Methodology for Calculating the Prospective Age Dataset

mz

16 August, 2018

Contents

Intro

This document describes the methodology used to calculate the variables in the Prospective Age Dataset. It describes the original data used, and the calculations preformed for both variables. See also the codebook for a description of the dataset itself. This methods file, the codebook and the dataset are deposited on figshare and updated automatically.

Data

Both datasets are part of the UN 2017 Revision of the World Population Prospects UN (2017) and were downloaded from the UN Population Division website on 13.08.2018.

The original variables are described here for both datasets:

UN World Population Prospects - Standard Projections

- **LocID** (numeric): numeric code for the location; for countries and areas, it follows the ISO 3166-1 numeric standard
- **Location** (string): name of the region, subregion, country or area
- **VarID** (numeric): numeric code for the variant
- **Variant** (string): projection variant name (Medium is the most used)
- **Time** (string): label identifying the single year (e.g. 1950) or the period of the data (e.g. 1950-1955)
- **MidPeriod** (numeric): numeric value identifying the mid period of the data, with the decimal representing the month (e.g. 1950.5 for July 1950)
- **AgeGrp** (string): label identifying the single age (e.g. 15) or age group (e.g. 15-19)

- **PopFemale**: Female population for the individual age (thousands)
- **PopTotal**: Total population for the individual age (thousands)
- **PopMale**: Male population for the individual age (thousands)

UN LifeTables - Mortality Indicators. [*ibid*]

Abridged life tables up to age 85 by sex and both sexes combined providing a set 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 period, from 1950-1955 to 2095-2100.

- **mx**: Central death rate, nmx, for the age interval (x, x+n)
- **qx**: Probability of dying (nqx), for an individual between age x and x+n
- **px**: Probability of surviving, (npx), for an individual of age x to age x+n
- **lx**: Number of survivors, (lx), at age (x) for 100000 births
- **dx**: Number of deaths, (ndx), between ages x and x+n
- **Lx**: Number of person-years lived, (nLx), between ages x and x+n
- **Sx**: Survival ratio (nSx) corresponding to proportion of the life table population in age group (x, x+n) who are alive n year later
- **Tx**: Person-years lived, (Tx), above age x
- **ex**: Expectation of life (ex) at age x, i.e., average number of years lived subsequent to age x by those reaching age x
- **ax**: Average number of years lived (nax) between ages x and x+n by those dying in the interval

---

Methodology

See Sanderson and Scherbov (2008) for more info on prospective measures of ageing.

Calculating the old-age thresholds

The *old-age threshold*, is the age at which the remaining life expectancy is 15 years. Calculating it was based on the abridged life tables which has life expectancy (e_x) values for five year age groups. I used splines to interpolate the age x where e_x equals 15.

I use the `R stats` function `splinefun()` and the monotone Hermite spline computation according to the method of Fritsch and Carlson: `method = "monoH.FC"` which produces identical results (to the second decimal point) as the ones published in the IIASA Ageing Demographic Data Sheet (Scherbov, Andruchowitz, and Sanderson 2018), for a random selection of a dozen country/year combinations. Although they do not provide details on their methodology, it makes most sense to use this method as it guarantees the interpolated values remain monotonically increasing/decreasing iff the input data is as well, which is what we would expect from life expectancy data short of any major disturbance.

Additionally to being abridged the life tables are for five-year time periods as well. So there are two steps in the interpolation:

1. finding the age x where e_x is 15 for every 5-year period
2. interpolating these for every individual year.

So we start with life expectancy values:

$e_x^{y=i_5}$

where the e is given for five year age groups (except for first two, and last):

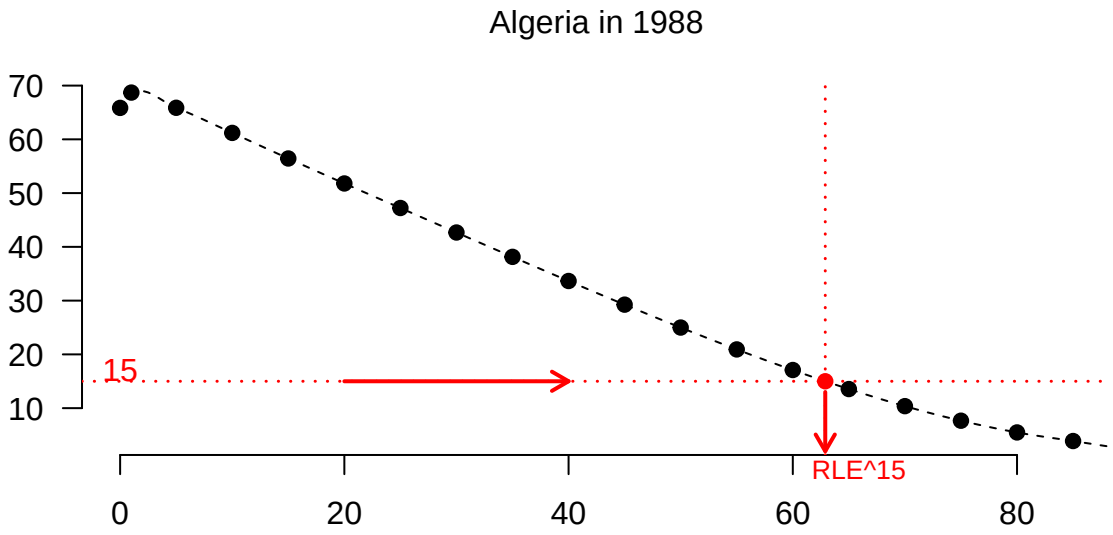

Figure 1: Interpolation of old-age threhsold (RLE^15): x where ex is 15.

$x = 0, 1, 5, 10, ...80, 85$

and for time periods of five years

$i_5 = 1950 - 55, 1955 - 60...2095 - 2100$

Now use splines to get the coefficients for life expectancy as a function of age, with which we can interpolate the age x at which $e_x = 15$ for each time period $y = i_5$. See also Figure 1 where black points represent known data and red the interpolated.

Now technically, this chart in Figure 1 is a misleading because I am not actually using the spline $g(e_x)$. I am doing the inverse and getting the spline function for $g(x)$, so that I can then enter $e_x = 15$ and get out the value of x i.e. the age. In doing this I am assuming that the function is monotonic–or rather that it is monotonic in the area that I'm interested in. Which is fine, because the only point at which it isn't monotonic is at the highest e_x values (i.e. at birth). But the value $e_x = 15$ occurs only once, so this is OK.

So let's call the old-age threshold at time $y = i$, following Sanderson and Scherbov (2008), $RLE^{15}_{y=i}$, and in this case we only have it for five year time periods so $RLE^{15}_{y=i_5}$. RLE stands for remaining life expectancy.

Then, for a smoother graph, and because we will need them in the next step, we then use splines to interpolate to single years instead of the five year periods. Here we use MidPeriod variable as the correct for each old-age threshold. So the RLE^{15} for 1950-1955 is actually the value in 1953, and the remaining years are then interpolated. Because of this I the first three and last two years might be treated with caution, since they are beyond the end points. The MidPeriod thresholds are therefore the input data for the interpolation of the individual year thresholds.

So we have $RLE^{15}_{y=i_5}$ as a function of y and interpolate using $g(RLE)$ to get individual years:

$RLE^{15}_{y=i}$, where $i = 1953, 1954...2098$

See also Figure 2 where again black points represent known data and red the interpolated.

Calculating the proportion over old-age threshold

So now we have the old-age threshold for every year.

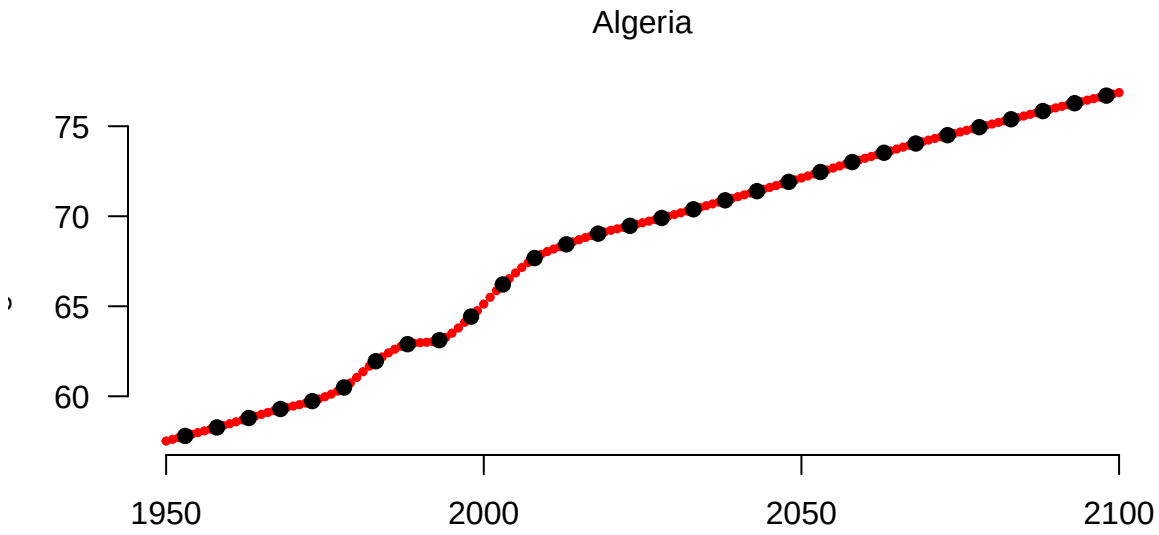

Figure 2: Interpolation of single year old-age thresholds

$RLE^{15}_{y=i}$

We also have the population data, single-year age groups, for every year.

$Pop^{x=k}_{y=i}$ where $k = 0, 1, 2...80+/100+$ and $y = 1950, 1951..2100$

(Some countries have data capped at 80, others at 100, but in neither case are the thresholds anywhere near that end, so this is not relevant to our calculations).[1]

So in order to get the proportion of the population over the old age-threshold, we now need to interpolate the population. E.g. there are 30,000 people aged 62 but not yet 63. How many people are aged 62.3? Again, this interpolation could be done linearly, but since we have information on the population at ages 61 and 63 etc, it makes sense to use that in the calculation and use splines again.

But of course (it turns out after a lot of odd results) I need to interpolate between the cumulative populations! So here is how this works via Figure 3. Here we start at the old-age threshold reading it off the x-axis–it's 62.89 for Algeria in 1988, and interpolate the cumulative population that is under that age from the y-axis. This cumulative population $Pop^{x<RLE^{15}_{y=i}}_{y=i}$ is then divided by the total population in year $y = i$ to get the proportion *under* the old-age threshold, which is subtracted from one to get the proportion *over* the threshold:

$$Prop > RLE^{15} = 1 - Prop < RLE^{15} = 1 - \frac{Pop^{x<RLE^{15}_{y=i}}_{y=i}}{\sum_{x=1}^{x=80+/100+} Pop^{x}_{y=i}}$$

References

Sanderson, Warren, and Sergei Scherbov. 2008. *Rethinking age and aging.* Population Reference Bureau Washington, DC.

Scherbov, S, S Andruchowitz, and W Sanderson. 2018. "Aging Demographic Data Sheet 2018." International Institute for Applied Systems Analysis.

[1]The UN Standard Projections dataset is inconsistent here in that the variable **AgeGrp** has the value "80+" in the case of the first set of countries but simply "100" in the second instead of "100+".

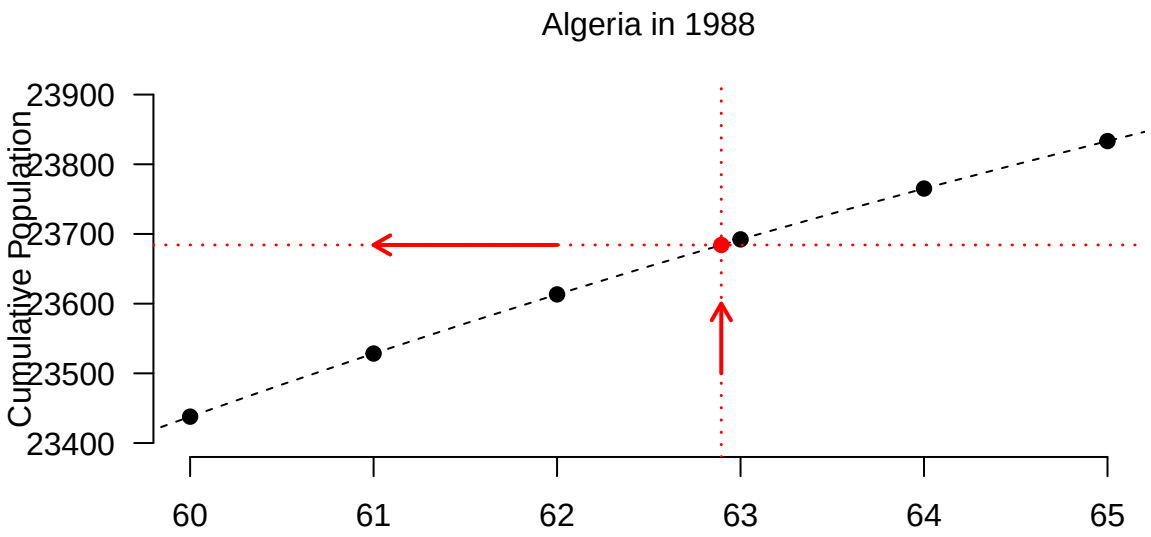

Figure 3: Figure 3: Interpolation of populaiton over old-age threhsold

UN. 2017. *World Population Prospects: The 2017 Revision.* New York: Deartment of Economic; Social Affairs, Population Division.