Exercise 1: Decomposing the crude death rate using Kitagawa decomposition

2022-05-06

- t_1 is the initial period and t_2 is the final period
- D_x = number of deaths at age x
- M_x = death rate
- N_x = is the mid-year population
- $N = $ total population over ages

Note that $D_x = M_x * N_x$ and that the difference between two crude death rates can be expressed as

$$\Delta CDR = \sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)} - \sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)}. \tag{1}$$

The aim is to disentangle the effects of changing rates and the effect of the change in the composition of the population.

The file Kitagawa_Data.RData contains mortality and exposure data for Israel in 1990 and 2016. The exercise consists in decomposing the changes in CDR following Kitagawa or equation (1)

Start by loading the data.

```
load('Kitagawa_Data.RData')
```

The exercises in this course will use the package tidyverse, which you should have already met with Tim Riffe.

We need to load this package too (and install it, if you haven't already).

```
#install.packages("tidyverse") #to install the package

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

The period 1 corresponds to 1990 and period 2 to 2016 in this case. Start by selecting deaths and exposure (approximated usually by the midyear population) for each of the periods and store them in vectors.

```
# time period 1
# Select deaths for first period
Dx1 <- Deaths %>%
  filter(Year==1990) %>%
  pull(Total)
# Select population for first period
```

```r
Nx1 <- Exposures %>%
  filter(Year==1990) %>%
  pull(Total)
# Do the same for period 2
Dx2 <- Deaths%>%
  filter(Year==2016) %>%
  pull(Total)
Nx2 <- Exposures%>%
  filter(Year==2016) %>%
  pull(Total)
```

Following equation one, we need to calculate the age-specific mortality rates, which is the quotient of deaths divided by the exposure for both periods.

```r
# time period 1
# Select deaths for first period
# get age-specific mortality rates
Mx1 <- Dx1/Nx1
# replace NA's with zero (just for simplicity)
Mx1 <- Mx1 %>%
  replace_na(replace=0)
# dot the same for the second period
Mx2 <- Dx2/Nx2
Mx2 <- Mx2 %>%
  replace_na(replace=0)
```

Now we have everything to calculate the crude death rate.

```r
# get the crude death rate in time 1
CDR1 <- sum(Dx1)/sum(Nx1)
#crude death expressed by 1000 in time 1
CDR1*1000
```

```
## [1] 6.686454
```

```r
# the same for period 2
CDR2 <- sum(Dx2)/sum(Nx2)
#crude death expressed by 1000 in time 2
CDR2*1000
```

```
## [1] 6.882632
```

```r
#change in CDR
(Dif <- (CDR2 - CDR1)*1000)
```

```
## [1] 0.1961774
```

**What would we have expected in this period? what do the CDRs suggest? Does it make sense? Let's decompose and see if effectively mortality went up or if it is a compositional effect.

$$\Delta CDR = \underbrace{\sum_x \left(\frac{M_x(t_2)+M_x(t_1)}{2}\right)\left(\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right)}_{\text{Changes in x-composition}} + \underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2}\right)(M_x(t_2)-M_x(t_1))}_{\text{Changes in rates}} \tag{2}$$

```
RC <- sum(0.5*(Nx2/sum(Nx2) + Nx1/sum(Nx1))*(Mx2-Mx1))
RC*1000
```

```
## [1] -1.898981
```

```
CC <- sum(0.5*(Mx2+Mx1)*(Nx2/sum(Nx2)-Nx1/sum(Nx1)))
CC*1000
```

```
## [1] 2.095158
```

Compare the decomposition results with the original difference

```
RC*1000 + CC*1000
```

```
## [1] 0.1961774
```

```
Dif
```

```
## [1] 0.1961774
```