

# European Doctoral School of Demography (EDSD)

## Decomposition Techniques - Final Assignment

Ainhoa-Elena Leger  
Gonzalo Daniel Garcia  
Liliana Patricia Calderon Bernal  
Marilyn-Anne Tremblay  
Özge Elif Özer

26/5/2021

### Challenge 1: proof Kitagawa decomposition (1995) without interactions

Define the difference between the crude death rates as  $\Delta$ .

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

I divide each of the terms into two equal parts and add and subtract some additional terms, thereby keeping the difference ( $\Delta$ ) constant.

$$\begin{aligned} \Delta\text{CDR} = & \frac{\sum_x M_x(t_2) \frac{N_x(t_2)}{N(t_2)}}{2} + \frac{\sum_x M_x(t_2) \frac{N_x(t_2)}{N(t_2)}}{2} - \frac{\sum_x M_x(t_1) \frac{N_x(t_1)}{N(t_1)}}{2} - \frac{\sum_x M_x(t_1) \frac{N_x(t_1)}{N(t_1)}}{2} \\ & + \frac{\sum_x M_x(t_1) \frac{N_x(t_2)}{N(t_2)}}{2} - \frac{\sum_x M_x(t_1) \frac{N_x(t_2)}{N(t_2)}}{2} + \frac{\sum_x M_x(t_2) \frac{N_x(t_1)}{N(t_1)}}{2} - \frac{\sum_x M_x(t_2) \frac{N_x(t_1)}{N(t_1)}}{2} \end{aligned}$$

I now combine the eight terms in  $\Delta$  into four:

$$\begin{aligned} \Delta\text{CDR} = & \sum_x \frac{N_x(t_2)}{N(t_2)} \left( \frac{M_x(t_2) + M_x(t_1)}{2} \right) - \sum_x \frac{N_x(t_1)}{N(t_1)} \left( \frac{M_x(t_2) + M_x(t_1)}{2} \right) \\ & + \sum_x M_x(t_2) \left( \frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) - \sum_x M_x(t_1) \left( \frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right). \end{aligned}$$

Finally, we combine the terms into two:

$$\Delta\text{CDR} = \sum_x \left( \frac{M_x(t_2) + M_x(t_1)}{2} \right) \left( \frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)} \right) + \sum_x \left( \frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) (M_x(t_2) - M_x(t_1)).$$

The first terms is the difference in age composition weighted by the average age-specific mortality, while the second term is the difference in rate schedules weighted by the average age composition. Therefore,  $\Delta$  is equal to the sum of the contribution of age compositional differences and the contribution of rate schedule differences.

**Challenge 2:** With data on fertility (e.g. HFD) select 3 countries and analyze the change in their crude fertility rate (CFR) in a recent period (10 years) and decompose these changes following Kitagawa's decomposition and describe your results. Then for the most recent period select the two countries (among the 3) with the highest and lowest CFR and decompose their difference and describe your results.

```
# Loading packages
library(tidyverse)
library(HMDHFDplus)

# Check countries
getHFDcountries()

## [1] "AUT"      "BLR"      "BGR"      "CAN"      "CHL"      "HRV"      "CZE"
## [8] "DNK"      "EST"      "FIN"      "FRATNP"   "DEUTNP"   "DEUTW"    "DEUTE"
## [15] "HUN"      "ISL"      "ITA"      "JPN"      "LTU"      "NLD"      "NOR"
## [22] "POL"      "PRT"      "KOR"      "RUS"      "SVK"      "SVN"      "ESP"
## [29] "SWE"      "CHE"      "TWN"      "UKR"      "GBR_NP"   "GBRTENW"  "GBR_SCO"
## [36] "GBR_NIR"  "USA"

# Downloading data for 3 countries: Spain, Bulgaria, Korea

# Spain
spain_birth <- readHFDweb(CNTRY = "ESP",
  item = "birthsTR",
  username = "gonzalo.fce@gmail.com",
  password = "fermat31416")

spain_exposure <- readHFDweb(CNTRY = "ESP",
  item = "exposTR",
  username = "gonzalo.fce@gmail.com",
  password = "fermat31416")

# Czechia
cze_birth <- readHFDweb(CNTRY = "CZE",
  item = "birthsTR",
  username = "gonzalo.fce@gmail.com",
  password = "fermat31416")

cze_exposure <- readHFDweb(CNTRY = "CZE",
  item = "exposTR",
  username = "gonzalo.fce@gmail.com",
  password = "fermat31416")

# Korea
kor_birth <- readHFDweb(CNTRY = "KOR",
  item = "birthsTR",
  username = "gonzalo.fce@gmail.com",
  password = "fermat31416")

kor_exposure <- readHFDweb(CNTRY = "KOR",
```

```

        item = "exposTR",
        username = "gonzalo.fce@gmail.com",
        password = "fermat31416")

# Wrangling time

# Spain
spain_birth2 <- spain_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()

spain_exposure2 <- spain_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

# Czechia
cze_birth2 <- cze_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()

cze_exposure2 <- cze_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

# Korea
kor_birth2 <- kor_birth %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Births = sum(Total, na.rm = TRUE)) %>%
  ungroup()

kor_exposure2 <- kor_exposure %>%
  filter(Age %in% c(13:54)) %>%
  group_by(Year, Age) %>%
  summarise(Exposure = sum(Exposure, na.rm = TRUE)) %>%
  ungroup()

# All together now
spain <- spain_exposure2 %>% left_join(spain_birth2, by = c("Year", "Age"))
czechia <- cze_exposure2 %>% left_join(cze_birth2, by = c("Year", "Age"))
korea <- kor_exposure2 %>% left_join(kor_birth2, by = c("Year", "Age"))

# All together in 1 data.frame
spain$cty <- "Spain"

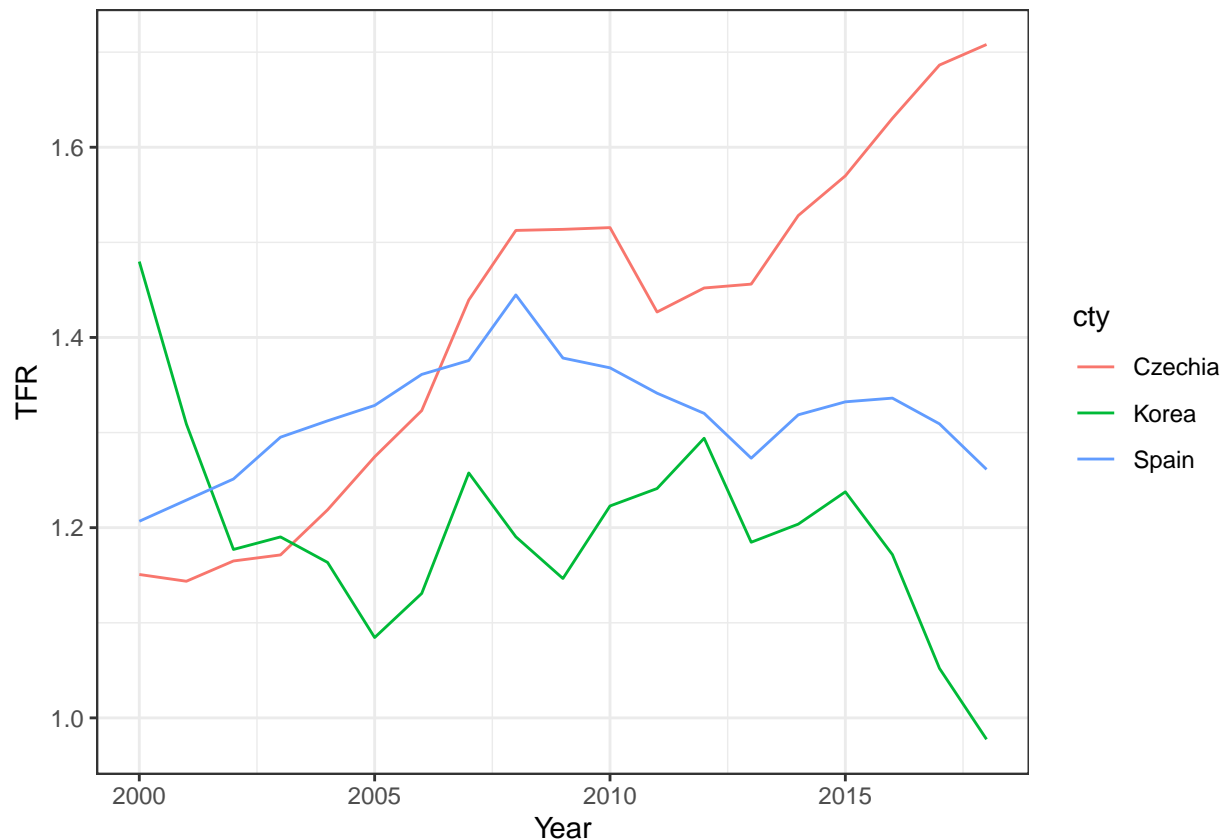
```

```

czechia$cty <- "Czechia"
korea$cty <- "Korea"
challenge2 <- rbind(spain, czechia, korea)

# Some plotting
challenge2 %>%
  filter(Age >= 15, Age <= 49) %>%
  mutate(TFR_age = Births/Exposure) %>%
  group_by(cty, Year) %>%
  summarise(TFR = sum(TFR_age, na.rm = T)) %>%
  filter(Year >= 2000) %>%
  ggplot(aes(x=Year, y = TFR, color = cty)) + geom_line() +
  theme_bw()

```



```

# Input: exposures and deaths for A and B
# Output: ASCDRs, CDRs, CDR(A)-CDR(B), compositional diff, ASR diff

stand.decom <- function(P.A, B.A, P.B, B.B){

  # CRUDE RATES
  A.crude <- sum(B.A) / sum(P.A)
  B.crude <- sum(B.B) / sum(P.B)
  Diff.crude <- A.crude - B.crude
  # age distribution of country A and B
  C.A <- P.A / sum(P.A)
  C.B <- P.B / sum(P.B)
}

```

```

# age-specific fertility rate in country A and B
F.A <- B.A / P.A
F.B <- B.B / P.B
# average age distribution
C.ave <- (C.A + C.B)/2

#### DECOMPOSITION OF DIFFERENCES BETWEEN RATES
comp.diff <- sum((C.A - C.B) * ((F.A + F.B)/2))
ASR.diff <- sum((F.A - F.B) * C.ave)

# preparing the outcomes
outcome <- c( Diff.crude = Diff.crude,
              Diff.comp = comp.diff,
              Diff.rates = ASR.diff)

# giving the outcomes
return(outcome)
}

decomp <- matrix(NA,11,3)

for(i in c(2007:2017)){

  spain_birth2a <- spain_birth2 %>%
    filter(Year==i) %>% pull(Births)
  spain_birth2b <- spain_birth2 %>%
    filter(Year==i+1) %>% pull(Births)

  spain_exposure2a <- spain_exposure2 %>%
    filter(Year==i) %>% pull(Exposure)
  spain_exposure2b <- spain_exposure2 %>%
    filter(Year==(i+1)) %>% pull(Exposure)

  # standardization + decomposition
  out <- stand.decom(P.A = spain_exposure2a, B.A = spain_birth2a,
                    P.B = spain_exposure2b, B.B = spain_birth2b)

  for(j in 1:3){ decomp[i-2006,j] <- out[[j]] }

}

decomp <- as.data.frame(decomp)
colnames(decomp) <- c("Diff.crude", "Diff.comp", "Diff.rates")
rownames(decomp) <- c("2007-2008", "2008-2009", "2009-2010", "2010-2011",
                     "2011-2012", "2012-2013", "2013-2014", "2014-2015",
                     "2015-2016", "2016-2017", "2017-2018")

decomp

##           Diff.crude    Diff.comp    Diff.rates
## 2007-2008 -0.0016370674 0.0001391127 -0.0017761801
## 2008-2009 0.0019547642 0.0002993576 0.0016554066
## 2009-2010 0.0005636163 0.0004150469 0.0001485694
## 2010-2011 0.0009715521 0.0004915310 0.0004800211
## 2011-2012 0.0010494831 0.0005735937 0.0004758894

```

```
## 2012-2013 0.0017861523 0.0006621398 0.0011240126
## 2013-2014 -0.0005179004 0.0007274552 -0.0012453557
## 2014-2015 0.0002855512 0.0007375776 -0.0004520265
## 2015-2016 0.0005510885 0.0007440104 -0.0001929219
## 2016-2017 0.0012558197 0.0007013645 0.0005544551
## 2017-2018 0.0016002491 0.0005962147 0.0010040344
```

**Challenge 4** Use the linear integral model to decompose the change in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death for 3 countries you might be interested in (over time or between them). Interpret the results of life expectancy alongside standard deviation. Make it interesting. You can use data from HCoD, HMD, WHO, GBD.