Life Table

Jan M Hoem, The Max Planck Institute for Demographic Research, Rostock, Germany

Abstract

In this article, modern statistical language is used to explain the basic notions of the oldest tool of demographic analysis, namely the life table. Following a long tradition in which demographers seek to establish sensible results out of a suboptimal data set, a real-life example is used to describe how one might produce a simple life table in the presence of digital preference (age heaping, grouped time reporting). An extension to multiple-decrement (or competing-risk) life tables is also provided.

The life table is perhaps the oldest tool developed for the analysis of survival patterns in human and other populations. Its roots go back at least to the famous book on the English Bills of Mortality that John Graunt published in 1662, a book that is sometimes cited as the beginning of systematic statistical science. Life-table techniques are described in detail in most introductory textbooks on the methods of actuarial statistics, biostatistics, demography, and epidemiology (see, e.g., Chiang, 1984; Elandt-Johnson and Johnson, 1980; Manton and Stallard, 1984; Hinde, 1998; Preston et al., 2001). In this article the basic notions of life-table construction are explained with a focus on issues that are sometimes underemphasized in other accounts. More of the underlying mathematical theory can be found in Hoem (1998) and in the textbooks. For the history of the topic, consult Seal (1977), Smith and Keyfitz (1977), and Dupâquier (1996).

A Cohort Life Table

Classical presentations of life-table techniques typically start with data in a format like that of the first four columns in **Table 1**. These particular data come from the Eritrean Demographic and Health Survey of 1995 and show the survival up through their fourth birthday of 7098 girls whose births and deaths were recorded in the survey. This data set is different from normal textbook examples in two ways. First, the data are given for intervals of single-month segments instead of the usual single-year intervals. Second, the present data are subject to strong heaping in the reported age at death for the girls that died. In part this is due to a typical inherent inaccuracy of age reporting in societies like the data source at the time of data collection, and in part due to the fact that (probably for that very reason) only integer years of age attained was asked except for very young ages at death. By contrast, an individual child's age when observation ended (age at censoring) could be computed accurately to the month, for in these data a child's record was censored only because the child's parent was interviewed; the age of a live child could then be calculated from its date of birth. This unconventional data set is used in order to illuminate how a number of practical issues that confront a demographer can be handled in a first analysis of mortality. A more complete investigation would subsequently use intensity-regression methods from event-history analysis (*see* Event History Analysis: Applications), as was done by Woldemicael (1999).

If there had not been any heaping in the reported ages at death, one would calculate $R_x = S_x - \frac{1}{2}(D_x + W_x)$ as an approximate number of person-months of exposure at age x. Here S_x is the number of children who survive to exact age x months, D_x is the number of deaths recorded between exact ages x and $x+1$ months, and W_x is the corresponding number of records that were censored during this interval. The approximation consists in assigning deaths and interviews to in the middle of the month on average. It has become known as the 'actuarial method.' The next step would be to compute occurrence/exposure rates D_x/R_x (*see* Demographic Techniques: LEXIS Diagram). The age heaping is abundantly evident in the plot of the 'raw' death rates (converted to 'deaths per 1000 person-years') in **Table 1**.

Such age heaping must have been caused by a shift of the real age at death to a reported near-by 'round' anchor age like 3 months, 6 months, or multiples of 12 months. The deaths in some interval $[x_0-a,\ x_0+b]$ around each anchor age x_0 have largely been assigned to the anchor age instead of to each age in the interval, but the total number of deaths $\sum_{t=-a}^{b} D_{x_0+t}$ in such an interval should be largely correct. Similarly, the total exposures $\sum_{t=-a}^{b} R_{x_0+t}$ in the interval should be about right, even though the individual term in the latter sum may be wide off the mark. Under these circumstances it is more sensible to compute death rates for groups of reported ages and list them in the row for the corresponding anchor age. By this reasoning,

$$m_3 = \sum_{x=2}^{4} D_x \Big/ \sum_{x=2}^{4} R_x$$

is entered for age 3 months,

$$m_6 = \sum_{x=5}^{7} D_x \Big/ \sum_{x=5}^{7} R_x$$

is entered for age 6 months, and so on, as indicated in column 7 of **Table 1**. For the intervening ages x, rates m_x have been computed by linear interpolation. The corresponding curve of rates is given with the label 'grouped' in **Table 1**.

The rest of the life-table computations are based on the following simple mathematical theory. Each individual child i has a lifetime T_i with some distribution function $F(t) = P\{T_i \le t\}$ (with $F(0) = 0$), a probability density $f(t) = \mathrm{d}F(t)/\mathrm{d}t$,

Table 1 Life-table computations for Eritrean girls aged 0 through 4 years

Age x in months	Observed survivors	Deaths D_x	Censored W_x	Exposures R_x	Raw rate D_x/R_x per 1000	Age grouping (months)	Grouped rates m_x	Death probab. $1000q_x$	Life-table survivors l_x
0	7098	208	15	6986.5	357.26	0	357.26[a]	29.33	100 000
1	6875	48	39	6831.5	84.32	1	84.32[a]	7.00	97 067
2	6788	36	39	6750.5	64.00		70.01[b]	5.82	96 387
3	6713	43	36	6673.5	77.32	2–4	55.70[a]	4.63	95 826
4	6634	14	29	6612.5	25.41		56.00[b]	4.66	95 383
5	6591	21	22	6569.5	38.36		56.30[b]	4.68	94 939
6	6548	50	42	6502	92.28	5–7	56.60[a]	4.71	94 494
7	6456	21	23	6434	39.17		54.98[b]	4.57	94 050
8	6412	26	31	6383.5	48.88		53.35[b]	4.44	93 620
9	6355	19	26	6332.5	36.00		51.73[b]	4.30	93 204
10	6310	16	23	6290.5	30.52		50.11[b]	4.17	92 803
11	6271	10	37	6247.5	19.21		48.48[b]	4.03	92 417
12	6224	144	30	6137	281.57	8–17	46.86[a]	3.90	92 044
13	6050	6	34	6030	11.94		45.45[b]	3.78	91 685
14	6010	7	31	5991	14.02		44.05[b]	3.66	91 339
15	5972	3	15	5963	6.04		42.64[b]	3.55	91 004
16	5954	5	35	5934	10.11		41.24[b]	3.43	90 681
17	5914	3	31	5897	6.10		39.83[b]	3.31	90 370
18	5880	25	34	5850.5	16.40		38.43[b]	3.20	90 071
19	5821	5	24	5806.5	10.33		37.02[b]	3.08	89 783
20	5792	2	28	5777	4.15		35.62[b]	2.96	89 506
21	5762	0	16	5754	0.00		34.21[b]	2.85	89 241
22	5746	0	29	5731.5	0.00		32.81[b]	2.73	88 987
23	5717	2	18	5707	4.21		31.40[b]	2.61	88 744
24	5697	148	25	5610.5	316.55	18–30	30.00[a]	2.50	88 512
25	5524	0	34	5507	0.00		29.28[b]	2.44	88 291
26	5490	0	32	5474	0.00		28.57[b]	2.38	88 076
27	5458	0	31	5442.5	0.00		27.85[b]	2.32	87 866
28	5427	0	27	5413.5	0.00		27.14[b]	2.26	87 663
29	5400	0	35	5382.5	0.00		26.42[b]	2.20	87 465
30	5365	0	33	5348.5	0.00		25.71[b]	2.14	87 272
31	5332	0	39	5312.5	0.00		24.99[b]	2.08	87 086
32	5293	0	28	5279	0.00		24.27[b]	2.02	86 904
33	5265	0	36	5247	0.00		23.56[b]	1.96	86 729
34	5229	0	24	5217	0.00		22.84[b]	1.90	86 559
35	5205	0	28	5191	0.00		22.13[b]	1.84	86 394
36	5177	109	22	5111.5	255.89	31–42	21.41[a]	1.78	86 235
37	5046	0	38	5027	0.00		20.44[b]	1.70	86 081
38	5008	0	20	4998	0.00		19.48[b]	1.62	85 935
39	4988	0	32	4972	0.00		18.51[b]	1.54	85 795
40	4956	0	31	4940.5	0.00		17.54[b]	1.46	85 663
41	4925	0	30	4910	0.00		16.57[b]	1.38	85 538
42	4895	0	30	4880	0.00		15.60[b]	1.30	85 420
43	4865	0	24	4853	0.00		14.63[b]	1.22	85 309
44	4841	0	26	4828	0.00		13.66[b]	1.14	85 205
45	4815	0	18	4806	0.00		12.70[b]	1.06	85 108
46	4797	0	20	4787	0.00		11.73[b]	0.98	85 018
47	4777	0	16	4769	0.00		10.76[b]	0.90	84 935
48[c]	4761	46	20	4728	116.75	43–54	9.79[a]	0.82	84 859

[a]Computed as (sum of deaths)/(sum of exposures) for age groups indicated. (Used for rows where age groups are indicated.)
[b]Computed by linear interpolation.
[c]Additional rows for ages 49–60 months have been deleted.

and a 'force of mortality' or 'death intensity' $\mu(t) = f(t)/\{1 - F(t)\}$. The various lifetimes T_i are independent of each other and they have the same distribution $F(t)$. Simple integration gives $F(x) = 1 - \exp\{-\int_0^x \mu(t)\mathrm{d}t\}$. The death probability is.

$$q_x = P\{T_i \leq x + 1 | T_i > x\} = \{F(x+1) - F(x)\}/F(x)$$

from which is derived

$$q_x = 1 - \exp\left\{-\int_0^1 \mu(x+t)\mathrm{d}t\right\} \qquad [1]$$

If one approximates the function $\mu(x+t)$ by the constant m_x for $0 < t < 1$ (the assumption of piecewise constancy), then (1) gives $q_x = 1 - e^{-m_x}$, which has been used to compute the next-to-last column in Table 1. Finally, a life-table 'survival function' is defined as $\ell(x) = \ell(0)\{1 - F(x)\}$, with $l(0) = 100\,000$. This is the expected number of survivors to exact age x out of an original cohort of $\ell(0)$ initial individuals. Simple manipulation with the definition of q_x gives

$$\ell(x+1) = \ell(x)(1 - q_x), \quad \text{for } x = 0, 1, \ldots \qquad [2]$$

This is a practical recursive formula by which we have computed the final column in Table 1. From the final entry in the $\ell(x)$ column in Table 1 we see that more than 15% of the original cohort of girl babies would die before age 4 (years) according to our data, because 1−0.848 59 = 0.151 41.

The mean number of months lived between birth and age n is

$$\mathring{e}_{0:\overline{n}|} = E\min(T_i, n) = \int_0^n \ell(t)\mathrm{d}t/\ell(0) \qquad [3]$$

If one partitions the integral here into a sum of n terms $\int_{x-1}^{x} \ell(t)\mathrm{d}t$ for $x = 1, 2, \ldots, n$, use the trapezoidal rule of numerical integration to see that this integral is approximately equal to $\frac{1}{2}\{\ell(x-1) + \ell(x)\}$, and collect terms suitably, we get the numerical approximation

$$\mathring{e}_{0:\overline{n}|} \simeq \left\{ \sum_{x=0}^{n} \ell(x) - \frac{1}{2} - \frac{1}{2}\ell(n) \right\} / \ell(0) \qquad [4]$$

The girls in the Eritrean data set lost 5 months of life on average before they reached their fourth birthday, for according to eqn [4] with $n = 48$ months they experienced a mean lifetime of 43 months.

Other Life-Table Procedures

Life-table construction consists in the estimation of parameters and tabulation of functions like those above from empirical data. The data can be for age at death to individuals, as in the illustrative example, but they can also be observations of duration until recovery from an illness, of intervals between births, of time until breakdown of some piece of machinery, or of any other positive duration variable. In general, a T_i is the duration until occurrence of some event that ends individual survival in a given status. The method of estimation depends on the character of the data. If accurate (ungrouped) individual-level data are available, then the Kaplan–Meier estimator (*see* Event History Analysis: Applications) can be used to estimate $\ell(x)$ for all relevant x, and estimates of the other life-table functions can then be computed subsequently. Alternatively a segment of the Nelson–Aalen estimator can be used to estimate each integral $\int_0^1 \mu(x+1)\mathrm{d}t$.

Equation [1] can then be used to estimate q_x for each integer x, and the rest of the computations follow suit. Sometimes the force of mortality is represented by some function $h(x;\theta)$, where θ is a vector of parameters. For instance, actuaries often use the classical Gompertz–Makeham function $h(x; a, b, c) = a + bc^x$ for the force of mortality in their life tables (*see* 'Demographic Models; Population Dynamics: Probabilistic Extinction, Stability, and Explosion Theorems).

From any given schedule of death probabilities $q_0, q_1, q_2, \ldots$, the $\ell(x)$ table is easily computed using the recursive eqn [2]. Much of the effort in life-table construction is, therefore, concentrated on providing a schedule $\{q_x\}$.

So far, this account has assumed that the data come from a group of independent individuals who have all been observed in parallel, essentially a cohort that is followed from a significant starting point (namely from birth in our empirical example) and that is diminished over time due to 'decrements' ('attrition') caused by the risk in question and also subject to reduction due to censoring. The empirical example displayed data only for the first 4 years of life of such a cohort. One would have had to follow the cohort until the end of the cohort's total life to produce a 'complete cohort life table.' It is more common to compute a q_x schedule from data collected for the members of a population during a limited time period and to use the mechanics of life-table construction to produce a 'period life table' from the q_x values. If real mortality patterns are tied to cohorts, individuals who live at widely differing ages in the observational period do not normally have the same risk structure, and the period taken by the life table to reflect the patterns of a 'synthetic (fictitious) cohort' exposed to the risks of the period at their various ages.

Whichever way the life-table survivor column has been produced, eqn [4] can always be used to calculate mean numbers of time-units lived under the risk for which the life-table was constructed. If the time unit is a year and n is chosen so large that no one lives more than n years (symbolized by letting $n = \infty$ and $\ell(\infty) = 0$), $\mathring{e}_{0:\overline{\infty}|}$ is called the 'life expectancy' of the life table and is normally denoted $\mathring{e}_0$. The subscript 0 indicates that the calculation is made for a newborn individual. Similar computations can be made at any age x and one gets expected remaining lifetimes of the form $\mathring{e}_{x:\overline{n}|} = \int_0^n \ell(x+t)\mathrm{d}t/\ell(x)$, with approximation formulas similar to eqn [4].

Multiple-Decrement Tables

When suitable data are available, the risk of death (or force of mortality) may be partitioned according to mutually exclusive causes of death. Let $\mu_k(t)$ be death intensity or force of mortality for cause k at age t, and let the total force of mortality be $\mu(t) = \sum_k \mu_k(t)$. Then the probability that an individual who is alive at age x will die from cause k before age $x + 1$ is

$$q_x^{(k)} = \int_0^1 {}_tp_x \mu_k(x+t)\mathrm{d}t \qquad [5]$$

where ${}_tp_x$ is the probability of surviving to age $x + t$ for an individual who is alive at age x. For given risk intensities $\{\mu_k(\cdot)\}$, $q_x^{(k)}$ can be computed by numerical integration in eqn [5]. A corresponding column of $q_x^{(k)}$ values may then be added to the life table for each k, and other cause-specific life-table functions may also be computed. Note that $q_x^{(k)}$ is influenced by risk intensities other than $\mu_k(\bullet)$, for

$$_tp_x = \exp\left\{ -\sum_j \int_0^t \mu_j(x+s)\mathrm{d}s \right\}$$

where the exponent depends on all $\mu_j(\bullet)$, including those for $j \neq k$.

Several further life-table functions can be defined by formal reduction or elimination of one or more of the risk intensities in formulas like eqn [5]. In particular, a 'single-decrement life table' can be computed for each cause k in order to show what the life table would look like in the hypothetical case where this cause was the only one operating in the study population and where it did so with the risk function estimated from the data. The purpose is to see the 'pure' effect of the risk in question on a fictitious cohort without interference from other causes. (The decrement probabilities and other features of the single-decrement life table do not depend on any other intensities than the one for which the table is computed.) This does not mean that one should believe that the total risk intensity can actually be reduced to the risk in focus or that this risk operates independently of other causes.

Most single-decrement life-table functions have straightforward interpretations. If, for instance, $\ell_k(x)$ denotes the corresponding survival function, then $\ell_k(x)/\ell_k(0)$ is the probability of surviving to age x in the fictitious table. Suppose, for example, that the study population is a group of unpartnered individuals subject to the competing risks of marital and nonmarital union formation as well as to the risk of death. One can then compute a single-decrement life table based on the risk of transition into a nonmarital union alone and find the probability that an individual would end up in such a union by age 50 years, say, if marriage were no alternative and death did not operate, even though in reality marriages continue to be formed by group members and unpartnered individuals continue to die. In such a table one may get $\ell_k(\infty) > 0$ because the event never occurs to some individuals, even if it operates alone. (We will all die but some individuals never enter a nonmarital partnership even if they live to age 100 years.)

Life expectancies $\overset{\circ}{e}_{15:\overline{n}|}$ computed in the three-decrement life table would be interpreted as the mean number of years (say) lived in the unpartnered state between ages 15 and $15 + n$ years. In the single-decrement life table for nonmarital-union formation, one possibility is to compute the mean age of entry into a nonmarital union only for those who do enter such a union by age n, that is, the conditional mean.

$$\int_0^n \ell_k(x)\mathrm{d}x / \{\ell_k(0) - \ell_k(n)\} \qquad [6]$$

Conversely one can compute a 'cause-deleted life table' by eliminating one or more of the cause intensities in the formulas. To 'remove' cause k, introduce $\mu_{-k}(t) = \mu(t) - \mu_k(t)$ and construct a 'normal' life table by replacing $\mu(t)$ by $\mu_{-k}(t)$ everywhere. This would show what the life table would look like if it were possible to eliminate cause k without changing the risk of any other cause. It could be used for example to see how partnering probabilities would appear if the partnering process were undisturbed by mortality. One would then remove the force of mortality from the life-table formulas and operate only with the sum $\mu_{-k}(t)$ of the intensities of marital and nonmarital union formation in the computations. Note that this is again a purely hypothetical construct; in fact it is counterfactual for mortality cannot of course be completely removed.

Life expectancies can be calculated for a cause-deleted life table as for any other life table. If $\ell_{-k}(x)$ is the table's survival function and $\ell_{-k}(\infty) = 0$, then a formula for the cause-eliminated life expectancy is $\overset{\circ}{e}_0^{(-k)} = \int_0^\infty \ell_{-k}(x)\mathrm{d}x/\ell_{-k}(0)$. The classical applications were to cause-of-death removal, and the difference $\overset{\circ}{e}_0^{(-k)} - \overset{\circ}{e}_0$ was interpreted as the (fictitious) gain in life expectancy produced by removal of cause k. For example, Westergaard (1907) found that according to the English life tables of 1881–90, life expectancy for females would increase by 2.7 years if tuberculosis deaths could be eliminated, by a year if cancer deaths could be removed, and by 1.1 years if deaths due to diarrhea and dysentery could be deleted separately. The effects were not additive, and if all three cause-of-death groups could be removed, the gain would be 5.6 years and not 4.8. The life expectancy for females in the original table was $\overset{\circ}{e}_0 = 47.2$ years.

See also: Age Structure; Demographic Models; Demographic Techniques: LEXIS Diagram; Event History Analysis: Applications; Motivation: Life Course and Sociological Perspectives.

Bibliography

Chiang, C.L., 1984. The Life Table and Its Applications. Krieger, Malabar, FL.
Dupâquier, J., 1996. L'invention de la table de mortalité. Presses universitaires de France, Paris.
Elandt-Johnson, R.C., Johnson, N.L., 1980. Survival Models and Data Analysis. Wiley, New York.
Graunt, J., 1662. Natural and Political Observations Made upon the Bills of Mortality. Roycroft, London.
Hinde, A., 1998. Demographic Methods. Arnold, London.
Hoem, J.M., 1998. The life table. In: Armitage, P., Colton, T. (Eds.), Encyclopedia of Biostatistics. Wiley, Chichester, pp. 2235–2239.
Hoem, J.M., 2001. Demographic analysis: Probabilistic approach. In: Neil J. Smelser, Paul B. Baltes (Eds.), International Encyclopedia of the Social and Behavioral Sciences, first ed., vol. 5, pp. 3428–3432.
Manton, K.G., Stallard, E., 1984. Recent Trends in Mortality Analysis. Academic Press, Orlando.
Preston, S.H., Heuveline, P., Guillot, M., 2001. Demography: Measuring and Modeling Populations. Blackwell, Oxford, UK.
Seal, H., 1977. Studies in history of probability and statistics, 35: multiple decrements or competing risks. Biometrika 64 (3), 429–439.
Smith, D., Keyfitz, N., 1977. Mathematical Demography: Selected Papers. Springer-Verlag, Heidelberg, Germany.
Westergaard, H., 1907. The horoscope of the population in the twentieth century. In: Compte-Rendu de la XI Session de l'Institut international de Statistique á Copenhague du 26 au 31 aout 1907.
Woldemicael, G., 1999. Infant and Child Mortality in Eritrea: Levels, Trends, and Determinants (Ph.D. thesis). Stockholm University.