Understanding World Population Dynamics

Assignment 1 - PSYC593

Jordan M Navarro

2023-09-07

Understanding population dynamics is important for many areas of social science. We will calculate some basic demographic quantities of births and deaths for the world's population from two time periods: 1950 to 1955 and 2005 to 2010. We will analyze the following CSV data files - `Kenya.csv`, `Sweden.csv`, and `World.csv`. Each file contains population data for Kenya, Sweden, and the world, respectively. The table below presents the names and descriptions of the variables in each data set.

Name	Description
`country`	Abbreviated country name
`period`	Period during which data are collected
`age`	Age group
`births`	Number of births in thousands (i.e., number of children born to women of the age group)
`deaths`	Number of deaths in thousands
`py.men`	Person-years for men in thousands
`py.women`	Person-years for women in thousands

Source: United Nations, Department of Economic and Social Affairs, Population Division (2013). *World Population Prospects: The 2012 Revision, DVD Edition.*

```
# Global variables

# Path variables

root_path <- rprojroot::find_root(has_dir("src")) # This is the root/working directory

# These are the subdirectories
code_path <- file.path(root_path, "src") # Using the base R function file.path
```

```r
docs_path <- file.path(root_path, "doc")
raw_data_path <- file.path(root_path, "data", "raw_data")
figs_path <- file.path(root_path, "results", "figures")
tables_path <- file.path(root_path, "results", "tables")
```

```r
# Read data
world_data  <- read.csv(file = file.path(raw_data_path, "World.csv"))
kenya_data  <- read.csv(file = file.path(raw_data_path, "Kenya.csv"))
sweden_data <- read.csv(file = file.path(raw_data_path, "Sweden.csv"))
```

The data are collected for a period of 5 years where *person-year* is a measure of the time contribution of each person during the period. For example, a person that lives through the entire 5 year period contributes 5 person-years whereas someone who only lives through the first half of the period contributes 2.5 person-years.

Question 1

We begin by computing *crude birth rate* (CBR) for a given period. The CBR is defined as:

$$\text{CBR} = \frac{\text{number of births}}{\text{number of person-years lived}}$$

Compute the CBR for each period, separately for Kenya, Sweden, and the world. Start by computing the total person-years, recorded as a new variable within each existing `data.frame` via the `$` operator, by summing the person-years for men and women. Then, store the results as a vector of length 2 (CBRs for two periods) for each region with appropriate labels. You may wish to create your own function for the purpose of efficient programming. Briefly describe patterns you observe in the resulting CBRs.

Answer 1

```r
# Create new variable py = total person years for each data set
world_data$py  <- world_data$py.men  + world_data$py.women
kenya_data$py  <- kenya_data$py.men  + kenya_data$py.women
sweden_data$py <- sweden_data$py.men + sweden_data$py.women
```

```r
# Function to compute the Crude Birth Rate (CBR)
compute_cbr <- function(population_data) {
  population_data %>%
    group_by(period) %>%
```

```r
    summarise(cbr = sum(births) / sum(py)) %>%
    pull()
}
```

```r
# Compute the CBR for each data set
(world_cbr  <- compute_cbr(world_data))
```

```
[1] 0.03732863 0.02021593
```

```r
(kenya_cbr  <- compute_cbr(kenya_data))
```

```
[1] 0.05209490 0.03851507
```

```r
(sweden_cbr <- compute_cbr(sweden_data))
```

```
[1] 0.01539614 0.01192554
```

The CBR for the world from 1950 to 1955 is greater than its CBR from 2005 to 2010. The CBR for Keyna from 1950 to 1955 is greater than its CBR from 2005 to 2010. The CBR for Sweden from 1950 to 1955 is very slightly greater than its CBR from 2005 to 2010. Kenya has the greatest CBR across itself, Sweden, and the world for both the periods 1950 to 1955 and 2005 to 2010.

Question 2

The CBR is easy to understand but contains both men and women of all ages in the denominator. We next calculate the *total fertility rate* (TFR). Unlike the CBR, the TFR adjusts for age compositions in the female population. To do this, we need to first calculate the *age specific fertility rate* (ASFR), which represents the fertility rate for women of the reproductive age range $[15, 50)$. The ASFR for age range $[x, x + \delta)$, where x is the starting age and δ is the width of the age range (measured in years), is defined as:

$$\text{ASFR}_{[x,\ x+\delta)} = \frac{\text{number of births to women of age } [x,\ x+\delta)}{\text{Number of person-years lived by women of age } [x,\ x+\delta)}$$

Note that square brackets, [and], include the limit whereas parentheses, (and), exclude it. For example, $[20, 25)$ represents the age range that is greater than or equal to 20 years old and less than 25 years old. In typical demographic data, the age range δ is set to 5 years. Compute the ASFR for Sweden and Kenya as well as the entire world for each of the two periods. Store the resulting ASFRs separately for each region. What does the pattern of these ASFRs say about reproduction among women in Sweden and Kenya?

Answer 2

```
# Function to compute Age specific fertility rate (ASFR)
compute_asfr <- function(population_data) {
  population_data %>%
    mutate(asfr = births / py.women)
}
```

```
# Compute ASFR for each data set
world_data  <- compute_asfr(world_data)
kenya_data  <- compute_asfr(kenya_data)
sweden_data <- compute_asfr(sweden_data)
```

```
# Compare ASFRs for Kenya and Sweden
kenya_data$asfr
```

```
 [1] 0.00000000 0.00000000 0.00000000 0.16884585 0.35596942 0.34657814
 [7] 0.28946367 0.20644016 0.11193267 0.03905205 0.00000000 0.00000000
[13] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
[19] 0.10057087 0.23583536 0.23294721 0.18087964 0.13126805 0.05626214
[25] 0.03815044 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
```

```
sweden_data$asfr
```

```
 [1] 0.0000000000 0.0000000000 0.0000000000 0.0389089519 0.1277108826
 [6] 0.1252436647 0.0873641591 0.0486037714 0.0162101857 0.0013418290
[11] 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
[16] 0.0000000000 0.0000000000 0.0000000000 0.0059709097 0.0507320271
[21] 0.1162085625 0.1322744621 0.0625923991 0.0121600765 0.0006143942
[26] 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
```

This pattern of ASFRs between Kenya and Sweden suggests that reproduction among women was greater in Kenya between 1950 and 1955 than in between 2005 and 2010, and that reproduction among women was greater in Sweden between 1950 and 1955 than in between 2005 and 2010. Across the two countries, reproduction among women was greater in Kenya than Sweden both between 1950 and 1955 *and* between 2005 and 2010.

Question 3

Using the ASFR, we can define the TFR as the average number of children women give birth to if they live through their entire reproductive age.

$$\text{TFR} = \text{ASFR}_{[15,\ 20)} \times 5 + \text{ASFR}_{[20,\ 25)} \times 5 + \cdots + \text{ASFR}_{[45,\ 50)} \times 5$$

We multiply each age-specific fertility rate rate by 5 because the age range is 5 years. Compute the TFR for Sweden and Kenya as well as the entire world for each of the two periods. As in the previous question, continue to assume that women's reproductive age range is $[15, 50)$. Store the resulting two TFRs for each country or the world as a vector of length two. In general, how has the number of women changed in the world from 1950 to 2000? What about the total number of births in the world?

Answer 3

```
# Function to compute the total fertility rate (TFR)
compute_tfr <- function(population_data) {
  population_data %>%
    group_by(period) %>%
    summarise(tfr = 5 * sum(asfr)) %>%
    pull()
}
```

```
# Compute the TFR for each data set
(world_tfr  <- compute_tfr(world_data))
```

```
[1] 5.007248 2.543623
```

```
(kenya_tfr  <- compute_tfr(kenya_data))
```

```
[1] 7.591410 4.879568
```

```
(sweden_tfr <- compute_tfr(sweden_data))
```

```
[1] 2.226917 1.902764
```

```r
# Compute totals of women and births in the world by period
(world_data %>%
    group_by(period) %>%
    summarise(total_women  = sum(py.women),
              total_births = sum(births)) ->
              totals_world)
```

```
# A tibble: 2 x 3
  period    total_women total_births
  <chr>           <dbl>        <dbl>
1 1950-1955    6555686.      488892.
2 2005-2010   16554781.      674581.
```

```r
# Compare how much these totals have changed
(changes_totals <- totals_world[2,-1] / totals_world[1,-1])
```

```
  total_women total_births
1    2.525256     1.379818
```

From 1950 to 2000, both the number of women in the world and the total number of births increased.

Question 4

Next, we will examine another important demographic process: death. Compute the *crude death rate* (CDR), which is a concept analogous to the CBR, for each period and separately for each region. Store the resulting CDRs for each country and the world as a vector of length two. The CDR is defined as:

$$\text{CDR} = \frac{\text{number of deaths}}{\text{number of person-years lived}}$$

Briefly describe patterns you observe in the resulting CDRs.

Answer 4

```r
# Function to compute the Crude death rate (CDR)
compute_cdr <- function(population_data) {
  population_data %>%
    group_by(period) %>%
    summarise(cdr = sum(deaths) / sum(py)) %>%
    pull()
}
```

```r
# Compute the CDR for each data set
(world_cdr  <- compute_cdr(world_data))
```

```
[1] 0.019318929 0.008166083
```

```r
(kenya_cdr  <- compute_cdr(kenya_data))
```

```
[1] 0.02396254 0.01038914
```

```r
(sweden_cdr <- compute_cdr(sweden_data))
```

```
[1] 0.009844842 0.009968455
```

The CDR for the world from 1950 to 1955 is greater than its CDR from 2005 to 2010. The CDR for Keyna from 1950 to 1955 is greater than its CDR from 2005 to 2010. The CDR for Sweden from 1950 to 1955 is about the same as its CDR from 2005 to 2010. Kenya has the greatest CDR across itself, Sweden, and the world for both the periods 1950 to 1955 and 2005 to 2010. However, Kenya's CDR for the period 2005 to 2010 is about the same as Sweden's.

Question 5

One puzzling finding from the previous question is that the CDR for Kenya during the period of 2005-2010 is about the same level as that for Sweden. We would expect people in developed countries like Sweden to have a lower death rate than those in developing countries like Kenya. While it is simple and easy to understand, the CDR does not take into account the age composition of a population. We therefore compute the *age specific death rate* (ASDR). The ASDR for age range $[x, x+\delta)$ is defined as:

$$\text{ASDR}_{[x,\ x+\delta)} = \frac{\text{number of deaths for people of age } [x,\ x+\delta)}{\text{number of person-years of people of age } [x,\ x+\delta)}$$

Calculate the ASDR for each age group, separately for Kenya and Sweden, during the period of 2005-2010. Briefly describe the pattern you observe.

Answer 5

```
# Function to compute Age specific death rate (ASDR)
compute_asdr <- function(population_data) {
  population_data %>%
    mutate(asdr = deaths / py)
}
```

```
# Compute ASDR for each data set
world_data  <- compute_asdr(world_data)
kenya_data  <- compute_asdr(kenya_data)
sweden_data <- compute_asdr(sweden_data)
```

During the period of 2005 to 2010, the ASDR is greater in Kenya than in Sweden for each age group.

Question 6

One way to understand the difference in the CDR between Kenya and Sweden is to compute the counterfactual CDR for Kenya using Sweden's population distribution (or vice versa). This can be done by applying the following alternative formula for the CDR.

$$\text{CDR} = \text{ASDR}_{[0,5)} \times P_{[0,5)} + \text{ASDR}_{[5,10)} \times P_{[5,10)} + \cdots$$

where $P_{[x,x+\delta)}$ is the proportion of the population in the age range $[x, x+\delta)$. We compute this as the ratio of person-years in that age range relative to the total person-years across all age ranges. To conduct this counterfactual analysis, we use $\text{ASDR}_{[x,x+\delta)}$ from Kenya and $P_{[x,x+\delta)}$ from Sweden during the period of 2005–2010. That is, first calculate the age-specific population proportions for Sweden and then use them to compute the counterfactual CDR for Kenya. How does this counterfactual CDR compare with the original CDR of Kenya? Briefly interpret the result.

Answer 6

```r
# Function to compute population proportion by period
compute_pop_prop <- function(population_data) {
  population_data %>%
    group_by(period) %>%
    mutate(popP = py / sum(py)) %>%
    ungroup()
}
```

```r
# Compute population proportion for each data set
world_data  <- compute_pop_prop(world_data)
kenya_data  <- compute_pop_prop(kenya_data)
sweden_data <- compute_pop_prop(sweden_data)
```

```r
# Compute Kenyas CDR Kenya had Sweden's population distribution
mutate(kenya_data,
      temp_cdr = asdr * sweden_data$popP) %>%
      group_by(period) %>%
      summarise(cdrresweden = sum(temp_cdr))
```

```
# A tibble: 2 x 2
  period    cdrresweden
  <chr>           <dbl>
1 1950-1955      0.0257
2 2005-2010      0.0232
```

During the period of 1950 to 1955, Kenya's counterfactual CDR is about the same as its original CDR. However, during the period of 2005 to 2010, Kenya's counterfactual CDR is greater than its original CDR; this suggests that, had Kenya had Sweden's population distribution at the time of measurement, Kenya's CDR would be considered more significant when compared with using its own population distribution.