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 function $\mu(t) = f(t)/S(t)$, 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 $l_x = \Pr(T > x)$ 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 $q_x = \Pr(T \in (x, x+1)/T > x) = d_x/l_x$, and the corresponding survival probability $p_x = 1 - q_x$. Simple integration gives

$$q_x = 1 - \exp\left(-\int_x^{x+1} \mu(s) \, ds\right).$$  \hspace{1cm} (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 $\int_x^y \mu(s) \, ds$; (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

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

and the crucial first step in the life-table computation has been achieved. Instead of (2), $\hat{\mu}_x \left( 1 + \frac{1}{2} \hat{\mu}_x \right)$ 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
Life Table

possible to partition the first year of life into sub-
tervals in each of which mortality can be taken as
constant, then it is statistically more efficient essen-
tially to build up a life table for this year. This leads
to an estimate like \( \hat{q}_0 = 1 - \exp(-\sum \hat{\mu}_k) \), 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 \( h(x; \theta) \), where \( \theta \) is a vector
of parameters. Actuaries most often use the classical
Gompertz–Makeham function \( h(x; a, b, c) = a + bc \) 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 \( h(\cdot; \theta) \) 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 \[ \hat{\mu}_x = A^{x+B} + D \exp[-E(\ln x - \ln F)^2] + GH^x. \]

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 life-
times have the same cdf. Staggered (delayed) entries
into the study population and voluntary exits (with-
drawals) from it are permitted provided they contain
no information about the risk in question, be it death,
recurrence of a disease, or something else. Never-
theless, 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
(attribution) caused by the risk’s operation. In demog-
raphy, this corresponds to following a birth cohort
through life or a marriage cohort while their mar-
rriages 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 sick-
ess can end in death or, alternatively, in recovery.
Suppose that an integer random variable \( K \) repre-
sents the cause of decrement and define \( F_k(t) =
\Pr(T \leq t, K = k) \), \( f_k(t) = dF_k(t)/dt \), 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 \( \mu(\cdot) = \sum \mu_k(\cdot) \) is
the total risk of decrement at time \( t \). For the multiple-
decrement table, we define the decrement probability
\[
q_x^{(k)} = \Pr(T \leq x + 1, K = k|T > x)
= \int_0^1 \exp \left[ - \int_0^t \mu(x + s) \, ds \right] \mu_k(x + t) \, dt.
\] (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
\( d_x^{(k)} = 1_t q_x^{(k)} \). 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 $\mu_{-k}(t) = \mu(t) - \mu_k(t)$,

$$
d_{x}^{(-k)} = \int_{0}^{t} \exp \left[ - \int_{0}^{s} \mu_{-k}(x + s) \, ds \right] \mu_{-k}(x + t) \, dt
$$

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

$$
\int_{0}^{t} \exp \left[ - \int_{0}^{s} \mu_{-k}(x + s) \, ds \right] \mu_j(x + t) \, dt,
$$

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

$$
\hat{e}_0 = E(T) = \int_{0}^{\infty} [1 - F(x)] \, dx = \int_{0}^{\infty} \frac{l_x}{\hat{e}_0} \, dx
$$

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

$$
\hat{e}_x = E(T - x | T > x) = \int_{0}^{\infty} S_x(u) \, du = \int_{0}^{\infty} \frac{l_{x+u}}{l_x} \, du.
$$

If $L_x = \int_{0}^{1} \frac{l_{x+t}}{l_x} \, dt$, we get $L_x \cong \frac{1}{2}(l_x + l_{x+1})$ by the trapezoidal rule of numerical integration, and

$$
\hat{e}_x = \sum_{t=0}^{\infty} L_{x+t} \cong \sum_{t=0}^{\infty} \frac{l_{x+t}}{l_x} - \frac{1}{2},
$$

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

Equivalent names for the life expectancies are mean survival time for $\hat{e}_0$ and mean residual survival time at age $x$ for $\hat{e}_x$. 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 $\hat{t}_x$, then it is defined by the relation

$$\hat{t}_x = \frac{1}{2} l_x.$$

The above functions can be computed for cohort life tables and for period life tables. Figure 1 shows plots of the function $\hat{e}_x$ 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 $\hat{e}_x$ increases with $x$ up to age 2 and remains above $\hat{e}_0$ 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](image-url)
disappeared in the younger table. Note that the expected total lifetime, \( x + \hat{\epsilon}_x \), 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 \( \hat{\epsilon}_x^{(-k)} \) from the decrement series of the cause-deleted life table for risk \( k \). The difference \( \hat{\epsilon}_x^{(-k)} - \hat{\epsilon}_x \) 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**


(See also Demography; Vital Statistics, Overview)