

Life Tables: 1948-2006

2024-09-02

- M1 MIDS MA7BY020
- [Université Paris Cité](#)
- Année 2024-2025
- [Course Homepage](#)
- [Moodle](#)



! Objectives

Loading

```
datafile <- 'tamed_life_table.Rds'
fpath <- str_c("../DATA/", datafile) # here::here('DATA', datafile) # check getwd() if probl

if (! file.exists(fpath)) {
  download.file("https://stephane-v-boucheron.fr/data/tamed_life_table.Rds",
               fpath,
               mode="wb")
}

life_table <- readr::read_rds(fpath)
```

Definitions

Check on <http://www.mortality.org> the meaning of the different columns.

See: *Demography: Measuring and Modeling Population Processes* by SH Preston, P Heuveline, and M Guillot. Blackwell. Oxford. 2001.

Document [Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle](#) contains detailed information on the construction of Life Tables for France.

Period tables versus cohort tables

Two kinds of Life Tables can be distinguished: *Period tables* (*Table du moment*) which contain for each period (here a period is a calendar year), the mortality risks at different age ranges (here, we have one year ranges) for that very period; and *Tables de génération* which contain for a given birthyear, the mortality risks at which an individual born during that year has been exposed.

The life tables investigated in this lab are *Table du moment*. According to the document by Vallin and Meslé, building the life tables required decisions and doctoring.

Lexis diagrams

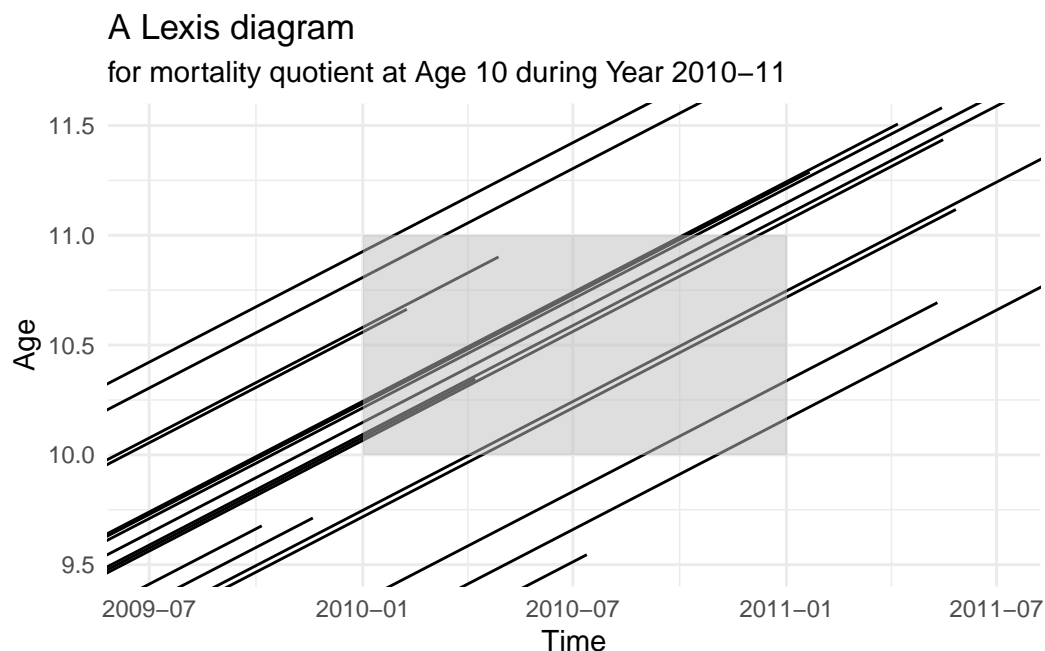
Lexis diagrams provide a graphical device that summarizes the construction of mortality quotients (and other rates in demography).

```
birth_dates <- as_date("1999-01-01") + duration(sample(2*365, size=20, replace=T),units="day")
death_dates <- as_date("2009-07-01") + duration(sample(3*365, size=20, replace=T),units="day")

b_period <- as_date("2010-01-01")
b_frame <- as_date(b_period - duration(1, units = "year"))
b_age <- 10L

tb_ld <- tibble(birth=birth_dates, death=death_dates)

tb_ld |>
  ggplot() +
  geom_segment(aes(x=b_frame,
                  xend=death,
                  y=interval(birth, b_frame)/years(1),
                  yend=interval(birth, death)/years(1))
  ) +
  annotate(geom="rect",
    xmin=b_period,
    xmax=as_date(b_period + duration(1, units = "year")),
    ymin=b_age,
    ymax=b_age + 1L,
    fill="grey",
    alpha=.5) +
  ylab("Age") +
  xlab("Time") +
  coord_cartesian(xlim=c(as_date(b_period - duration(6, units = "months")),
                        as_date(b_period + duration(18, units = "months"))),
    ylim=c(b_age - .5, b_age+1.5)) +
  labs(
    title="A Lexis diagram",
    subtitle = "for mortality quotient at Age 10 during Year 2010-11"
  )
```



Each line represents the *life line* of an individual born during years 1999 and 2000 and deceased between mid 2009 and mid 2012. In order to compute the mortality quotient at age 10 for year 2010, we have to compute the relevant *number of occurrences*, that is the number of segments ending in the grey rectangle, and the *sum of exposure times*, which is proportional to the sum of the lengths of the segments crossing the grey rectangle.

Have a look at [Lexis diagram](#) or at Preston *et al.*

Definitions can be obtained from www.lifeexpectancy.org. We translate it into mathematical (rather than demographic) language.

The mortality quotients define a probability distribution over \mathbb{N} . This probability distribution is a *construction* that reflects the health situation in a population at a given time. This probability distribution does not describe the sequence of sanitary situations experienced by a *cohort* (people born during a specific year).

One works with a period, or current, life table (*table du moment*). This summarizes the mortality experience of persons across all ages in a short period, typically one year or three years. More precisely, the death probabilities q_x for every age x are computed for that short period, often using census information gathered at regular intervals. These q_x 's are then applied to a hypothetical cohort of 100000 people over their life span to produce a life table.

```
small_tb <- life_table |>
  filter(Country=='France', Year== 2010, Gender=='Female', Age < 10 | between(Age, 80, 89)) |>
  select(Age, qx, mx, lx, dx, Lx, Tx, ex)

small_tb
```

```
# A tibble: 20 x 8
```

```
  Age    qx    mx    lx    dx    Lx    Tx    ex
```

	<int>	<dbl>	<dbl>	<int>	<int>	<int>	<int>	<dbl>
1	0	0.00324	0.00325	100000	324	99722	8465207	84.6
2	1	0.00032	0.00032	99676	32	99660	8365484	83.9
3	2	0.00015	0.00015	99645	15	99637	8265824	83.0
4	3	0.00011	0.00011	99630	11	99624	8166187	82.0
5	4	0.00008	0.00008	99619	8	99615	8066563	81.0
6	5	0.00005	0.00005	99611	5	99608	7966948	80.0
7	6	0.00008	0.00008	99606	8	99602	7867339	79.0
8	7	0.00008	0.00008	99598	8	99594	7767737	78.0
9	8	0.00008	0.00008	99590	8	99586	7668143	77
10	9	0.00007	0.00007	99582	7	99578	7568557	76
11	80	0.0298	0.0302	75619	2252	74493	802295	10.6
12	81	0.0346	0.0352	73367	2535	72099	727802	9.92
13	82	0.0398	0.0406	70832	2818	69423	655702	9.26
14	83	0.0464	0.0475	68014	3158	66435	586280	8.62
15	84	0.0539	0.0554	64856	3493	63109	519845	8.02
16	85	0.0610	0.0630	61362	3745	59490	456736	7.44
17	86	0.0699	0.0725	57617	4029	55603	397246	6.89
18	87	0.0793	0.0826	53588	4249	51464	341643	6.38
19	88	0.0922	0.0966	49339	4547	47066	290180	5.88
20	89	0.105	0.111	44792	4706	42440	243114	5.43

i Question

The table above is not as readable as it should. Use package `gt` to get a more tunable output.

Reorder and filter the columns so that **Age** comes first (they identify rows), then **qx**, **mx** up to **ex**. You can use `select` or `relocate`, or both to do this. Note that **Gender**, **Country**, **Year** are constant in this tibble and need to be reported in the table header, but nowhere else.

Columns **qx** and **mx** (for mortality quotient and central death rate) should be displayed in scientific notation so that the fact that their range extends over several orders of magnitude shows up.

Columns **lx**, **dx**, **Lx**, **Tx** contain integer values.

Column **ex** (residual life expectancy) is a (fictional) decimal number of years

Understanding the columns of the life table

In the sequel, we denote by F_t the *cumulative distribution function* for year t . We agree on $\overline{F}_t = 1 - F_t$ and $F_t(-1) = 0$. Henceforth, \overline{F} is called the *survival* function.

qx (age-specific) risk of death at age x , or mortality quotient at given age x for given year t .

i About the definition of $q_{\{t,x\}}$

Defining and computing $q_{\{t,x\}}$ does not boil down to knowing the number of people at age x at the beginning of year t and knowing how many of them died during year t . If we want to be rigorous, we need to know all life lines in the Lexis diagram, or equivalently, how many people at Age x were alive on each day of Year t .

🔥 Mortality quotients define a probability distribution

For a given year t , the sequence of mortality quotients define a survival function \bar{F}_t using the following recursion:

$$q_{t,x} = \frac{\bar{F}_t(x) - \bar{F}_t(x+1)}{\bar{F}_t(x)}$$

with boundary condition $\bar{F}_t(-1) = 1$.

This recursion can also be read as:

$$\bar{F}_t(x+1) = \bar{F}_t(x) \times (1 - q_{t,x+1}).$$

This artificial probability distribution is used to define and compute life expectancies.



$q_{t,x}$ is the *hazard rate* of \bar{F}_t at age x .

mx *central death rate* at age x during year t . This is connected with $q_{t,x}$ by

$$m_{t,x} = -\log(1 - q_{t,x}),$$

or equivalently

$$q_{t,x} = 1 - \exp(-m_{t,x})$$

i About central death rate

If we want to define a continuous probability distribution G over $[0, \infty)$ so that G and F coincide over integers and G has piecewise constant hazard rate, we can pick $m_{t,x}$ as the piecewise constant hazard rate.

1x the so-called *survival function*: the scaled proportion of persons alive at age x . These values are computed recursively from the $q_{t,x}$ values using the formula

$$l_t(x+1) = l_t(x) \times (1 - q_{t,x}),$$

with $l_{t,0}$, the *radix* of the table, (arbitrarily) set to 100000. In the table **1x** is rounded to the next integer

Function $l_{t,\cdot}$ and \bar{F}_t are connected by

$$l_{t,x+1} = l_{t,0} \times \overline{F}_t(x).$$

dx $d_{t,x} = q_{t,x} \times l_{t,x}$. The fictitious number of deaths occurring at age x during year t . Again this is a rounded quantity.

Tx Total number of person-years lived by the cohort from age x to $x+1$. This is the sum of the years lived by the $l_{t,x+1}$ persons who survive the interval, and the $d_{t,x}$ persons who die during the interval. The former contribute exactly 1 year each, while the latter contribute, on average, approximately half a year, so that $L_{t,x} = l_{t,x+1} + 0.5 \times d_{t,x}$. This approximation assumes that deaths occur, on average, half way in the age interval x to $x+1$. Such is satisfactory except at age 0 and the oldest age, where other approximations are often used.

Compare with the denominator in the definition of **qx** and its description using the Lexis diagram.

We will stick to a simplified vision $L_{t,x} = l_{t,x+1}$

ex: Residual Life Expectancy at age x and year t

This is the expectation of $X - x$ for a random variable X distributed according to \overline{F}_t conditionnally on the event $\{X \geq x\}$. That is $e_{t,x}$ is the expectation of the probability distribution defined by $\overline{F}_t(\cdot + x - 1) / \overline{F}_t(x - 1)$.

Question

Check dependencies between columns


Western countries in 1948

Several pictures share a common canvas:

Question

Plot mortality quotients (**qx**) against age using a logarithmic scale on the y axis. Countries are identified by aesthetics (**shape**, **color**, **linetype**).

1. Use facetting to plot **qx** of all countries at all ages for years 1950, 1960, ..., 2010.
2. Use **plotly** to build an animated plot using **Year** for the **frame** aesthetics.

 Abiding to the DRY principle, define a prototype **ggplot** (alternatively **plotly**) object. The prototype will then be fed with different datasets and decorated and arranged for the different figures.

Question

Plot ratios between mortality quotients (**qx**) in European countries and mortality quotients in the USA in 1948.

i Comment

This animation reveals less than the preceding one since we just have ratios with respect to the USA. But the patterns followed by European societies emerge in a more transparent way. The divide between northern and southern Europe at the onset of the period is even more visible. The ratios are important across the continent: there is a factor of 10 between spanish and swedish infant mortality rates. But the ratios at ages 50 and above tend to be similar. By the early 60s, the gap between southern and northern Europe has shrunk. By now, the ratios between mortality quotients tend to be within a factor of 2 across all ages, and even less at ages 50 and above.

Death rates evolution since WW II

Question

Plot mortality quotients (column `qx`) for both genders as a function of **Age** for years 1946, 1956, ... up to 2016 . Use aesthetics to distinguish years. You will need to categorize the **Year** column (`forcats::` may be helpful).

1. Facet by **Gender** and **Country**
2. Pay attention to axes labels, to legends. Assess logarithmic scales.

i Question

Write a function `ratio_mortality_rates` with signature `function(df, reference_year=1946, target_years=seq(1946, 2016, 10))` that takes as input:

- a dataframe with the same schema as `life_table`,
- a reference year `ref_year` and
- a sequence of years `target_years`

and that returns a dataframe with schema:

Column Name	Column Type
Year	integer
Age	integer
mx	double
mx.ref_year	double
Country	factor
Gender	factor

where `(Country, Year, Age, Gender)` serves as a *primary key*, `mx` denotes the central death rate at **Age** for **Year** and **Gender** in **Country** whereas `mx_ref_year` denotes central death rate at **Age** for argument `reference_year` in **Country** for **Gender**.

i Question

Draw plots displaying the ratio $q_{x,t}/q_{x,1946}$ for ages $x \in 1, \dots, 90$ and year t for $t \in 1946, \dots, 2013$ where $q_{x,t}$ is the mortality quotient at age x during year t .

1. Handle both genders and all countries
2. One properly faceted plot is enough.

i Comment.

During the last seventy years, death rates decreased at all ages in all seven countries. This progress has not been uniform across ages, genders and countries. Across most countries, infant mortality dramatically improved during the first post-war decade while death rates at age 50 and above remained stable until the mid seventies.

Trends

We noticed that mortality quotients did not evolve in the same way across all ages: first, the decay has been much more significant at low ages; second, the decay of mortality quotients at old ages (above 60) mostly took place during the last four decades. It is worth digging separately at what happened for different parts of life.

Question

Plot mortality quotients at ages 0, 1, 5 as a function of time. Facet by Gender and Country

i Comment

All European countries achieved the same infant mortality rates after year 2000. The USA now lag behind.

During years 1940-1945, in the Netherlands and France, gains obtained before 1940 were reversed. Year 1945 was particularly difficult in the Netherlands.

i Question

Plot mortality quotients at ages 15, 20, 40, 60 as a function of time. Facet by **Gender** and **Country**

i Comment.

While death rates at ages 15 and 20 among women are close across all societies, death rates are higher at age 20 than at age 15 among men. In France, at age 20, death rates declined from 1945 until 1960, and then increased back to their initial level until 1980. Male death rates at age 60 started to decline around 1980. Female death rates at age 60 declined steadily throughout the 7 decades. Years 1940-1945 exhibit disruptions with different shapes and intensities in Italy, France, England & Wales, and the Netherlands.