

Demography 101: Life Table Basics

The *life table* is a tool for describing levels and age patterns of mortality in a population using a standard set of descriptive statistics. Mastery of life table concepts and methods is essential for serious analysis of mortality. The importance of life table methods is more general, however, because the ideas that apply to the analysis of length of life apply in many other settings.

Life tables may present a stumbling block even to serious and determined students of demography. It is worth considering why this is so, and several possibilities may be suggested. One is the sheer volume of material to be absorbed. Reasonable mastery requires weeks of effort, not hours or days.

The notation used for life table columns and formulas requires a familiarity with basic mathematics that comes more easily for some than for others. A solid understanding of the distinction between period and cohort life tables requires experience using life table methods to study mortality in actual human populations.

A realistic appreciation of the effort required to learn life table methods may be the best way to avoid discouragement. That said, let's begin.

1. Cohort and period life tables

A *cohort* life table describes the mortality experience of a birth cohort, a group of persons born during a particular year or other time period. A cohort life table cannot be constructed until all or nearly all of the persons in the cohort have died. This means, firstly, that the persons must have been born a very long time ago, and secondly, that the mortality risks described by the table amalgamate mortality conditions over a very long period of time, nearly a century.

If we are interested in mortality during particular year or time period, we want a "period" life table. A *period* life table looks just like a cohort life table, but instead of describing what happened in a birth cohort over the past 100 or so years, it describes the possible *future* mortality of the cohort born during the period to which the life table refers, this on the assumption that persons in this cohort experience the mortality risks described by the life table throughout their lives.

Most life tables are period life tables, partly because few countries have the data required to construct cohort life tables, and partly because there is more interest in current than in past mortality.

The importance of cohort life tables is mainly conceptual. Life table concepts are more easily explained and understood for cohort life tables. Learning how cohort life tables are defined and constructed is the best preparation for learning period life tables.

The following section presents an example of a life table. Like most life tables encountered in practice, it is a period table, but to explain the meaning of the table the reader is asked to imagine that it is a cohort life table constructed from a list of the ages at death of persons born during a year in the distant past.

Table 1. Life table for Austrian Males, 1992

x	n	${}_n m_x$	${}_n a_x$	${}_n q_x$	l_x	${}_n d_x$	${}_n L_x$	T_x	e_x
0	1	0.008743	0.068	0.008672	1.00000	0.00867	0.99192	72.8891	72.889
1	4	0.000370	1.626	0.001479	0.99133	0.00147	3.96183	71.8972	72.526
5	5	0.000153	2.500	0.000765	0.98986	0.00076	4.94742	67.9354	68.631
10	5	0.000193	3.143	0.000965	0.98910	0.00095	4.94375	62.9880	63.682
15	5	0.000976	2.724	0.004869	0.98815	0.00481	4.92980	58.0442	58.740
20	5	0.001285	2.520	0.006405	0.98334	0.00630	4.90108	53.1144	54.014
25	5	0.001135	2.481	0.005659	0.97704	0.00553	4.87128	48.2133	49.346
30	5	0.001360	2.601	0.006778	0.97151	0.00658	4.84176	43.3420	44.613
35	5	0.001882	2.701	0.009369	0.96493	0.00904	4.80385	38.5003	39.900
40	5	0.002935	2.663	0.014575	0.95589	0.01393	4.74687	33.6964	35.251
45	5	0.004849	2.698	0.023977	0.94195	0.02259	4.65778	28.9496	30.733
50	5	0.007133	2.676	0.035083	0.91937	0.03225	4.52189	24.2918	26.422
55	5	0.011263	2.645	0.054860	0.88711	0.04867	4.32096	19.7699	22.286
60	5	0.018600	2.624	0.089064	0.83845	0.07468	4.01481	15.4489	18.426
65	5	0.028382	2.619	0.132927	0.76377	0.10153	3.57713	11.4341	14.971
70	5	0.041238	2.593	0.187572	0.66225	0.12422	3.01224	7.8570	11.864
75	5	0.071634	2.518	0.304102	0.53803	0.16362	2.28404	4.8448	9.005
80	5	0.112324	2.423	0.435547	0.37441	0.16307	1.45182	2.5607	6.839
85	∞	0.190585	5.247	1.000000	0.21134	0.21134	1.10889	1.1089	5.247

Source Preston, Heuveline, and Guillot [3], page 49. The life table columns were calculated from the age-specific death rates and ${}_n a_x$ values in the source using the formulas given in the source and the parameters in Table 3.3. Minor discrepancies between the numbers shown here and the numbers in the source are due to rounding error.

2. Life table for Austrian males, 1992

Table 1 shows a life table for Austrian males based on mortality risks observed in 1992. Understanding the table requires a few definitions. A person’s *age* at any given time is the time elapsed since the person’s birth. *Age in completed years* (or “age at last birthday”) is the greatest integer less than age.

Age may be referred to as *exact age* when it is unclear whether age or age in completed years is intended. In practice, “age” tends to be used without qualification, the reader being expected to figure out from context what meaning is intended. This is almost always possible, and not difficult, but it requires experience and practice.

Rows of the table correspond to age intervals. The number in the x column is the (exact) age at which the age interval begins. The number in the n column is the width of the interval. The age intervals are understood to be what mathematicians call “left closed, right open”, meaning that they include persons whose exact age is x but exclude persons whose exact age is $x + n$. The last interval is nearly always “open-ended”, in this case age 85 years and over.

The subscript notation in the column headings was invented by actuaries and has been in use for well over a century. The following subscript x denotes exact age. If the values in the column refer to an age interval, the leading subscript, sometimes called a “prescript”, gives the width of the interval. The prescript is omitted if the width is one year. Putting the width of the open-ended interval to ∞ is a convenient formalism.

The ${}_n m_x$ column is not strictly a life table column, rather it is the primary input from which most period life tables are constructed. The ${}_n a_x$ column, used in section 5 and discussed in detail in section 7, provides ancillary information used to construct the table. These two columns may shown or not, depending on how the life table

was constructed and on how much detail the presenter choses to show.

The remaining six columns of Table 1 tend to appear in all life tables. They are discussed in the following three sections.

3. The l_x and ${}_n d_x$ columns

To learn the meaning of the life table columns it is useful to imagine that Table 1 is a cohort life table for persons born during a year in the distant past and was constructed from a list of the exact ages at death of all persons born during this year. The ${}_n m_x$ column is not needed if the life table is constructed in this way.

The ${}_n q_x$, l_x , and ${}_n d_x$ columns are defined as follows. The l_x and ${}_n d_x$ columns are discussed in this section, the ${}_n q_x$ column in the following section.

$$l_x = \frac{\text{Number of persons surviving to exact age } x}{\text{Number of persons in cohort}} \quad (1a)$$

$${}_n d_x = \frac{\text{Number of persons dying between exact age } x \text{ and exact age } x + n}{\text{Number of persons in cohort}} \quad (1b)$$

$${}_n q_x = \frac{\text{Number of persons dying between exact age } x \text{ and exact age } x + n}{\text{Number of persons surviving to exact age } x} \quad (1c)$$

If we were really calculating a cohort life table from a list of ages at death, we would

1. tally the number of deaths in each age interval and divide by the number of persons in the cohort to get the ${}_n d_x$ column values,
2. cumulate the ${}_n d_x$ column upward from the last row to get the l_x column values, and
3. divide the ${}_n d_x$ values by the l_x values, row by row, to get the ${}_n q_x$ column values.

If the life table were indeed constructed in this way it would be natural to put the columns in this order.

The different order of the columns in Table 1 is due to its being, in fact, a period life table. As explained in section 8, construction of a period life table begins with calculation of the ${}_n q_x$ from age-specific death rates ${}_n m_x$ followed by calculation of the l_x and ${}_n d_x$ columns. This explains the order of the columns in Table 1.

The ${}_n d_x$ and l_x columns will be familiar to students of introductory statistics. The ${}_n d_x$ column gives the frequency distribution of age at death for members of the birth cohort. The number in the row for given x and n is the proportion of persons in the cohort for whom $x \leq x_i < x + n$.

The l_x column is the cumulative frequency distribution, except that the direction of cumulation is reversed: instead of the proportion of persons dying before age x , we have the proportion surviving to age x , which is the proportion dying after age x . All persons “survive to age 0”, so $l_0 = 1$.

The solid dots in Figure 1 show the frequency distribution of deaths (${}_n d_x$) in Table 1. A common procedure for plotting values that refer to intervals is to plot the point

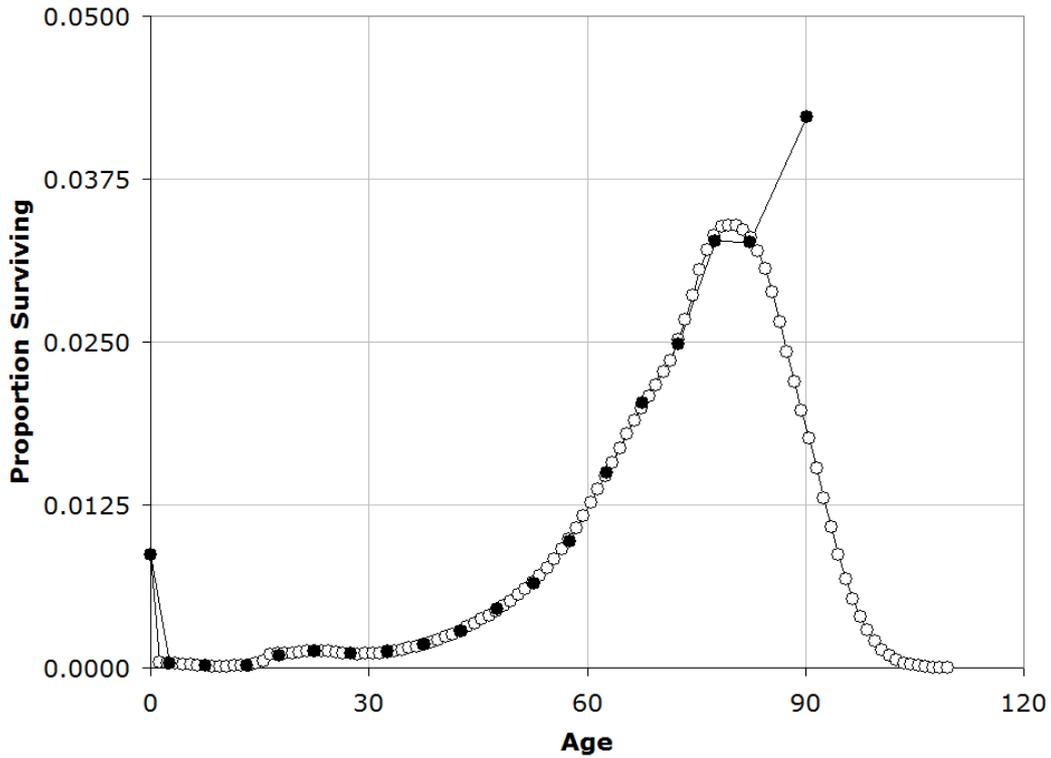


Figure 1. Frequency distribution of deaths ${}_n d_x$
Source Table 1. **Note** See text for discussion.

at the midpoint of the interval. A slightly refined alternative is to plot the point for an interval at the mean of the values in the interval. This has the advantage of providing an age at which to plot values for the open-ended interval, for which the midpoint is undefined. For the distribution in Table 1, the means of the values in each age interval are given by $x + {}_n a_x$. The ${}_n d_x$ in Figure 1 are plotted against these ages.

Figure 1 shows that many deaths occur at ages considerably above 85 years, far too many to justify truncating the table at this age. Life tables that truncate mortality experience too early are common, partly because conventional upper age limits were adopted long ago when expectations of life were lower.

The hollow circles in Figure 1 show d_x values from a life table constructed by extrapolating the age-specific death rates (${}_n m_x$) in Table 1 to older ages by fitting the *Human Mortality Database Log Quadratic* model [4] and then interpolating to single years of age. These points suggest that an open-ended group beginning at no less than 100 years of age is necessary to get a satisfactory picture of the frequency distribution of deaths.

4. The ${}_n q_x$ column

The distribution of ages at death of members of a birth cohort is distinctive in that deaths occur at particular times, as well as at a particular ages. For the study of mortality, the time is as important as the age. This characteristic of the distribution of cohort deaths by age is not shared by frequency distributions in general, and it is essential to understanding the meaning of the ${}_n q_x$ values.

The formula defining ${}_n q_x$, (1c), shows that its numerator is the same as the nu-

erator of ${}_n d_x$, the number of deaths in the age interval $[x, x + n)$. The difference between ${}_n q_x$ and ${}_n d_x$ is the denominator, all persons in the cohort for ${}_n d_x$ versus persons surviving to age x for ${}_n q_x$.

To see the significance of this difference, suppose for simplicity that persons in the cohort are all born at the same time t , so that deaths in the age interval $[x, x + n)$ occur between time $t + x$ and time $t + x + n$. What does ${}_n d_x$ tell us about the level of mortality for this age group during this time period?

The answer turns out to be—*nothing*. The value of ${}_5 d_{60}$ (say) may be low because most cohort members have died at earlier ages (high mortality), or because most cohort members die at older ages (low mortality). The ${}_n d_x$ values taken individually are not in general useful indicators of the level of mortality.

Dividing the number of deaths to the cohort in the age interval $[x, x + n)$ by the number of survivors at age x controls for the history of mortality in the cohort by eliminating the dependence on the numbers of deaths below age x . The value of ${}_n d_x$ cannot be larger than the number of persons surviving to age x . The value of ${}_n q_x$ may be any number between 0 and 1, whatever the number of deaths before age x , provided only that there are *some* survivors at age x . This is one way of understanding the significance of ${}_n q_x$.

Another way is to ask what is the maximum number of people in the cohort who *could* die in the interval $[x, x + n)$. The most fundamental principle of demographic analysis is that numbers of events tend to be larger when the numbers of persons who might experience these events are larger. Dividing the number of deaths by the number of persons who might have died gives a relative number that controls for *large* differences in numbers of persons. The influence of small differences in numbers of persons may be outweighed by other influences on numbers of deaths.

The number of persons in the cohort who die in the interval $[x, x + n)$ is ${}_n d_x$ times the number in the cohort. The number who *might* die is the proportion l_x who survive to age x times the number of in the cohort. Dividing the first number by the second gives the definition in formula (1c). This is a second way of understanding ${}_n q_x$.

A third way to understand ${}_n q_x$ is as a conditional probability. By definition, the conditional probability of event A given event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2)$$

Let A be the event “died in the age interval $[x, x + n)$ ” and B the event “survived to age x ”, which is the same as “died in the interval $[x, \infty)$ ”. The intersection of these two events is death in the interval $[x, x + n)$, so ${}_n q_x$ is the conditional probability of dying in the interval $[x, x + n)$ given survival to age x .

To personalize this, suppose that I am a cohort member, that I have just reached age x , and that I want to know how likely it is that I will die before reaching age $x + 5$ years. The ${}_n q_x$ column of the life table tells me. If I am 30 years old, the chance of dying before age 35 is 0.7%. If I am 70 years old, the chance of dying before age 75 is nearly 20%—this of course on the assumption that my mortality risks are represented by the life table shown in Table 1.

Given any one of ${}_n q_x$, l_x , and ${}_n d_x$ columns, the remaining two columns may be

calculated. This may be demonstrated by elementary algebra or by thinking through how one would get from any one of the columns to the other two. Given the ${}_n d_x$ column, for example, cumulating from the bottom gives the l_x column and dividing ${}_n d_x$ by l_x gives ${}_n q_x$. The information in the three columns is redundant in this narrow sense. As descriptive statistics, the columns provide different and useful information.

Note finally that the ${}_n q_x$, l_x , and ${}_n d_x$ columns do not depend on the age distribution of deaths within age intervals. The ${}_n L_x$, T_x , and e_x columns do depend on the age distribution of deaths within age intervals. This dependence is the basis for dividing the six basic life table columns into these two groups.

5. The ${}_n L_x$, T_x , and e_x columns

The values in these columns are defined as follows.

$${}_n L_x = \frac{\text{Person years lived by cohort members in the interval } [x, x+n)}{\text{Number of persons in the cohort}} \quad (3a)$$

$$T_x = \frac{\text{Person years lived by cohort members after age } x}{\text{Number of persons in the cohort}} \quad (3b)$$

$$e_x = \text{Average years of life remaining for a person at age } x = \frac{T_x}{l_x} \quad (3c)$$

The ‘‘Person years lived’’ in the numerator of (3a) is the sum of years lived in the age interval by each person who survives to age x . This sum may be divided into two components.

(a) Persons who survive to age $x+n$ contribute n person years to the sum. Multiplying the number of these persons by n and dividing by the total number of persons in the cohort gives $n \cdot l_{x+n}$.

(b) Persons who die at age $x \leq x_i < x+n$ contribute $x_i - x$ person years to the sum. Multiplying the number of these persons by the average of the $x_i - x$ values, ${}_n a_x$, and dividing by the total number of persons in the cohort gives ${}_n a_x \cdot {}_n d_x$.

Combining these two expressions gives

$${}_n L_x = n \cdot l_{x+n} + {}_n a_x \cdot {}_n d_x. \quad (4)$$

This formula may be used to calculate the ${}_n L_x$ column from the l_x column if the ${}_n a_x$ column is available (noting that ${}_n d_x = l_x - l_{x+n}$).

Person years lived is additive over age intervals, so the T_x column may be calculated from the ${}_n L_x$ column using

$$T_x = {}_n L_x + {}_{n+1} L_{x+n} + \dots, x \geq 5 \quad (5a)$$

$$T_0 = L_0 + T_1 \quad (5b)$$

$$T_1 = {}_4 L_1 + T_5 \quad (5c)$$

The value of $e_x = T_x/l_x$ is the *expectation of life at age x* —the average number of years persons who have survived to age x live beyond this age. For $x = 0$, this becomes *expectation of life at birth* or *life expectancy at birth*. Life expectancy at birth for the cohort is the same as the mean age at death of cohort members.

6. How the life table got its columns

Readers familiar with introductory statistics may be puzzled by the definition of life expectancy at birth in the preceding section. If it is simply the mean age at death for cohort members, why not calculate it using one of the formulas for the mean found in introductory statistics textbooks? Why “person years lived” and the cumbersome and apparently unnecessary three column calculation?

The answer may surprise readers who grew up with personal computers and spreadsheet programs. Life tables go back more than 350 years, beginning with John Graunt’s *Natural and Political Observations Made Upon the Bills of Mortality*, a study of statistics of deaths in contemporary London [1]. The life table in its modern form is newer, but the columns in Table 1 go back more than a century [2].

This means that life tables were developed when arithmetic was done using pencil and paper—no calculating machines, to say nothing of computers. When arithmetic is done by hand, addition is simplest and least error-prone. Subtraction, to say nothing of multiplication and division, is more difficult and to be avoided if at all possible.

Calculations are therefore organized to minimize the use of subtraction, multiplication and division. This explains the seemingly peculiar manner in which life expectancy is calculated. Manual calculation also explains the layout of the life table, the first purpose of which was not to present results, but to organize the calculations in a way that minimizes the chance of careless error.

The difficulties of manual arithmetic explain the origin of the life table columns, but two other factors explain their persistence. Insurance and annuity calculations use e_x values at all ages, and if these are wanted, the traditional organization of their calculation is as good as any other. It has the advantage, moreover, of providing ${}_nL_x$ values, which turn out to be useful in connection with stationary populations and survivorship problems.

Life expectancy at birth and mean age at death *for a cohort* are two expressions for the same quantity, but “mean age at death” for a *period* might reasonably be understood as the mean age at which deaths to the population during a time period. This mean reflects the population age distribution as well as the age-specific risks of mortality and is an extremely poor indicator of mortality. “Expectation of life at birth” avoids this confusion.

7. The ${}_na_x$ column

The ${}_na_x$ column shows the average years lived in the age interval $[x, x + n)$ by persons who die in this interval. It was used in section 5 to calculate ${}_nL_x$ from l_x using formula 4.

Substituting $l_x - l_{x+n}$ for ${}_nd_x$ in formula 4 and rearranging terms gives

$${}_n a_x = \frac{{}_n L_x - n \cdot l_x}{l_x - l_{x+n}}. \quad (6)$$

If a life table shows all three columns, this formula provides a check of the consistency between them. If a life table does not show ${}_n a_x$ values, they may be calculated from the ${}_n L_x$ and l_x values.

Formula 4 also implies

$$l_{x+n} = \frac{{}_n L_x - (n - {}_n a_x) \cdot l_x}{{}_n a_x}. \quad (7)$$

Since $l_0 = 1$, this shows that the l_x column may be iteratively calculated from the ${}_n L_x$ and ${}_n a_x$ columns.

Formulas 4, 6, and 7 together show that the information in the l_x , ${}_n L_x$, and ${}_n a_x$ columns is redundant in the sense that any one of these columns may be calculated from the other two.

8. Period life table construction

Preceding sections discussed the life table columns in Table 1 as though they were calculated from a list of the ages at death of a cohort of persons born in the distant past, so distant that all of the cohort members have died. This pedagogical device facilitates understanding of the meaning of the columns and their interrelationships.

As noted in the introduction, however, most life tables encountered in practice are period life tables. They show what *would* happen to a cohort of persons born during a recent time period if these persons experienced the mortality risks observed during the period in which they were born. The idea is not that mortality risks will remain in fact constant, but that this assumption provides a useful representation of current mortality risks.

Construction of period life tables involves many technicalities. The methods used vary with the data available. A thorough discussion is far beyond the scope of this letter. It will be useful, however, to present a generic procedure for calculating a period life table from a set of age-specific death rates.

Current mortality risks are represented by ***age-specific death rates***, defined as the number of deaths of persons in an age interval $[x, x+n)$ during a time period divided by the person years lived by persons in the age group during the time period. The ${}_n m_x$ column in Table 1 shows age-specific death rates for Austrian males for 1992.

The simplest way to calculate a period life table is to use a formula that translates age-specific death rates into ${}_n q_x$ values. The ${}_n q_x$ column values are then used to calculate the l_x column values, and the l_x column values are used to calculate the ${}_n q_x$ column values.

Calculation of the ${}_n L_x$, T_x , and e_x columns requires information on the age distribution of deaths within age intervals. This is not provided by the ${}_n q_x$, l_x or ${}_n d_x$ columns and so must be exogenously supplied. A set of ${}_n a_x$ values, calculated from available data or taken from another life table, may be used for this purpose (see the discussion in section 3.2 of [3]).

Given a set of age-specific death rates ${}_n m_x$ and a set of ${}_n a_x$ values, ${}_n q_x$ values may be calculated as

$${}_n q_x = \frac{n \cdot {}_n m_x}{1 + (n - {}_n a_x) \cdot {}_n m_x}. \quad (8)$$

A derivation of this formula is given in section 3.1 of [3].

References

- [1] Guillot, Michelle. “Life tables” entry in Volume 2 of the *Encyclopedia of Population*, Ed. Paul Demeny and Geoffrey McNicoll. New York: Macmillan Reference USA.
- [2] King, George. 1902. Statistical applications of the mortality table. Excerpted in David P. Smith and Nathan Keyfitz, *Mathematical Demography: Selected Papers*, second, revised edition, Ed. Kenneth W. Wachter and Herv Le Bras. Demographic Research Monographs: A Series of the Max Planck Institute for Demographic Research. New York: Springer. Available at www.demogr.mpg.de/books/, accessed 13 July 2016.
- [3] Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2000. *Demography: Measuring and Modeling Population Processes*. Oxford, United Kingdom: Blackwell Publishers Inc.
- [4] Wilmoth, John R., Sarah Zureick, Vladimir Canudas-Romo, Mie Inoue, and Cheryl Sawyer. 2012. A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies* **66**(2):1-28.

DSITL is an occasional email letter on demography, statistics, and information technology by Griffith Feeney. It is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License](https://creativecommons.org/licenses/by-nc-nd/3.0/). To subscribe (unsubscribe), send email to ‘feeney at gfeeney dot com’ with “subscribe” (“unsubscribe”) in the subject line.