Life Tables in R using the tidyverse

Monica Alexander

Contents

1 Columns of the lifetable 2

1.1 Survivorship lx . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

1.2 Deaths ndx . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.3 Probability of dying, nqx, and of surviving, npx . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1.4 Average years lived, nax . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1.5 Person-years lived, nLx . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1.6 Person-years lived above age x, Tx . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

1.7 Life expectancy at age x, ex . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

2 Period life tables 9

2.1 Construction from period mortality rates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.2 Getting values for nax . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.3 Interpretation of period life table measures . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

3 R: Make your own life table 11

Life tables are a fundamental tool in demography. A life table describes the mortality experiences for a certain population. Usually a life table is composed of sets of values showing the mortality experience of a hypothetical group of infants born at the same time and subject throughout their lifetime to the specific mortality rates of a given year. Life tables are how we calculate life expectancy, probably one of the most common mortality summary measures. They are useful to compare populations and also tells us something about the implied stationary population. Each column refers to a different measure of survivorship. There are different ways of describing survivorship: for example, probability still alive, life expectancy, etc, so a life table has many different columns. This module explains the main columns of a lifetable and demonstrates how to construct a lifetable in R using the tidyverse syntax. Let’s load in the packages we need:



1 Columns of the lifetable

**1.1 Survivorship lx**

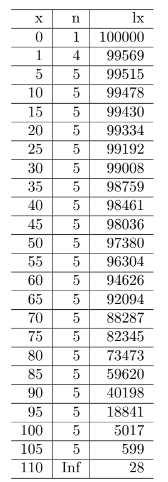
Every row of a life table refers to a different age or age group: if the later, the table is referred to as an abridged life table. We will define x to be age and n to be the length of the interval.

A usual place to start is survivorship, lx, which is defined as the number of people still left alive at age x. The value of l0 is the starting size of the population, and is called the radix. In practice, the radix is usually equal to 1, 100, or 100,000. If l0 = 1 then then lx is a probability of survival to age x. Note that for now we are implicitly assuming this l0 relates to a cohort of people moving through time, so the life table documents cohort mortality. However, later on we will look at period mortality.

Here’s the estimated lx values for females in Ontario in 2015. The data are from the Canadian Human Mortality Database. Here, the radix is 100,000. By age 110, out of the original population of 100,000, it is estimated that 28 will survive.



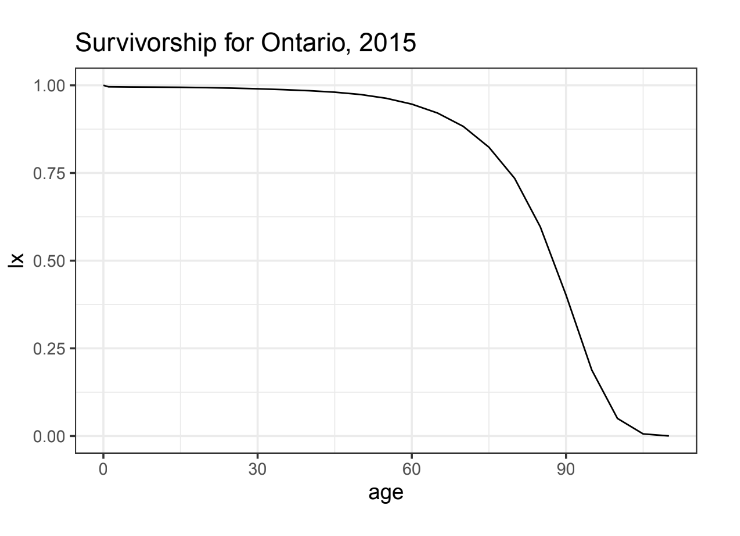
Let’s also plot the lx and divide through by 100,000 so the lx can be interpreted as the proportion of the population surviving at age x.



Let’s also plot the lx and divide through by 100,000 so the lx can be interpreted as the proportion of the

population surviving at age x.





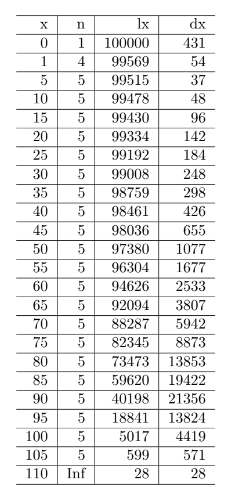
**1.2 Deaths ndx**

The next column, ndx, is the number of deaths between ages x and x + n. Note that we using the ‘duration-age’ notation, because it refers to deaths over an interval. In contrast, lx refers to survivors at a particularage x.

By definition, the number of deaths over an interval must be the number of survivors at the start of the interval, minus the number of survivors at the end, i.e.

ndx = lx − lx+n.

Let’s look at the estimated ndx values for Ontario in 2015:



Notice the structure of the life table in terms of where the the values of lx and ndx line up within the rows. lx always starts at the radix, so the first row represents the total population before any deaths. The first row of ndx represents the deaths in the first interval. So the second row of lx is equal to the previous row of lx minus the previous row of ndx, and so on. Note also the last interval: everyone who survived to the last age group must die.1

**1.3 Probability of dying, nqx, and of surviving, npx**

The next column, nqx is the probability of dying between ages x and x + n. Note that this is a conditional probability, so it’s the probability of dying in that interval given you survived to age x. nqx can be calculated as

nqx = ndx/lx

The complement of nqx is the probability of survival, npx

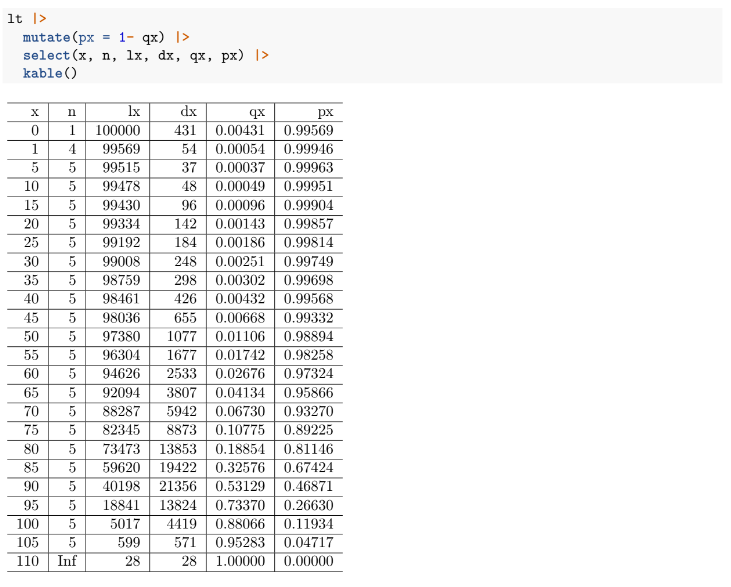
npx = 1 − nqx

Again, this is a conditional probability, so it’s the probability of survival between ages x and x + n given you survived to age x. Given the relationship for nqx and ndx, we can also calculate npx as

npx = 1 − nqx = 1 − (ndx/lx) = (lx − lx + lx+n)/lx = (lx+n)/lx

i.e. the probability of survival is the ratio of the the survivors at the end and start of the interval.

Looking again at the data for Ontario in 2015, notice the probability of death in the last age group is 1, because again, everyone must die eventually.



**1.4 Average years lived, nax**

nax is the number of years lived by those who died between ages x and x + n. So for example, if 1a0 = 0.25, then for that population, those infants who died in the first year on average died after 0.25 years = 3 months. To calculate the exact value for nax requires a lot of data: you would need to know the exact lengths of life for each individual in the cohort. Approximations to nax are discussed below in the period life table section.

**1.5 Person-years lived, nLx**

nLx is the number of person-years lived between ages x and x + n. The total number of person-years lived (PYL) in an interval is the sum of

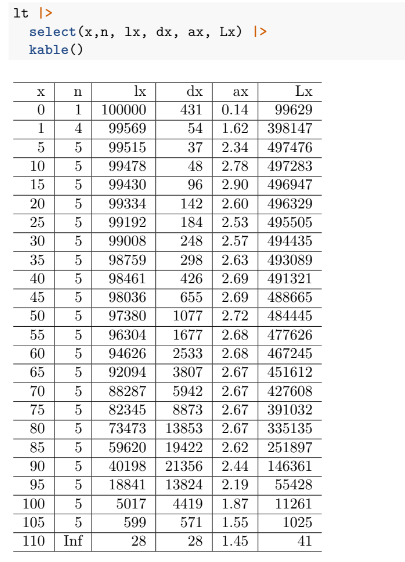
1. the PYL by those who survived and

2. the PYL by those who died in the interval.

The first piece is just the interval length: n multiplied by the number of survivors at the end of the intervals, lx+n. The second piece is the average time spent alive in the interval by those who died, nax multiplied by the number of people who died in the interval, ndx. So:

nLx = n · lx+n +n ax ·n dx.

Note for the last interval there are no survivors, so ∞Lx = ∞ax · ∞dx. Adding nax and nLx to the Ontario life table:

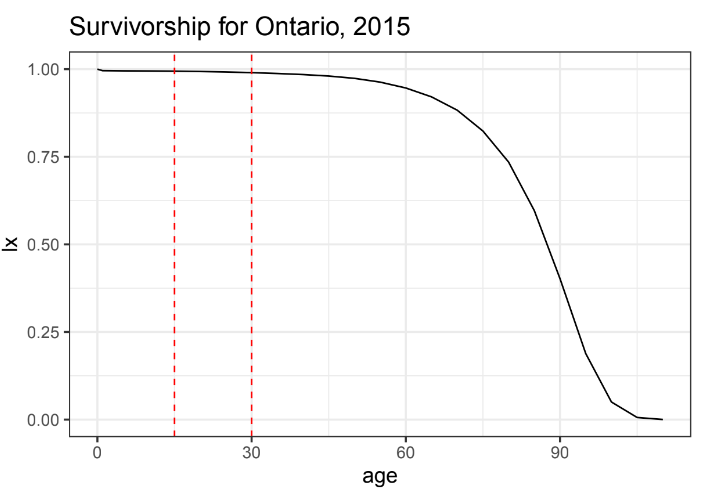


1.5.1 nLx graphically and the relationship with lx

nLx is essentially the number of survivors times the average length of time they survived in a particular interval. For a given radix l0 and interval length n, the maximum nLx could be is l0 ·n, if everyone survived.

How does nLx relate to lx? It is the area under the lx curve for the interval [x, x + n], as illustrated below by the red dashed lines, for 15L15. It may help to think about the units here: nLx has units person-years. lx has units of persons. The x-axis on the graph below has units years.



With this in mind, we can represent nLx in continuous form as



In practice we usually have to calculate nLx in the discrete form, but it’s often useful to think about it in continuous form, i.e. the area under the survivorship curve.

**1.6 Person-years lived above age x, Tx**

Whereas nLx is the person-years lived in a specific interval, Tx is the person-years lived above a specific age x (so notice it does not have the duration/age notation). It is defined as the sum of the relevant nLx:



In a similar fashion to nLx, Tx can be thought of as the area under the lx curve above age x. In addition, We can write Tx in continuous form as

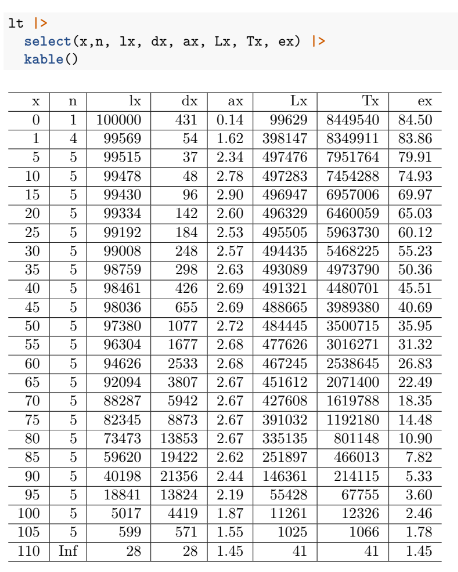


1.7 Life expectancy at age x, ex

The final column we will introduce for now is probably the most well-known: ex, the average number of remaining years of life for those who reach age x, or the life expectancy at age x. Note that the ‘expectancy’ terminology is related to the expected value in the statistical sense. ex is calculated as



We can do a quick check of the units here to make sure it makes sense: Tx has units person-years, lx has units persons, so ex has units of years. You are probably most familiar with life expectancy at birth, e0. Note that it is again a conditional measure, that is, it’s the average number of remaining years given a person has already survived to age x. As such, the value of ex need not decrease monotonically over age. In practice it usually does, unless infant mortality is relatively high. The filled-in life table with all columns discussed:



**2 Period life tables**

A life table as defined above refers to tracking the mortality of a cohort of people as they age. However, it is often not practical or useful just to consider the mortality of a cohort, because in order to to build a complete table, we have to wait to observe everyone in the cohort die. So for the 1990 birth cohort, for example, we would probably have to wait around until at least 2090 before a reasonable cohort life table could be built. This is not very useful to study current mortality conditions.

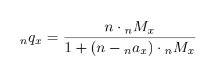
In addition to constructing cohort life tables, we can construct period life tables. They refer to the period in the sense that they are constructed using mortality conditions in a particular period. This means the lifetable refers to a synthetic cohort, a hypothetical group of people that experience the mortality conditions of the period of interest throughout their entire life. Why is this hypothetical and potentially unrealistic? Because mortality conditions change over time (and in general, are getting better). So for example, if I live until I’m 70, the mortality conditions I am subject to in the future are likely to be different to the mortality conditions that a current 70-year-old is being subjected to. However, period life tables are still useful to compare mortality outcomes for different populations in a more up-to-date way.

**2.1 Construction from period mortality rates**

The key to constructing period life tables is converting the observed period mortality rates nMx to probabilities of death, nqx. The mortality rate is the number deaths divided by person years lived, so in life table notation this is:



We can use the following nMx to nqx conversion formula to get the nqx column, after which all other columns can be derived based on the relationships discussed above (and choosing a radix). The formula is:



How did this come about? By rewriting



rearranging to get lx and rewriting nqx with this denominator.

**2.2 Getting values for nax**

The conversion formula requires information only on period mortality rates and values of the average number of years lived for those who died, nax. How do we get these values? As mentioned above, the data required to calculate nax exactly is usually not available, so we need to approximate it somehow. Preston, Heuveline and Guilot (2000) has a good overview of the options here (section 3.2, page 44), but the most common and easiest approach is, for most age groups, assume:

nax = n/2

That is, on average those who die, die half-way through the interval. So for an abridged life table with five-year intervals, nax = 2.5. This assumption is fine for most age groups, except for the very young and very old ages.

At younger ages, typically in abridged life tables the first five years is split out into the first year, and years 1-5, so we need values for 1a0 and 4a1. For mortality at younger ages, we expected comparatively more deaths to occur at the start of the interval, so nax < n/2. The following two approximations are common, and used e.g. by Wachter in Essential Demographic Methods (2014):

1a0 = 0.07 + 1.71M0

and

4a1 = 1.5

The equation for 1a0 says that infants who die on average die at about 1 month plus a bit, where the bit depends on the over level of infant mortality. This is based on the observation that as infant mortality declines, infants that are dying are more likely to die from pre-existing conditions rather than exogenous factors, so deaths occur relatively early.

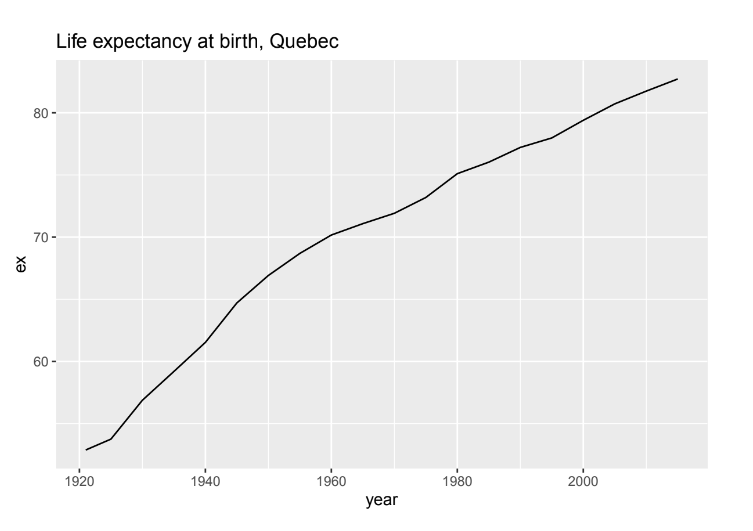
For the last age group, we can assume nax is the inverse of the mortality rate:

∞aω = 1/∞Mω

where ∞Mω is the age-specific mortality rate for the last age interval; ω refers to the last age.

The smaller the interval of the open-ended age group, the better approximation.

**2.3 Interpretation of period life table measures**

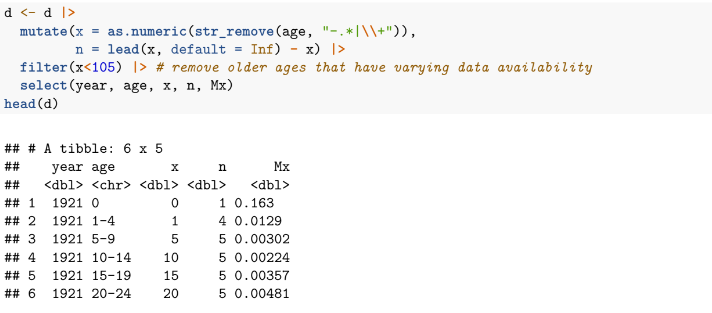
As mentioned atthe start of this section, period life tables are constructed from a synthetic cohort of people that hypothetically would go through life experiencing the age-specific mortality conditions of the current period. Period life table measures, and in particular, period life expectancy at birth, are the most commonly published and discussed. However, period measures of life expectancy are often misinterpreted. The technical definition is the expected number of years of life for a newborn who would be subject to the current mortality conditions for their entire life. But life expectancy is usually just talked about as ‘how long you’re expected to live for’. Part of the confusion comes from the name, but the ‘expected’ part refers to the fact that it is an expected value in the statistical sense; in particular, E[Tx] = ex.

**3 R: Make your own life table**

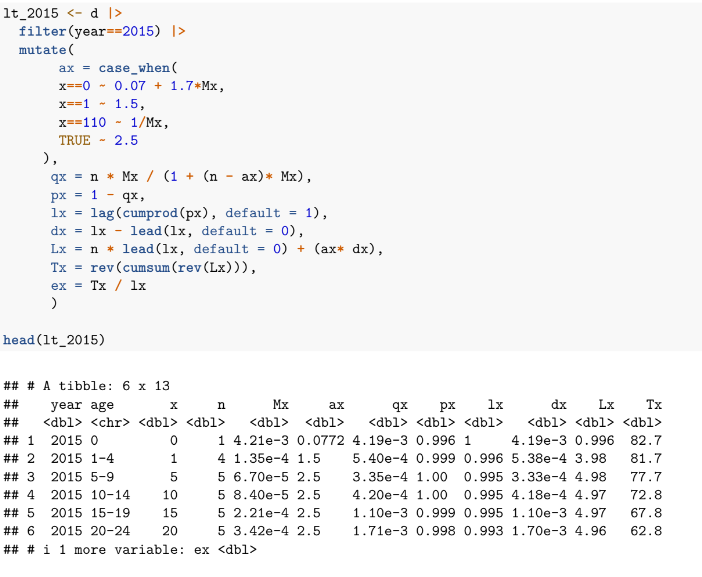
Let’s make a period life table for females in Quebec in 2015 using data from the Canadian Human Mortality Database. Read in the data directly from the website, and filter out what we need (the variables year, age, and total):



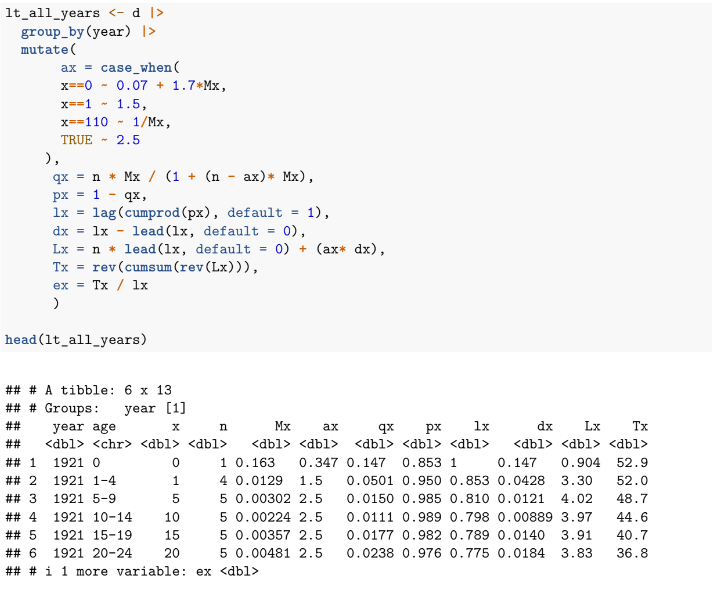
Now we can use tidyverse to calculate the columns in the life table, based on the equations presented in previous sections. I set the radix l0 to be zero and filter to just include the year 2015. This code makes use of the case\_when function, which allows to define different values of nax based on age group. Formulas for other columns are based on equations stated above. The formula for Tx is implemented by first reversing the nLx column, taking the cumulative sum, and then reversing the result.



We can extend this to calculate life tables for every year using the group\_by function.



We can extend this to calculate life tables for every year using the group\_by function



Let’s plot the lx curve over time; notice as mortality improves, the drop in lx after the first year of life becomes less noticeable, and the curve becomes more ‘rectangular’2



