Life Table

A life table is a tabular representation of central features of the distribution of a positive **random variable**, say T, with an absolutely continuous distribution. It may represent the lifetime of an individual, the failure time of a physical component, the remission time of an illness, or some other duration variable. In general, T is the time of occurrence of some event that ends individual survival in a given status. Let its cumulative distribution function (cdf) be $F(t) = \Pr(T \leq t)$ and let the corresponding survival function be $S(t) = 1 - F(t)$, where $F(0) = 0$. If $F(\cdot)$ has the probability density function (pdf) $f(\cdot)$, then the risk of event occurrence is measured by the **hazard** $\mu(t) = f(t)/S(t)$, for t where $S(t) > 0$. Because of its sensitivity to changes over time and to risk differentials between population subgroups, $\mu(t)$ is a centerpiece of interest in empirical investigations.

In applications to human mortality, which is where life tables originated, the time variable normally is a person's attained age and is denoted x. The function $\mu(x)$ is then called the *force of mortality* or *death intensity* (*see* **Hazard Rate**). The life-table function $l_x = 100\,000\ S(x)$ is called the *decrement function* and is tabulated for integer x in *complete life tables*; in *abridged life tables* it is tabulated for sparser values of x, most often for five-year intervals of age. The *radix* l_0 is selected to minimize the need for decimals in the l_x table; a value different from 100 000 is sometimes chosen. Other life-table functions are the expected number of deaths $d_x = l_x - l_{x+1}$ at age x (i.e. between age x and age $x + 1$), the single-year death probability $q_x = \Pr(T \leq x + 1 | T > x) = d_x/l_x$, and the corresponding survival probability $p_x = 1 - q_x$. Simple integration gives

$$q_x = 1 - \exp\left[-\int_x^{x+1} \mu(s)\,\mathrm{d}s\right]. \qquad (1)$$

Life-table construction consists in the estimation and tabulation of functions of this nature from empirical data. If ungrouped individual-level data are available, then the **Kaplan–Meier estimator** can be used to estimate l_x for all relevant x and estimators of the other life-table functions can then be computed subsequently. Alternatively, a segment of the **Nelson–Aalen estimator** can be used to estimate $\int_x^{x+1} \mu(s)\,\mathrm{d}s$; (1) can then be used to estimate q_x for each x, and the rest of the computations follow suit. From any given schedule of death probabilities $q_0, q_1, q_2, \ldots$, the l_x table is easily computed sequentially by the relation $l_{x+1} = l_x(1 - q_x)$ for $x = 0, 1, 2, \ldots$. Much of the effort in life-table construction therefore is concentrated on providing such a schedule $\{q_x\}$.

More conventional methods of life-table construction use **grouped survival times**. Suppose for simplicity that the range of the lifetime T is subdivided into intervals of unit length and that the number of failures observed during interval x is D_x. Let the corresponding total person-time recorded under risk of failure in the same interval be R_x. Then, if $\mu(t)$ is constant over interval x (the assumption of *piecewise constancy*), then the *death rate* $\hat{\mu}_x = D_x/R_x$ is the **maximum likelihood** estimator of this constant. Relation (1) can again be used to provide an estimator

$$\hat{q}_x = 1 - \exp(-\hat{\mu}_x), \qquad (2)$$

and the crucial first step in the life-table computation has been achieved. Instead of (2), $\hat{\mu}_x \big/ \left(1 + \frac{1}{2}\hat{\mu}_x\right)$ is often used to estimate q_x. This solution is of older vintage and may be regarded as an approximation to (2).

Two kinds of problems may arise: (i) the exact value of R_x may not be known, and (ii) the constancy assumption for the hazard may be violated.

When the exact risk time R_x is not known, some approximation is often used. An Anglo-Saxon tradition is to use the midyear population in the age interval. Alternatively, suppose that the number N_x of survivors to exact age x and the number W_x of withdrawals (losses to follow-up) in the age interval are known. What has become known as the **actuarial method** then consists in approximating R_x by $N_x - \frac{1}{2}(D_x + W_x)$. If there are no withdrawals and N_x is known, then D_x/N_x is the maximum likelihood estimator of q_x, and this provides a suitable starting point for the life-table computations.

For the case where only grouped data are available and the piecewise-constancy assumption for the intensity function is implausible, various methods have been developed to improve on (2). For an overview, see Keyfitz [12]. Even if single-year age groups are used, mortality drops too fast in the first year of life to merit an assumption of constancy over this interval. Demographers often use $\hat{\mu}_0/[1 + (1 - a_0)\hat{\mu}_0]$ to estimate q_0, where a_0 is some small figure, say between 0.1 and 0.15 [2]. If it is

possible to partition the first year of life into subintervals in each of which mortality *can* be taken as constant, then it is statistically more efficient essentially to build up a life table for this year. This leads to an estimate like $\hat{q}_0 = 1 - \exp(-\sum_i \hat{\mu}_i)$, where the sum is taken over the first-year intervals. See Dublin et al. [5, p. 24] for an example.

The force of mortality is sometimes represented by a function $h(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. Actuaries most 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* **Parametric Models in Survival Analysis**). When individual-level data are available, it would be statistically most efficient to estimate the parameters by the maximum likelihood method, but most often they are estimated by fitting $h(\cdot;\boldsymbol{\theta})$ to a schedule of death rates $\{\hat{\mu}_x\}$, perhaps by **least squares**, minimum chi-square (*see* **Ban Estimates**), or some **method of moments**. This approach is called *analytic graduation*; for its statistical theory, see [11]. One of many alternatives to modeling the force of mortality is to let [10]

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D\exp[-E(\ln x - \ln F)^2] + GH^x.$$

So far we have tacitly assumed that the data come from a group of independent individuals who have all been observed in parallel and whose lifetimes have the same cdf. **Staggered** (delayed) **entries** into the study population and voluntary exits (withdrawals) from it are permitted provided they contain no information about the risk in question, be it death, recurrence of a disease, or something else. Nevertheless, the basic idea is that of a connected cohort of individuals that is followed from some significant starting point (like birth or the onset of some disease) and which is diminished over time due to *decrements (attrition)* caused by the risk's operation. In demography, this corresponds to following a **birth cohort** through life or a marriage cohort while their marriages last, and the ensuing tables are called *cohort life tables*.

Because such tables can only be terminated at the end of a cohort's life, it is more common to compute age-specific attrition rates $\hat{\mu}_x$ from data collected for the members of a population during a limited period and to use the mechanics of life-table construction to produce a *period life table* for the population from such rates. If mortality patterns are tied to cohorts, then individuals who live at widely differing ages in the period of observation cannot be expected to have the same risk structure, and the period table is said to reflect the patterns of a *synthetic* (fictitious) cohort exposed to the risk of the period at the various ages.

Multiple-decrement Tables

When two or more mutually exclusive risks operate on the study population (*see* **Competing Risks**), one may correspondingly compute a *multiple-decrement table* to reflect this. For instance, a period of sickness can end in death or, alternatively, in recovery. Suppose that an integer random variable K represents the *cause of decrement* and define $F_k(t) = \Pr(T \le t, K = k)$, $f_k(t) = \mathrm{d}F_k(t)/\mathrm{d}t$, and $\mu_k(t) = f_k(t)/S(t)$, assuming that all $F_k(\cdot)$ are absolutely continuous. Then $\mu_k(\cdot)$ is the cause-specific hazard (intensity) for risk cause k and $\mu(t) = \sum_k \mu_k(t)$ is the total risk of decrement at time t. For the multiple-decrement table, we define the decrement probability

$$\begin{aligned} q_x^{(k)} &= \Pr(T \le x+1, K = k | T > x) \\ &= \int_0^1 \exp\left[-\int_0^t \mu(x+s)\,\mathrm{d}s\right] \mu_k(x+t)\,\mathrm{d}t. \end{aligned} \tag{3}$$

For given risk intensities, $q_x^{(k)}$ can be computed by numerical integration in (3). The expected number of decrements at age x as a result of cause k is $d_x^{(k)} = l_x q_x^{(k)}$. When estimates are available for the cause-specific risk intensities, one or two columns can therefore be added to the life table for each cause to include estimates of $d_x^{(k)}$ and possibly $q_x^{(k)}$.

Several further life-table functions can be defined by formal reduction or elimination of one or more of the intensity functions in formulas like those above. In this manner, a *single-decrement life table* can be computed for each cause k, depicting what the normal life table would look like *if* cause k were the only one that operated in the study population and *if* it did so with the risk function estimated from the data. The purpose is to see the effect of the risk cause in question without interference from other causes. Some demographers call this abstraction the risk's *pure* effect. No assumption is made that in practice the total attrition risk can actually be reduced to the level of the one which is in focus or that this cause operates independently of other causes. For instance, a single-decrement life table of recovery from an

illness reflects the pure timing effect of the duration structure of the intensity of recovery even though the elimination of mortality is unattainable.

A single-decrement life table is at an extreme end of a class of tables produced by deleting one (or more) of the cause intensities in formulas like those above. To obtain a *cause-deleted life table*, where only cause k has been eliminated, one may introduce $\mu_{-k}(t) = \mu(t) - \mu_k(t)$,

$$q_x^{(-k)} = \int_0^1 \exp\left[-\int_0^t \mu_{-k}(x+s)\,\mathrm{d}s\right]\mu_{-k}(x+t)\,\mathrm{d}t$$
$$= 1 - \exp\left[\int_x^{x+1} \mu_{-k}(s)\,\mathrm{d}s\right], \qquad (4)$$

and so on, and a "normal" life table may be computed with $\mu(t)$ replaced by $\mu_{-k}(t)$ everywhere. A corresponding cause-deleted multiple-decrement life table may be based on reduced cause-specific decrement probabilities like

$$\int_0^1 \exp\left[-\int_0^t \mu_{-k}(x+s)\,\mathrm{d}s\right]\mu_j(x+t)\,\mathrm{d}t,$$
$$\text{for } j \neq k.$$

Such a table would show what a normal table would look like *if* it were possible to eliminate cause k without changing the risk of any other cause. Again no assumption needs to be made about the feasibility of such elimination in real life nor about cause independence. The computations are based on a pure abstraction. The interpretation for real-life applications must be based on substantive considerations and is a different matter.

Life Expectancy

An individual's **life expectancy** (at birth) is the expected value

$$\mathring{e}_0 = \mathrm{E}(T) = \int_0^\infty [1 - F(x)]\,\mathrm{d}x = \int_0^\infty \frac{l_x}{l_0}\,\mathrm{d}x$$

of his or her lifetime T, computed for the probability distribution $F(\cdot)$ operating at the time of birth. When the individual has survived to (exact) age x, his or her remaining lifetime, $U = T - x$, is positive and has the survival function $S_x(u) = S(x+u)/S(x) = l_{x+u}/l_x$, and the *residual life expectancy* is

$$\mathring{e}_x = \mathrm{E}(T - x | T > x) = \int_0^\infty S_x(u)\,\mathrm{d}u = \int_0^\infty \frac{l_{x+u}}{l_x}\,\mathrm{d}u.$$

If $L_x = \int_0^1 l_{x+t}\,\mathrm{d}t$, we get $L_x \cong \frac{1}{2}(l_x + l_{x+1})$ by the trapezoidal rule of numerical integration, and

$$\mathring{e}_x = \sum_{t=0}^\infty L_{x+t} \cong \sum_{t=0}^\infty \frac{l_{x+t}}{l_x} - \frac{1}{2}, \qquad (5)$$

which is normally used to compute values for $\mathring{e}_x$.

Equivalent names for the life expectancies are *mean survival time* for $\mathring{e}_0$ and *mean residual survival time at age x* for $\mathring{e}_x$. The *median length of life* is the median in the distribution of T; it used to be called the *probable length of life* (*see* **Median Survival Time**). Correspondingly, the *median residual length of life* at age x used to be called the *probable residual length of life*. If we denote the latter by ξ_x, then it is defined by the relation $l_{x+\xi_x} = \frac{1}{2}l_x$.

The above functions can be computed for cohort life tables and for period life tables. Figure 1 shows plots of the function $\mathring{e}_x$ according to the mortality experience for Swedish women in 1891–1900 and 1990–1994. The life expectancy at birth has increased from 53.6 years in the older table to 80.8 some one hundred years later. Note that in the older table $\mathring{e}_x$ increases with x up to age 2 and remains above $\mathring{e}_0$ up through age 11. When mortality is high at very young ages, surviving the first part of life *increases* your expected remaining lifetime. As a consequence of mortality improvements for very young children, these features have

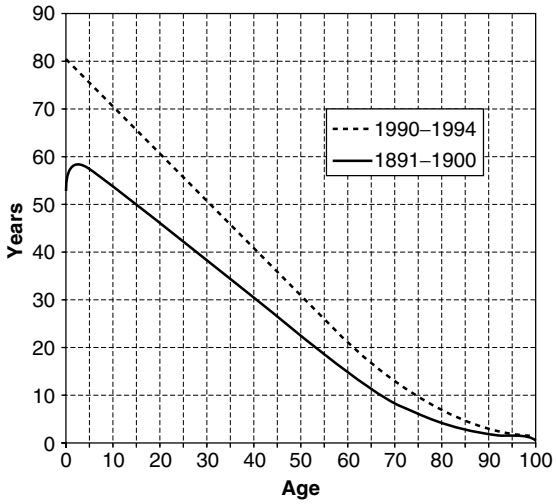

Figure 1 Residual life expectancy for Swedish women, 1891–1900 and 1990–1994

disappeared in the younger table. Note that the expected *total* lifetime, $x + \mathring{e}_x$, always increases with x throughout the human lifespan. (One can show that the derivative of this function is always positive.) The longer you have lived already, the longer you can expect the total length of your life to be.

In a multiple-decrement situation, formula (5) can be used to compute a residual life expectancy $\mathring{e}_x^{(-k)}$ from the decrement series of the cause-deleted life table for risk k. The difference $\mathring{e}_x^{(-k)} - \mathring{e}_x$ is the gain one would get in residual life expectancy at age x *if* it were possible to eliminate risk cause k without changing the risk intensity of any other cause of decrement. Dublin et al. [5, p. 96] note that according to the cause-specific mortality of the US in 1939–1941 the gain would be 9.01 years for white men and 8.80 years for white women at age 0 if one could eliminate the risk of death due to cardiovascular–renal diseases at all ages (and change no other cause-specific mortality risks). The gains from eliminating the risk of death in cancer alone were much less (1.39 years for men and 2.05 years for women).

History and Literature

The first step toward the development of the life table was taken when **Graunt** [9] published his famous *Bills of Mortality*. There were subsequent contributions by **Halley**, Huygens, Leibniz, Euler, and others. Deparcieux [4] clarified the definition of the life expectancy and identified the need for separate tables for men and women. Wargentin [17] was the first to publish real age-specific death rates, and the first to do so for a whole country. Price [14] included most of the columns now associated with the life table, and the tables by Duvillard [7] contained them all. The basic notions of cause-eliminated life tables go back to Bernoulli [1]. Cournot [3] developed the essentials of their mathematics. See Dupâquier [6] and Seal [15] for historical overviews. Smith & Keyfitz [16] have collected extracts from many original texts.

Life-table techniques are described in most introductory textbooks on the methods of actuarial statistics, biostatistics, demography, or epidemiology. See for example, Chiang [2], Elandt-Johnson & Johnson [8], or Manton & Stallard [13].

References

[1] Bernoulli, D. (1766). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Académie Royale des Sciences, Mémoires, Année 1760*, pp. 1–45.

[2] Chiang, C.L. (1984). *The Life Table and its Applications*. Krieger, Malabar.

[3] Cournot, A. (1843). *Exposition de la théorie des chances et des probabilités*. Hachette, Paris.

[4] Deparcieux, A. (1746). *Essai sur les probabilités de la durée de la vie humaine*. Guérin Frères, Paris.

[5] Dublin, L.I., Lotka, A.J. & Spiegelman, M. (1947). *Length of Life.* Ronald Press, New York.

[6] Dupâquier, J. (1996). *L'invention de la table de mortalité*. Presses Universitaires de France, Paris.

[7] Duvillard, E. (1806). Analyse des tableaux de l'influence de la petite vérole, et de celle qu'un préservatif tel que la vaccine peut avoir sur la population et la longévité. Paris.

[8] Elandt-Johnson, R.C. & Johnson, N.L. (1980). *Survival Models and Data Analysis*. Wiley, New York.

[9] Graunt, J. (1662). *Natural and Political Observations Made Upon the Bills of Mortality*. London.

[10] Heligman, L. & Pollard, J. (1980). The age pattern of mortality, *Journal of the Institute of Actuaries* **107**, 49–75.

[11] Hoem, J.M. (1972). Analytic graduation, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability 1970*, Vol. 1, L.M. Le Cam, J. Neyman & E.L. Scott, eds. University of California Press, Berkeley, pp. 569–600.

[12] Keyfitz, N. (1982). Keyfitz method of life-table construction, in *Encyclopedia of Statistical Sciences*, Vol. 4, S. Kotz & N.L. Johnson, eds. Wiley, New York, pp. 371–372.

[13] Manton, K.G. & Stallard, E. (1984). *Recent Trends in Mortality Analysis*. Academic Press, New York.

[14] Price, R. (1783). *Observations of Reversionary Payments: On Schemes for Providing Annuities for Widows, and for Persons in Old Age; and on the National Debt*. Cadell & Davies, London.

[15] Seal, H. (1977). Studies in the history of probability and statistics, XXV: multiple decrements or competing risks, *Biometrika* **64**, 429–439.

[16] Smith, D. & Keyfitz, N. (1977). *Mathematical Demography: Selected Papers*. Springer-Verlag, Heidelberg.

[17] Wargentin, P. (1766). *Mortaliteten i Sverige, i anledning af Tabell-Verket*. Kongl. Vetenskaps-Academiens Handlingar, Stockholm.

(*See also* **Demography**; **Vital Statistics, Overview**)

JAN M. HOEM