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** 1

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:

**library**(tidyverse) *# data manipulation and ggplot functions*

**library**(kableExtra) *# format tables*

**library**(janitor) *# to easily clean up column names*

**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 *l*0 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 *l*0 = 1 then then *lx* is a probability of survival to age *x*. Note that for now we are implicitly assuming this *l*0 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.

lt <- **read\_table**("http://www.prdh.umontreal.ca/BDLC/data/ont/fltper\_5x5.txt", skip = 2) lt <- lt **|>**

**filter**(Year**==**"2015-2019") **|>**

**mutate**(x = **c**(0,1,**seq**(5, 110, by = 5)),

n = **lead**(x, default = Inf)**-**x)

lt **|>**

**select**(x,n, lx) **|>**

**kable**()

2

x lx

| n |
| --- |
| 1 |
| 4 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| 5 |
| Inf |

0 100000

1 99569

5 99515

10 99478

15 99430

20 99334

25 99192

30 99008

35 98759

40 98461

45 98036

50 97380

55 96304

60 94626

65 92094

70 88287

75 82345

80 73473

85 59620

90 40198

95 18841

100 5017

105 599

110 28

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*.

lt **|>**

**mutate**(lx = lx**/**100000) **|>**

**ggplot**(**aes**(x, lx)) **+**

**geom\_line**() **+**

**xlab**("age") **+**

**theme\_bw**(base\_size = 14) **+**

**ggtitle**("Survivorship for Ontario, 2015")

3

x

l

1.00 0.75 0.50 0.25 0.00

Survivorship for Ontario, 2015

0 30 60 90 age

**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 certain age *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:

lt **|>**

**select**(x,n, lx, dx) **|>**

**kable**()

4

x dx

| n | lx |
| --- | --- |
| 1 | 100000 |
| 4 | 99569 |
| 5 | 99515 |
| 5 | 99478 |
| 5 | 99430 |
| 5 | 99334 |
| 5 | 99192 |
| 5 | 99008 |
| 5 | 98759 |
| 5 | 98461 |
| 5 | 98036 |
| 5 | 97380 |
| 5 | 96304 |
| 5 | 94626 |
| 5 | 92094 |
| 5 | 88287 |
| 5 | 82345 |
| 5 | 73473 |
| 5 | 59620 |
| 5 | 40198 |
| 5 | 18841 |
| 5 | 5017 |
| 5 | 599 |
| Inf | 28 |

0 431

1 54

5 37

10 48

15 96

20 142

25 184

30 248

35 298

40 426

45 655

50 1077

55 1677

60 2533

65 3807

70 5942

75 8873

80 13853

85 19422

90 21356

95 13824

100 4419

105 571

110 28

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

*lx*=*lx − lx* + *lx*+*n*

*npx* = 1 *− nqx* = 1 *−ndx*

*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*memento mori*.

5

lt **|>**

**mutate**(px = 1**-** qx) **|>**

**select**(x, n, lx, dx, qx, px) **|>**

**kable**()

x px

| n | lx | dx | qx |
| --- | --- | --- | --- |
| 1 | 100000 | 431 | 0.00431 |
| 4 | 99569 | 54 | 0.00054 |
| 5 | 99515 | 37 | 0.00037 |
| 5 | 99478 | 48 | 0.00049 |
| 5 | 99430 | 96 | 0.00096 |
| 5 | 99334 | 142 | 0.00143 |
| 5 | 99192 | 184 | 0.00186 |
| 5 | 99008 | 248 | 0.00251 |
| 5 | 98759 | 298 | 0.00302 |
| 5 | 98461 | 426 | 0.00432 |
| 5 | 98036 | 655 | 0.00668 |
| 5 | 97380 | 1077 | 0.01106 |
| 5 | 96304 | 1677 | 0.01742 |
| 5 | 94626 | 2533 | 0.02676 |
| 5 | 92094 | 3807 | 0.04134 |
| 5 | 88287 | 5942 | 0.06730 |
| 5 | 82345 | 8873 | 0.10775 |
| 5 | 73473 | 13853 | 0.18854 |
| 5 | 59620 | 19422 | 0.32576 |
| 5 | 40198 | 21356 | 0.53129 |
| 5 | 18841 | 13824 | 0.73370 |
| 5 | 5017 | 4419 | 0.88066 |
| 5 | 599 | 571 | 0.95283 |
| Inf | 28 | 28 | 1.00000 |

0 0.99569

1 0.99946

5 0.99963

10 0.99951

15 0.99904

20 0.99857

25 0.99814

30 0.99749

35 0.99698

40 0.99568

45 0.99332

50 0.98894

55 0.98258

60 0.97324

65 0.95866

70 0.93270

75 0.89225

80 0.81146

85 0.67424

90 0.46871

95 0.26630

100 0.11934

105 0.04717

110 0.00000

**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 1*a*0 = 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.*

6

Note for the last interval there are no survivors, so *∞Lx* = *∞ax · ∞dx*.

Adding *nax* and *nLx* to the Ontario life table:

lt **|>**

**select**(x,n, lx, dx, ax, Lx) **|>**

**kable**()

x Lx

| n | lx | dx | ax |
| --- | --- | --- | --- |
| 1 | 100000 | 431 | 0.14 |
| 4 | 99569 | 54 | 1.62 |
| 5 | 99515 | 37 | 2.34 |
| 5 | 99478 | 48 | 2.78 |
| 5 | 99430 | 96 | 2.90 |
| 5 | 99334 | 142 | 2.60 |
| 5 | 99192 | 184 | 2.53 |
| 5 | 99008 | 248 | 2.57 |
| 5 | 98759 | 298 | 2.63 |
| 5 | 98461 | 426 | 2.69 |
| 5 | 98036 | 655 | 2.69 |
| 5 | 97380 | 1077 | 2.72 |
| 5 | 96304 | 1677 | 2.68 |
| 5 | 94626 | 2533 | 2.68 |
| 5 | 92094 | 3807 | 2.67 |
| 5 | 88287 | 5942 | 2.67 |
| 5 | 82345 | 8873 | 2.67 |
| 5 | 73473 | 13853 | 2.67 |
| 5 | 59620 | 19422 | 2.62 |
| 5 | 40198 | 21356 | 2.44 |
| 5 | 18841 | 13824 | 2.19 |
| 5 | 5017 | 4419 | 1.87 |
| 5 | 599 | 571 | 1.55 |
| Inf | 28 | 28 | 1.45 |

0 99629

1 398147

5 497476

10 497283

15 496947

20 496329

25 495505

30 494435

35 493089

40 491321

45 488665

50 484445

55 477626

60 467245

65 451612

70 427608

75 391032

80 335135

85 251897

90 146361

95 55428

100 11261

105 1025

110 41

**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 *l*0 and interval length *n*, the maximum *nLx* could be is *l*0 *·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 15*L*15. 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.

lt **|>**

**mutate**(lx = lx**/**100000) **|>**

**ggplot**(**aes**(x, lx)) **+**

**geom\_line**() **+**

**xlab**("age") **+**

**theme\_bw**(base\_size = 14) **+**

**geom\_vline**(xintercept = 15, lty = 2, color = "red")**+**

**geom\_vline**(xintercept = 30, lty = 2, color = "red")**+**

**ggtitle**("Survivorship for Ontario, 2015")

7

x

l

1.00 0.75 0.50 0.25 0.00

Survivorship for Ontario, 2015

0 30 60 90 age

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

*nLx* =1*l*0Z *x*+*n x*

*lxdx*

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*:

*Tx* =X*∞ x*

*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*

*Tx* =1*l*0Z *∞ x*

*lxdx*

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’

8

terminology is related to the expected value in the statistical sense. *ex* is calculated as

*ex* =*Tx*

*lx*

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, *e*0. 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:

lt **|>**

**select**(x,n, lx, dx, ax, Lx, Tx, ex) **|>**

**kable**()

x ex

| n | lx | dx | ax | Lx | Tx |
| --- | --- | --- | --- | --- | --- |
| 1 | 100000 | 431 | 0.14 | 99629 | 8449540 |
| 4 | 99569 | 54 | 1.62 | 398147 | 8349911 |
| 5 | 99515 | 37 | 2.34 | 497476 | 7951764 |
| 5 | 99478 | 48 | 2.78 | 497283 | 7454288 |
| 5 | 99430 | 96 | 2.90 | 496947 | 6957006 |
| 5 | 99334 | 142 | 2.60 | 496329 | 6460059 |
| 5 | 99192 | 184 | 2.53 | 495505 | 5963730 |
| 5 | 99008 | 248 | 2.57 | 494435 | 5468225 |
| 5 | 98759 | 298 | 2.63 | 493089 | 4973790 |
| 5 | 98461 | 426 | 2.69 | 491321 | 4480701 |
| 5 | 98036 | 655 | 2.69 | 488665 | 3989380 |
| 5 | 97380 | 1077 | 2.72 | 484445 | 3500715 |
| 5 | 96304 | 1677 | 2.68 | 477626 | 3016271 |
| 5 | 94626 | 2533 | 2.68 | 467245 | 2538645 |
| 5 | 92094 | 3807 | 2.67 | 451612 | 2071400 |
| 5 | 88287 | 5942 | 2.67 | 427608 | 1619788 |
| 5 | 82345 | 8873 | 2.67 | 391032 | 1192180 |
| 5 | 73473 | 13853 | 2.67 | 335135 | 801148 |
| 5 | 59620 | 19422 | 2.62 | 251897 | 466013 |
| 5 | 40198 | 21356 | 2.44 | 146361 | 214115 |
| 5 | 18841 | 13824 | 2.19 | 55428 | 67755 |
| 5 | 5017 | 4419 | 1.87 | 11261 | 12326 |
| 5 | 599 | 571 | 1.55 | 1025 | 1066 |
| Inf | 28 | 28 | 1.45 | 41 | 41 |

0 84.50

1 83.86

5 79.91

10 74.93

15 69.97

20 65.03

25 60.12

30 55.23

35 50.36

40 45.51

45 40.69

50 35.95

55 31.32

60 26.83

65 22.49

70 18.35

75 14.48

80 10.90

85 7.82

90 5.33

95 3.60

100 2.46

105 1.78

110 1.45

**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.

9

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 proba bilities of death, *nqx*. The mortality rate is the number deaths divided by person years lived, so in life table

notation this is:

*nMx* =*ndx nLx*

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:

*nqx* =*n · nMx*

1 + (*n − nax*) *· nMx*

How did this come about? By rewriting

*nLx* = *n · lx*+*n* + *nax · ndx* = *n*(*ln − ndx*) + *nax · ndx*

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 1*a*0 and 4*a*1. 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):

1*a*0 = 0*.*07 + 1*.*71*M*0

and

4*a*1 = 1*.*5

10

The equation for 1*a*0 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 at the 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 live 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:

Mx <- **read\_table**("http://www.prdh.umontreal.ca/BDLC/data/que/Mx\_5x5.txt", skip = 2, col\_types = 'ccddd'

d <- Mx **|>**

**mutate**(year = **as.numeric**(**substr**(Year, 1, 4))) **|>**

**select**(year, Age, Total) **|>**

**clean\_names**() **|>**

**rename**(Mx = total)

**head**(d)

## # A tibble: 6 x 3

## year age Mx

## <dbl> <chr> <dbl>

## 1 1921 0 0.163

## 2 1921 1-4 0.0129

## 3 1921 5-9 0.00302

## 4 1921 10-14 0.00224

## 5 1921 15-19 0.00357

## 6 1921 20-24 0.00481

The age column is a character, so let’s make and age x and interval length n column:

11

d <- d **|>**

**mutate**(x = **as.numeric**(**str\_remove**(age, "-.\*|**\\**+")),

n = **lead**(x, default = Inf) **-** x) **|>**

**filter**(x**<**105) **|>** *# remove older ages that have varying data availability* **select**(year, age, x, n, Mx)

**head**(d)

## # A tibble: 6 x 5

## year age x n Mx

## <dbl> <chr> <dbl> <dbl> <dbl>

## 1 1921 0 0 1 0.163

## 2 1921 1-4 1 4 0.0129

## 3 1921 5-9 5 5 0.00302

## 4 1921 10-14 10 5 0.00224

## 5 1921 15-19 15 5 0.00357

## 6 1921 20-24 20 5 0.00481

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 *l*0 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.

lt\_2015 <- d **|>**

**filter**(year**==**2015) **|>**

**mutate**(

ax = **case\_when**(

x**==**0 **~** 0.07 **+** 1.7**\***Mx,

x**==**1 **~** 1.5,

x**==**110 **~** 1**/**Mx,

TRUE **~** 2.5

),

qx = n **\*** Mx **/** (1 **+** (n **-** ax)**\*** Mx),

px = 1 **-** qx,

lx = **lag**(**cumprod**(px), default = 1),

dx = lx **- lead**(lx, default = 0),

Lx = n **\* lead**(lx, default = 0) **+** (ax**\*** dx),

Tx = **rev**(**cumsum**(**rev**(Lx))),

ex = Tx **/** lx

)

**head**(lt\_2015)

## # A tibble: 6 x 13

## year age x n Mx ax qx px lx dx Lx Tx ## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> ## 1 2015 0 0 1 4.21e-3 0.0772 4.19e-3 0.996 1 4.19e-3 0.996 82.7 ## 2 2015 1-4 1 4 1.35e-4 1.5 5.40e-4 0.999 0.996 5.38e-4 3.98 81.7 ## 3 2015 5-9 5 5 6.70e-5 2.5 3.35e-4 1.00 0.995 3.33e-4 4.98 77.7 ## 4 2015 10-14 10 5 8.40e-5 2.5 4.20e-4 1.00 0.995 4.18e-4 4.97 72.8 ## 5 2015 15-19 15 5 2.21e-4 2.5 1.10e-3 0.999 0.995 1.10e-3 4.97 67.8 ## 6 2015 20-24 20 5 3.42e-4 2.5 1.71e-3 0.998 0.993 1.70e-3 4.96 62.8 ## # i 1 more variable: ex <dbl>

12

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

lt\_all\_years <- d **|>**

**group\_by**(year) **|>**

**mutate**(

ax = **case\_when**(

x**==**0 **~** 0.07 **+** 1.7**\***Mx,

x**==**1 **~** 1.5,

x**==**110 **~** 1**/**Mx,

TRUE **~** 2.5

),

qx = n **\*** Mx **/** (1 **+** (n **-** ax)**\*** Mx),

px = 1 **-** qx,

lx = **lag**(**cumprod**(px), default = 1),

dx = lx **- lead**(lx, default = 0),

Lx = n **\* lead**(lx, default = 0) **+** (ax**\*** dx),

Tx = **rev**(**cumsum**(**rev**(Lx))),

ex = Tx **/** lx

)

**head**(lt\_all\_years)

## # A tibble: 6 x 13

## # Groups: year [1]

## year age x n Mx ax qx px lx dx Lx Tx ## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> ## 1 1921 0 0 1 0.163 0.347 0.147 0.853 1 0.147 0.904 52.9 ## 2 1921 1-4 1 4 0.0129 1.5 0.0501 0.950 0.853 0.0428 3.30 52.0 ## 3 1921 5-9 5 5 0.00302 2.5 0.0150 0.985 0.810 0.0121 4.02 48.7 ## 4 1921 10-14 10 5 0.00224 2.5 0.0111 0.989 0.798 0.00889 3.97 44.6 ## 5 1921 15-19 15 5 0.00357 2.5 0.0177 0.982 0.789 0.0140 3.91 40.7 ## 6 1921 20-24 20 5 0.00481 2.5 0.0238 0.976 0.775 0.0184 3.83 36.8 ## # i 1 more variable: ex <dbl>

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

lt\_all\_years **|>**

**ggplot**(**aes**(x, lx, color = year, group = year)) **+**

**geom\_line**()**+**

**ggtitle**("Survivorship over time, Quebec")

2Indeed, this phenomenon is called ‘the rectangularization of the survival curve’ and is related to the study of ‘compression’ of mortality, meaning that deaths become more concentration at older ages. See for example Wilmoth and Horiuchi.

13

x

l

Survivorship over time, Quebec

1.00

0.75

0.50

0.25

0.00

0 25 50 75 100 x

year

2000

1980

1960

1940

We can also plot life expectancy at birth over time:

lt\_all\_years **|>**

**filter**(x**==**0) **|>**

**ggplot**(**aes**(year, ex)) **+**

**geom\_line**() **+**

**ggtitle**("Life expectancy at birth, Quebec") 14

x e

Life expectancy at birth, Quebec

80

70

60

1920 1940 1960 1980 2000 year

15