Exercise 2

2023-04-17

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 TFR. Moreover, we will decompose along two dimensions: age and partnership status. This exercise is based on Nishikido, Cui and Esteve (2022), although the data have been simulated for privacy purposes.

The elements we need are the same as in Exercise 1:

- t_1 is the initial period and t_2 is the final period
- b_{ix} = number of births for age i and group x
- $ASFR_x$ = age-specific fertility rate for group x
- N_{ix} = is the mid-year population for age i and group x
- $N = $ total population

We know that $TFR_x = \sum_x ASFR(x)$ and that the difference between two TFRs can be expressed as

$$\Delta TFR = \sum_x \sum_p ASFR(x,p,t=t_2) \times \frac{N(x,p,t=t_2)}{N(p,t=t_2)} - \sum_x \sum_p ASFR(x,p,t=t_1) \times \frac{N(x,p,t=t_1)}{N(p,t=t_1)} \quad (1)$$

Note the double summation, one for age and one for partnership status. Because we are decomposing along two dimensions, we need a double summation also in the Kitagawa formula:

$$\begin{aligned}\Delta TFR = &\sum_x \sum_p (ASFR(x,p,t=t_2) - ASFR(x,p,t=t_1)) \times (\frac{N(x,p,t=t_2)}{N(p,t=t_2)} + \frac{N(x,p,t=t_1)}{N(p,t=t_1)}) \times 0.5 + \\ &\sum_x \sum_p (ASFR(x,p,t=t_2) + ASFR(x,p,t=t_1)) \times 0.5 \times (\frac{N(x,p,t=t_2)}{N(p,t=t_2)} - \frac{N(x,p,t=t_1)}{N(p,t=t_1)})\end{aligned}$$

The file "NishikidoCuiEsteve_Kitagawa.RData" contains simulated data on births and exposures for women born between 1965 and 1969 in Spain and Sweden by age and partnership status (kindlly supplied by the authors). The exercise consists in decomposing the changes in TFR following Kitagawa. Once again, the aim is to disentangle the contribution of differences in rates and the contribution of differences in the composition of the population.

Start by loading the data and the tidyverse package.

```
load('NishikidoCuiEsteve_Kitagawa.RData')

library(tidyverse)
```

Then select births and exposures for each of the countries and store them in vectors.

```
# Select births for Spain, for unpartnered and partnered women
bx0_ESP <- data %>%
  filter(pop=="ESP", partnered==0) %>%
```

```r
  pull(births)
bx1_ESP <- data %>%
  filter(pop=="ESP", partnered==1) %>%
  pull(births)
# Select population for Spain, for unpartnered and partnered women
Nx0_ESP <- data %>%
  filter(pop=="ESP", partnered==0) %>%
  pull(exposure)
Nx1_ESP <- data %>%
  filter(pop=="ESP", partnered==1) %>%
  pull(exposure)
# Do the same for Sweden
bx0_SWE <- data %>%
  filter(pop=="SWE", partnered==0) %>%
  pull(births)
bx1_SWE <- data %>%
  filter(pop=="SWE", partnered==1) %>%
  pull(births)
Nx0_SWE <- data %>%
  filter(pop=="SWE", partnered==0) %>%
  pull(exposure)
Nx1_SWE <- data %>%
  filter(pop=="SWE", partnered==1) %>%
  pull(exposure)
```

First, we look at the difference in TFRs.

```r
# Spain
TFR_ESP <- sum(bx1_ESP+bx0_ESP)/(Nx1_ESP+Nx0_ESP)[1]
# the same for Sweden
TFR_SWE <- sum(bx1_SWE+bx0_SWE)/(Nx1_SWE+Nx0_SWE)[1]
# difference in TFR
(diff <- TFR_SWE - TFR_ESP)
```

```
## [1] 0.1306009
```

Let's decompose it and see if the difference in the fertility rate between Spain and Sweden is due to a composition effect in the population or to differences in the partnership- and age-specific fertility rates.

We need to calculate the partnership- and age-specific fertility rates, which are the number of births per subpopulation divided by the exposure for both countries. We also need the exposures, by country, partnership status and age.

```r
# Partnership-specific ASFR for Spain
ASFR0_ESP <- bx0_ESP/Nx0_ESP
ASFR1_ESP <- bx1_ESP/Nx1_ESP
# And Sweden
ASFR0_SWE <- bx0_SWE/Nx0_SWE
ASFR1_SWE <- bx1_SWE/Nx1_SWE
```

Now, we calculate the Kitagawa components. They need to be specific to each age and partnership status combination.

```r
# Unpartnered women
RC0 <- 0.5*(Nx0_SWE/(Nx0_SWE+Nx1_SWE) + Nx0_ESP/(Nx0_ESP+Nx1_ESP))*(ASFR0_SWE-ASFR0_ESP)
CC0 <- 0.5*(ASFR0_SWE+ASFR0_ESP)*(Nx0_SWE/(Nx0_SWE+Nx1_SWE)-Nx0_ESP/(Nx0_ESP+Nx1_ESP))
```

```r
# Partnered women
RC1 <- 0.5*(Nx1_SWE/(Nx0_SWE+Nx1_SWE) + Nx1_ESP/(Nx0_ESP+Nx1_ESP))*(ASFR1_SWE-ASFR1_ESP)
CC1 <- 0.5*(ASFR1_SWE+ASFR1_ESP)*(Nx1_SWE/(Nx0_SWE+Nx1_SWE)-Nx1_ESP/(Nx0_ESP+Nx1_ESP))
```

Then, we sum across partnership status. This is the first (most internal) summation.

```r
# Sum across partnership status (first summation)
RC_age <- RC0+RC1
CC_age <- CC0+CC1
```

Then, we sum across ages, to get the total rate and composition effects.

```r
# Sum across ages (second summation)
RC <- sum(RC_age)
CC <- sum(CC_age)
```

Let's check the results

```r
RC+CC
```

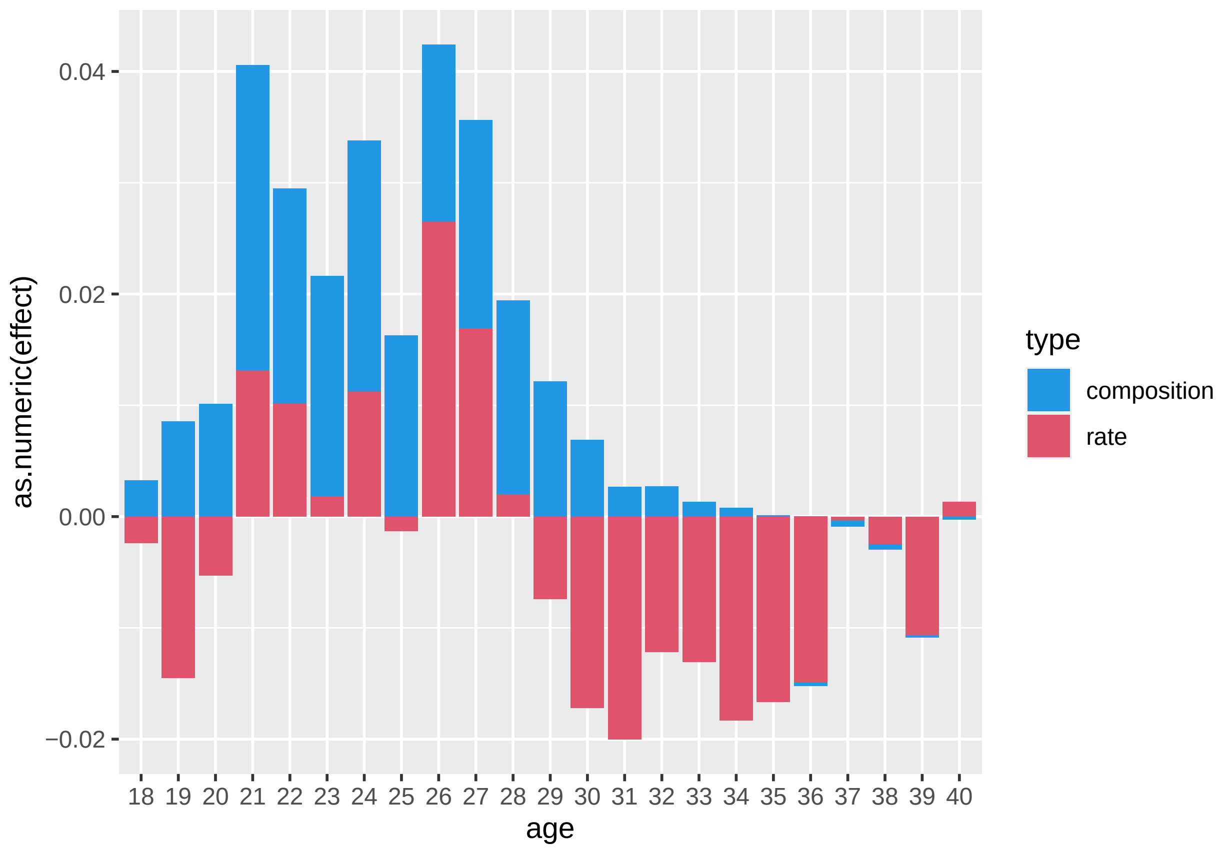
```
## [1] 0.1306009
```

```r
diff
```

```
## [1] 0.1306009
```

And we can make a nice plot to get the contributions by age, divided in rate and composition effects.

```r
cbind(age=18:40,type="rate",effect=RC_age) %>%
  rbind(cbind(age=18:40,type="composition",effect=CC_age)) %>%
  as.data.frame() %>%
ggplot() +
  geom_bar(aes(x=age, y=as.numeric(effect), fill=type), stat="identity", position="stack") +
  scale_fill_manual(values=c(4,2))
```


0.04
0.02
0.00
−0.02
as.numeric(effect)
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
age
type
composition
rate