# Day 2, exercise 2: Decomposing the crude marriage rate using Kitagawa decomposition

2022-05-06

In Exercise 1, we used the Kitagawa method to decompose the crude death rate by age. However, this method can be used to decompose any crude rate along any dimension. In this exercise we will use the Kitagawa decomposition to analyse differences in crude marriage rates by country of birth. The equations and method do not change:

- $t_1$ is the initial period and $t_2$ is the final period
- $m_x$ = number of marriages for group $x$
- $M_x$ = marriage rate
- $N_x$ = is the mid-year population
- $N = $ total population

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

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

The file "MarriageKitagawa.RData" contains data on marriages and exposure data for Spain and the Netherlands in 2019 by country of birth (extracted from Eurostat). The exercise consists in decomposing the changes in CMR following Kitagawa. Once again, the aim is to disentangle the contribution of changing rates and the contribution of the change in the composition of the population.

Start by loading the data and the tidyverse package.

```r
load('MarriageKitagawa.RData')

library(tidyverse)
```

Start by selecting marriages and exposure (approximated here by the population at 1st January) for each of the countries and store them in vectors.

```r
# Select marriages for Spain
mx_ESP <- data %>%
  filter(country=="Spain") %>%
  pull(marriages)
# Select population for Spain
Nx_ESP <- data %>%
  filter(country=="Spain") %>%
  pull(exposure)
# Do the same for the Netherlands
mx_NLD <- data %>%
  filter(country=="Netherlands") %>%
  pull(marriages)
Nx_NLD <- data %>%
  filter(country=="Netherlands") %>%
  pull(exposure)
```

Now we have everything to calculate the crude marriage rates.

```
# get the crude marriage rate in Denmark
CMR_ESP <- sum(mx_ESP)/sum(Nx_ESP)
#crude marriage expressed by 1000 in Denmark
CMR_ESP*1000
```

```
## [1] 3.203679
```

```
# the same for Slovenia
CMR_NLD <- sum(mx_NLD)/sum(Nx_NLD)
#crude marriage expressed by 1000 in Slovenia
CMR_NLD*1000
```

```
## [1] 3.225619
```

```
#change in CDR
Dif <- (CMR_NLD - CMR_ESP)*1000
```

What did you expect? What do the CMRs suggest? Does it make sense?

Let's decompose and see if the difference in crude marriage rates between Spain and the Netherlands is due to a compositional effect in the population or to differences in the marriage rates.

Again, remember Kitagawa's formula:

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

We need to calculate the country of birth-specific marriage rates, which are the quotient of marriages per subpopulation divided by the exposure for both countries.

```
# get country-of-birth-specific marriage rates
# Spain
Mx_ESP <- mx_ESP/Nx_ESP
# Netherlands
Mx_NLD <- mx_NLD/Nx_NLD
```

Get the components.

```
RC <- sum(0.5*(Nx_NLD/sum(Nx_NLD) + Nx_ESP/sum(Nx_ESP))*(Mx_NLD-Mx_ESP))
RC*1000
```

```
## [1] 0.03630046
```

```
CC <- sum(0.5*(Mx_NLD+Mx_ESP)*(Nx_NLD/sum(Nx_NLD)-Nx_ESP/sum(Nx_ESP)))
CC*1000
```

```
## [1] -0.01436017
```

What do these results indicate? How are marriage rates different in Spain and the Netherlands? How much of the difference in crude marriage rates is linked to to compositional differences?

Compare the decomposition results with the original difference

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

```
## [1] 0.02194029
```

```r
Dif
```

```
## [1] 0.02194029
```