Life expectancy: a global health index

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 problem

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)
```

> **References**
>
> For definitions of column, check on http://www.mortality.org the meaning of the different columns.
> See also *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.

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.

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 ear 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 $\overline{F}_t$ using the following recursion:

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

with boundary condition $\overline{F}_t(-1) = 1$.
This recursion can also be read as:

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

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

☛

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

`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 $\overline{F}_t(\cdot + x - 1)/\overline{F}_t(x-1)$.

Rearrangement

Question

From dataframe `life_table`, compute another dataframe called `life_table_pivot` with primary key `Country`, `Gender` and `Year`, with a column for each `Age` from `0` up to `110`. For each age column, the entry should be the central death rate at the age defined by column, for `Country`, `Gender` and `Year` identifying the row.
You may use functions `pivot_wider`, `pivot_longer` from `tidyr::` package.
The resulting schema should look like:

Column Name	Type
Country	factor
Gender	factor
Year	integer
`0`	double
`1`	double
`2`	double
`3`	double
⋮	⋮

Question

Using `life_table_pivot` compute life expectancy at birth for each Country, Gender and Year using formula

$$e_{t,0} = \sum_{x=0}^{\infty} \overline{F}_t(x)$$

Life expectancy and window functions

Question

Write a function that takes as input a vector of mortality quotients, as well as an age, and returns the residual life expectancy corresponding to the vector and the given age.

Question

Write a function that takes as input a dataframe with the same schema as `life_table` and returns a data frame with columns `Country`, `Gender`, `Year`, `Age` defining a primary key and a column `res_lex` containing *residual life expectancy* corresponding to the pimary key.

In order to compute residual life expectancies, you may consider using `window` functions over apropriately defined windows. The next window function suffices to compute life expectancy at birth. It computes the logarithm of survival probabilities for each `Country`, `Year`, `Gender` (partition) at each `Age`. Note that the expression mentions an aggregation function `sum` and that the correction of the result is ensured by a correct design of the `frame` argument.

> **Question**
>
> Compute residal life expectancies at all ages using window functions
> You can use `slider::slide()`.

Computing residual life expectancies using window functions and `accumulate`

> The official calculation of residual life expectancies assumes that except at age 0 and great age, people die uniformly at random between age x and $x+1$:
>
> $$e_{t,x} = (1 - q_{t,x}) \times (1 + e_{t,x+1}) + \frac{1}{2} \times q_{t,x}$$
>
> This recursion suggests a more efficient to compute *residual life expectancies* at all ages. Indeed, `purrr::accumulate()` allows to compute all values for $e_{t,x}$ using exactly one pass over the table.
> See https://purrr.tidyverse.org/reference/accumulate.html

> **Question**

> **Question**

> **Question**
>
> Compute and display residual life expectancies for ages 0 to 9 for year 1972

> **Question**
>
> Plot residual life expectancy as a function of `Year` at ages 60 and 65, facet by `Gender` and `Country`.

> **Question**