data analysis methods for MENA factscheet PH 14.2 - May 2018

mz

15 August, 2018

Data

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

Calculations

Old-age threshold

Calculating the *old-age threshold*, which is the age at which the remaining life expectancy is 15 years, was based on the abridged life tables which has **ex** values for five year groups, so I used splines to interpolate the

age where **ex** equals 15.

Although even a linear interpolation would not change the calculations dramatically, so this is a minor point, but still..

I use the R stats function `splinefun()` and the monotone Hermite spline cumputation 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.

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

Furthermore, the life tables are for five-year time periods as well. So after step one–finding the age at which ex in 15, for every 5-year period–comes step two, 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):

$x = 0, 1, 5, 10, ...75, 80$

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 funciton of age, with which we can interpolate the age x at which $e_x = 15$ for each time period $y = i_5$.

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 funciton is monotonic–or rather that it is monotonic in the area that I'm interested in. Which is fine, because teh 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), $RLE_{y=i}^{15}$, and in this case we only have it for five year time periods so $RLE_{y=i_5}^{15}$. RLE stands for remainling 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 the MidPeriod variable is assumed as the correct for each old-age threshold, and the intervening periods are then interpolated. Because of this I don't have data for the first and last couple of years: the first time period is 1950-55 with 1953 as the MidPeriod. This is treated as the input data for the interpolation of the individual year thresholds.

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

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

Proportion over old-age threshold

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

$RLE_{y=i}^{15}$

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

$Pop_{y=i}^{x=k}$ where $k = 0, 1, 2...80$ and $y = 1950, 1951..2100$

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.
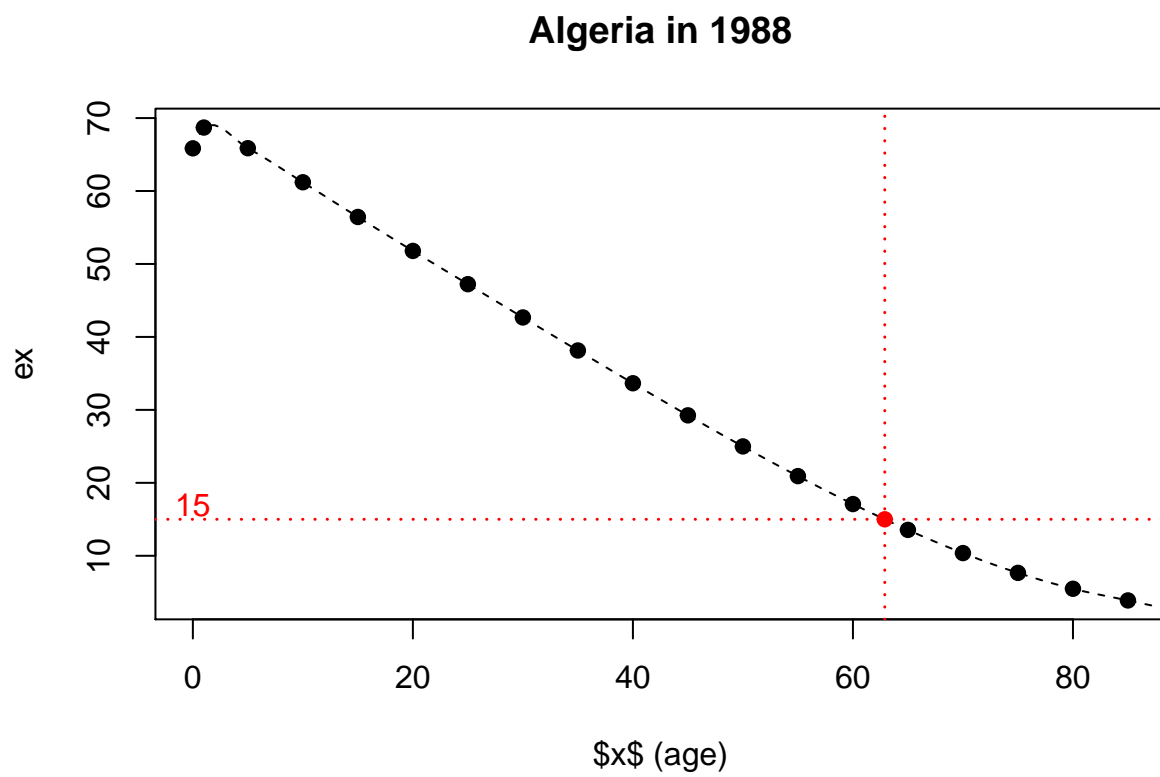
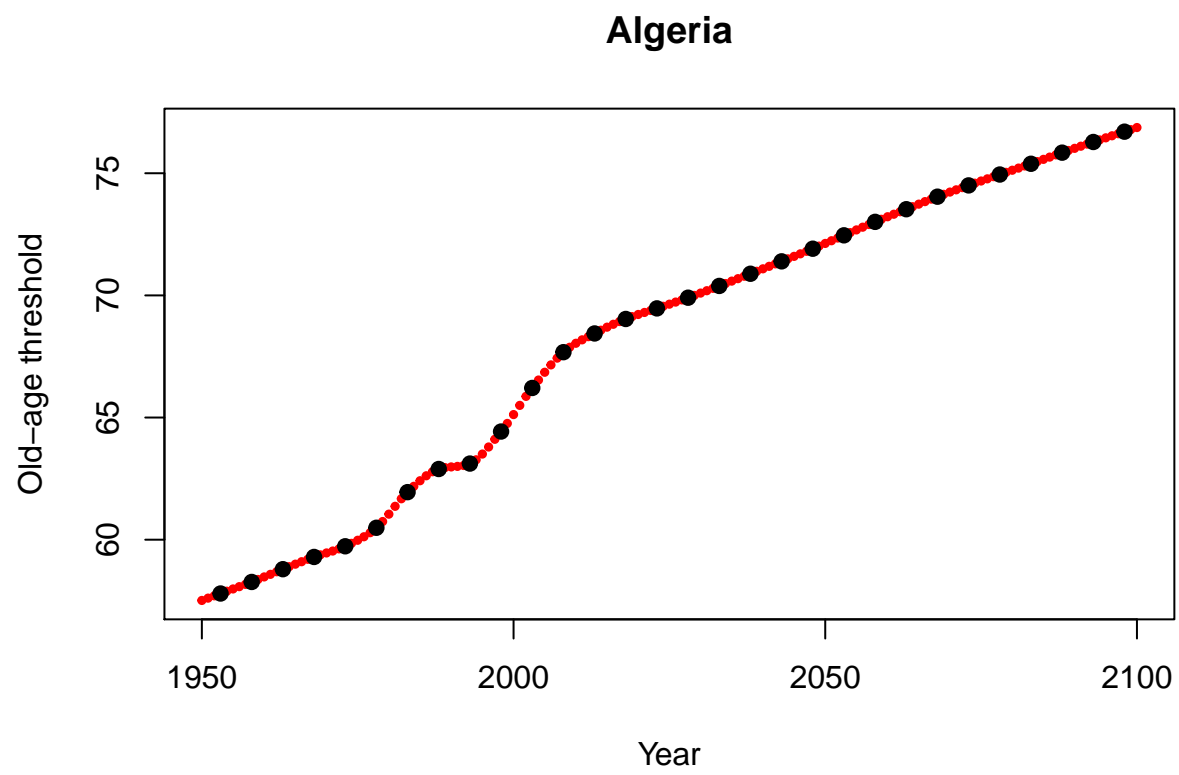Figure 1: Figure 1:: Interpolation of old-age threhsold: x where ex is 15.

Figure 2: Interpolation of single year old-age thresholds
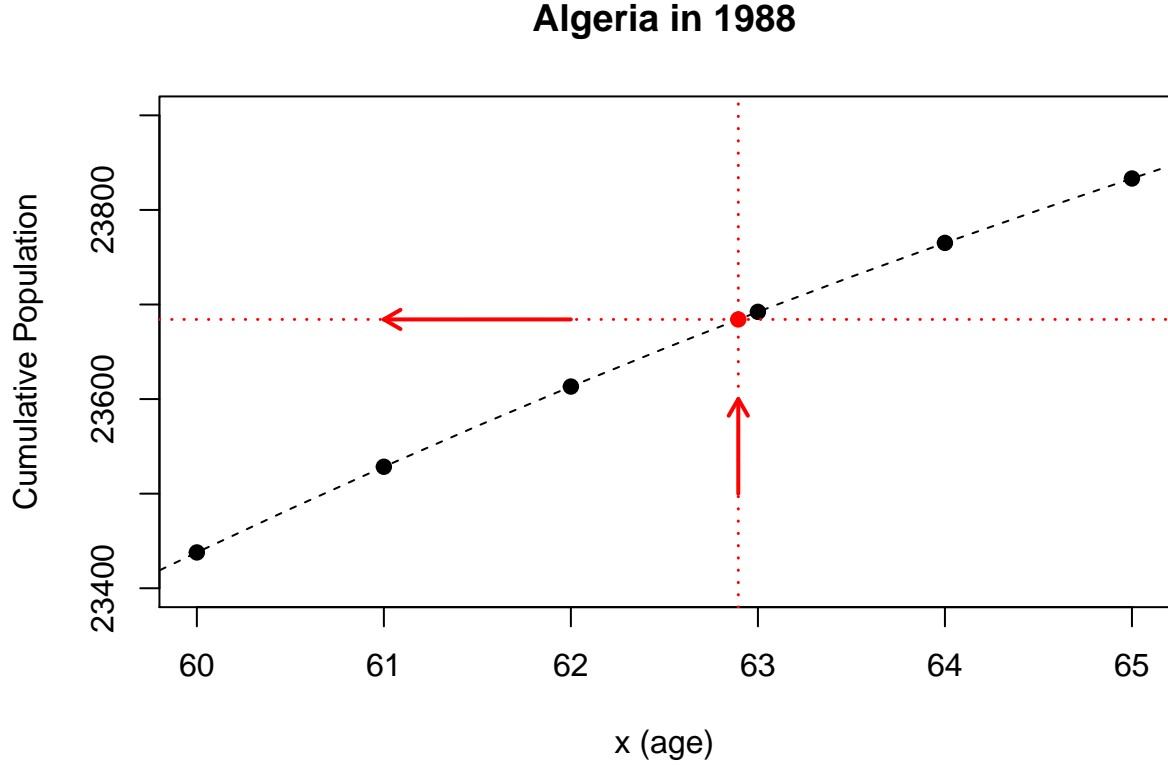
Algeria in 1988

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

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 $Pop_{y=i}^{x<RLE_{y=i}^{15}}$ is then divided by the total population in year $y = i$ to get the proportion over the old-age threshold.

$$Prop > RLE^{15} = \frac{Pop_{y=i}^{x<RLE_{y=i}^{15}}}{\sum_{x=1}^{x=80+} Pop_{y=i}^x}$$

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.