = l_{x+t} / {}_t p_x$, $t \in [1,4]$ (using the right string values for ${}_t p_x$), on each five years interval and determining the mean of those values. For all the obtained data we consider a cubic interpolation and the result is presented in Fig. 1.

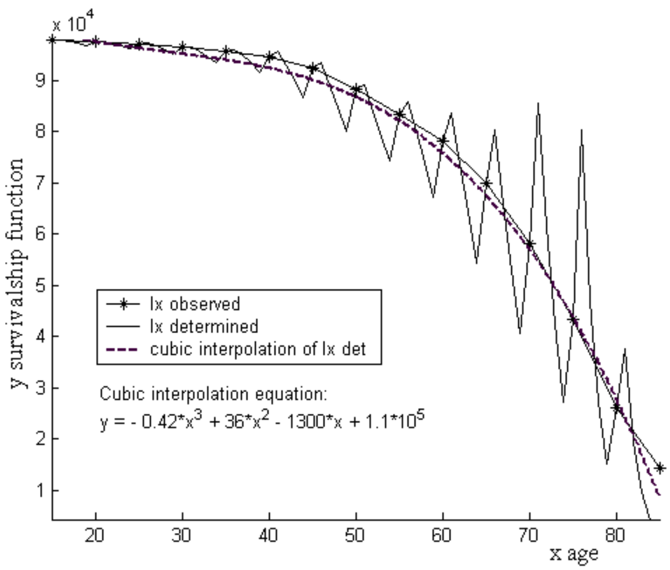

Fig. 1: The survival function for the population of Romania -2002; comparison between observed data at reduced level, annual determined data and interpolated data.

Our goal is to find the survival function as $l_x = K \cdot s^x \cdot g^{c^x}$, $K = l_{x_0} \cdot e^{a x_0} \cdot e^{b \cdot c^{x_0} / \ln c}$, $s = e^{-a}$, $g = e^{-b/\ln c}$ if we consider for the force of mortality a function type Makeham - Gompertz: $\mu_x = a + b \cdot c^x$. Using relation $\ln p_x = \ln s + c^x (c-1) \ln g$ for estimation of parameters a, b, c by King and Hardy method (where are considered three series of observations at the same length *n*, for persons who have consecutively ages, with $A_x = \sum_{y=x}^{x+n-1} \ln p_y$, the parameters are $c = \sqrt[n]{\frac{A_{x+2n} - A_{x+n}}{A_{x+n} - A_x}}$, $g = \exp(\frac{A_{x+n} - A_x}{c^x (c^n - 1)^2})$, $a = -\ln(p_x g^{c^x (1-c)})$, $b = -\ln g \cdot \ln c$). The result for $n \in \overline{10,23}$ and for the start age $x = 15$ with an averaging of the obtained values is précised in Fig. 2.

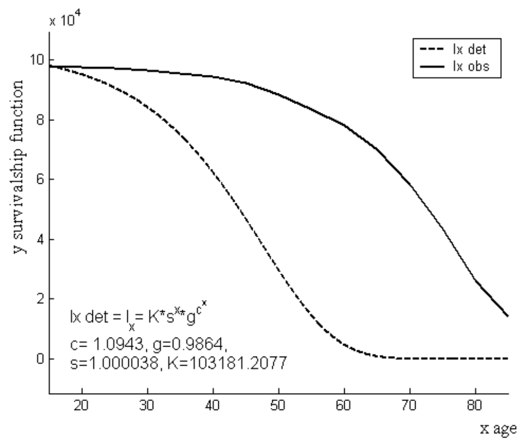

Fig. 2: The survival function for the population of Romania – 2002
Observed data at reduced level and determined data.

In order to determine the values of life expectancy we use $e_x = 1/2 + \sum_{n=1}^{\omega-x-1} l_{x+n} / l_x$ with l_x as third grade polynomial function obtained by cubic interpolation. We use also a fitting factor $\text{dif} = (e_x \text{obs} - e_x \text{det})(15) = 6.9$ (see Fig. 3.).

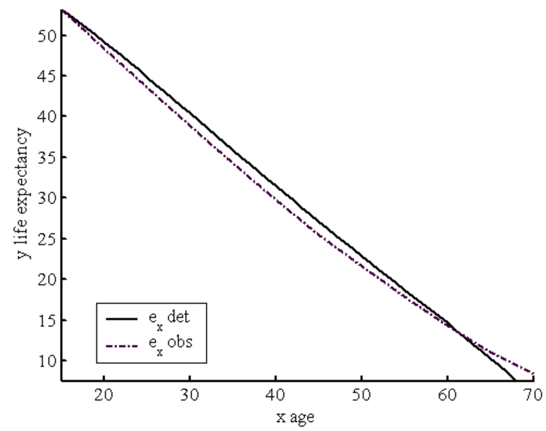

Fig. 3: Life expectancy for the population of Romania – 2002, Observed data and determined data.

6. Conclusions

We made an algorithm for the completion of the abridged survival table, according to the available data for Romania. This table contains: the survival function, the number of deaths, dying probability, the number of persons – years that survived, the number of person – years that survived after age x, the life expectancy, the force of mortality and the probability of survival.

We use Sprague interpolation for p_x observed function because the grade of the polynomial function that approximate with a lowest error is five (grade obtained if one use the lowest small squares method searching *r* such that $\mathrm{Var}(\Delta^r qx)/\binom{2r}{r} - \mathrm{Var}(\Delta^{r+1} qx)/\binom{2(r+1)}{r+1} < \varepsilon \to 0$, $\binom{2r}{r}$ the binomial coefficients, see H. Herero [1], T. Haavelmo [4]).

From Figure 2 we observe that Makeham-Gompertz function is not a very good expression for the force rate of mortality for Romania population. One can try an estimation of parameters for Perks distribution: $\mu_x = \dfrac{a + b \cdot c^x}{1 + d \cdot c^x}$ (see Horiuchi, Coale [5]).

Acknowledgements: The paper was presented at the communication session in duty to professors: Târnoveanu and I. Roşculeţ.

R E F E R E N C E S

[1]. H.C. Herrero, *An experience constructing a complete and an abridged life table using a mathematical formula for a small population,* Departamento de Economía Aplicada, Universidad de Valladolid, 2002.

[2]. I. Purcaru, I. Mircea, Gh. Lazăr: *Asigurări de persoane şi de bunuri. Aplicaţii. Cazuri. Soluţii*, Ed. Economică, Buc., 1998.

[3]. H. Tudor, O. Popescu: *Matematici financiare şi actuariale*, Ed. Albastră, Cluj-Napoca, 2004.

[4]. T. Haavelmo, *A note on the variate difference method*, Econometrica, Vol. 9, Nr. 1, 1941, pp. 74-79.

[5]. S. Horiuchi, A.J. Coale 1990, *Age Pattern of Mortality for Older Women: An Analysis using the Age-Specific Rate of Mortality Change with Age*, Mathematical Population Studies, Vol.2 Nr. 4, 245-267.