

Day 1, exercise 1: Decomposing the general fertility rate using Kitagawa decomposition

Serena Vigizzi

2025-05-27

- t_1 is the initial period and t_2 is the final period
- B_x = number of deaths at age x
- $ASFR_x$ = age-specific fertility rate
- N_x = is the mid-year population
- N = total population over ages

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

$$\Delta GFR = \sum_x ASFR_x(t_2) \frac{N_x(t_2)}{N(t_2)} - \sum_x ASFR_x(t_1) \frac{N_x(t_1)}{N(t_1)}. \quad (1)$$

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 births and exposure (for women between ages 12 and 55) data for Chile and Ireland in 2020. The exercise consists in decomposing the changes in GFR following Kitagawa.

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 core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Population 1 corresponds to Chile and population 2 to Ireland in this case. Start by selecting deaths and exposures (approximated usually by the midyear population) for each of the populations and store them in vectors.

```

# Select births for population 1
Bx1 <- data %>%
  filter(pop=="CHL") %>%
  pull(births)
# Select population for population 1
Nx1 <- data %>%
  filter(pop=="CHL") %>%
  pull(exposure)
# Do the same for population 2
Bx2 <- data %>%
  filter(pop=="IRL") %>%
  pull(births)
Nx2 <- data %>%
  filter(pop=="IRL") %>%
  pull(exposure)

```

We can calculate the general fertility rates directly from these.

```

# get the general fertility rate in population 1
GFR1 <- sum(Bx1)/sum(Nx1)
#general fertility expressed by 1000 in population 1
GFR1*1000

```

```
## [1] 36.14496
```

```

# the same for period 2
GFR2 <- sum(Bx2)/sum(Nx2)
#general fertility expressed by 1000 in population 2
GFR2*1000

```

```
## [1] 38.39096
```

```

#difference in GFR
(Diff <- (GFR2 - GFR1)*1000)

```

```
## [1] 2.246003
```

What would we have expected in this period? What do the GFRs suggest? Does it make sense?

Let's decompose and see if effectively fertility rates declined or if it is a compositional effect.

Remember Kitagawa's formula:

$$\begin{aligned}
 \Delta GFR = & \underbrace{\sum_x \left(\frac{ASFR_x(t_2) + ASFR_x(t_1)}{2} \right) \left(\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)} \right)}_{\text{Changes in x-composition}} + \\
 & \underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) (ASFR_x(t_2) - ASFR_x(t_1))}_{\text{Changes in rates}}
 \end{aligned} \tag{2}$$

We need to calculate the age-specific fertility rates, which are the age-specific births divided by the age-specific exposures for both periods.

```

# population 1
# Select births for first population

```

```

# get age-specific fertility rates
ASFR1 <- Bx1/Nx1
# replace NA's with zero (just for simplicity)
ASFR1 <- ASFR1 %>%
  replace_na(replace=0)
# do the same for the second population
ASFR2 <- Bx2/Nx2
ASFR2 <- ASFR2 %>%
  replace_na(replace=0)

RC <- sum(0.5*(Nx2/sum(Nx2) + Nx1/sum(Nx1))*(ASFR2-ASFR1))
RC*1000

```

```
## [1] 4.906327
```

```

CC <- sum(0.5*(ASFR2+ASFR1)*(Nx2/sum(Nx2)-Nx1/sum(Nx1)))
CC*1000

```

```
## [1] -2.660324
```

What do they tell us about differences in fertility between Chile and Ireland? What was the role of population composition vs fertility rates?

Let us check that the decomposition results fit with the original difference.

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

```
## [1] 2.246003
```

```
Diff
```

```
## [1] 2.246003
```