Understanding World Population Dynamics

Assignment 1 - PSYC593

Biraj Neupane

2024-10-02

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.*

Loading required packages for our analysis.

```
# Load packages with message suppressed ----
suppressPackageStartupMessages(library(tidyverse))
# packages for the path
suppressPackageStartupMessages(library(here))
# package for path
suppressPackageStartupMessages(library(rprojroot))
```

Global and local path variables to make sure the working directory and path directory are consistent throughout the project.

```
# directories ----------
root_path <- here::here() # to verify it as a RStudio project
crit_path <- rprojroot::find_root(has_dir("src")) # directory structure
# sub directory to access and create files
code_path <- file.path(root_path, "src") # R and qmd code file path
data_path <- file.path(root_path, "data") # path for data set
figs_path <- file.path(crit_path, "results", "figures") # path for figures
```

Loading Detasets required for our analysis

Data sets for our analysis are located in data folder and raw_data sub folder of our folders. We need to access that and read it uniformly. The original data are in CSV format we are creating a data frame from it.

This code will read our datsets.

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

## Creating Data frame for each country ----

world_data <- data.frame(world_data)
kenya_data <- data.frame(kenya_data)
sweden_data <- data.frame(sweden_data)
```

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. Before you begin this exercise, it would be a good idea to directly inspect each data set. In R, this can be done with the `View` function, which takes as its argument the name of a `data.frame` to be examined. Alternatively, in RStudio, double-clicking a `data.frame` in the `Environment` tab will enable you to view the data in a spreadsheet-like view.

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

Calculation of total person years to compute the CBR from our data set. for this we add men person year and women person year and create a new column py(person year) for each period.

This code will create a new column For the period of 1950-1955 with person year.

```r
# New variable py = total person years for each data set period 1950-1955----
world_data$py_1950_1955 <- ifelse(world_data$period == "1950-1955",
                                  world_data$py.men + world_data$py.women, NA)
kenya_data$py_1950_1955 <- ifelse(kenya_data$period == "1950-1955",
                                kenya_data$py.men + kenya_data$py.women, NA)
sweden_data$py_1950_1955 <- ifelse(sweden_data$period == "1950-1955",
                                kenya_data$py.men + sweden_data$py.women, NA)
```

For the period of 2005-2010.

```r
# New variable py = total person years for each data set period 2005-2010----
world_data$py_2005_2010 <- ifelse(world_data$period == "2005-2010",
                            world_data$py.men + world_data$py.women, NA)
kenya_data$py_2005_2010 <- ifelse(kenya_data$period == "2005-2010",
                            kenya_data$py.men + kenya_data$py.women, NA)
sweden_data$py_2005_2010 <- ifelse(sweden_data$period == "2005-2010",
                             sweden_data$py.men + sweden_data$py.women, NA)
```

for the whole data set.

```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
```

Creating a function to calculate the CBR for our data set, to do this we create a function and use the formula for CBR enlisted up above.

```r
# Function to compute the Crude Birth Rate (CBR) and grouping it by each period
compute_cbr <- function (pop_data) {
  pop_data%>%
    group_by(period) %>%
      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
```

###Results Interpretation

From the results, we can see that the CBR of 0.037 (37) births per 1,000 people in the world during the period of 1950-1955 which decreased to 0.020 (20) per 1,000 people in the period of 2005-2010. Similarly, we observe a CBR of 0.052 (52) births per 1,000 people in Kenya for 1950-1955, and a CBR of 0.038 (38) births per 1,000 people in 2005-2010. In Sweden, the CBR was 0.015 (15) births per 1,000 people in 1950-1955, and it decreased to 0.0119 (12) births per 1,000 people in 2005-2010.

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

Creating a function to calculate Age-Specific Fertility Rate, this code will help us to calculate asfr.

```r
# Function to compute Age specific fertility rate (ASFR) ----
compute_asfr <- function(pop_data) {
  pop_data %>%
    filter(age %in% c("15-19", "20-24", "25-29", "30-34",
                      "35-39", "40-44", "45-49")) %>%
    group_by(period) %>%
    summarise(asfr = sum(births) / sum(py.women), .groups = "drop") %>%
    pull(asfr)
}
```

This code is computing ASFR for each data set and add new column for each dataset with variable name asfr.

```r
# Compute ASFR for each data set -----
(world_data$asfr <- compute_asfr(world_data))
```

```
[1] 0.1517275 0.0777116
```

```r
(kenya_data$asfr <- compute_asfr(kenya_data))
```

```
[1] 0.2345367 0.1583425
```

```r
(sweden_data$asfr <- compute_asfr(sweden_data))
```

```
[1] 0.0628168 0.0530123
```

Creating a dummy variables to total the ASFRs by Period to make it easy for interpretation.

```r
# summing up ASFRS for Kenya and Sweden by period --------
sum_asfrs_kenya <- kenya_data %>%
  group_by(period) %>%
  summarise(total_asfr_kenya = sum(asfr, na.rm = TRUE))

sum_asfrs_sweden <- sweden_data %>%
                    group_by(period) %>%
                      summarise(total_asfr_sweden = sum(asfr, na.rm = TRUE))
```

this will generate a results of ASFRs for kenya and sweden

```r
#Compare ASFRs for Kenya and Sweden --------
sum_asfrs_kenya
```

```
# A tibble: 2 x 2
  period    total_asfr_kenya
  <chr>                <dbl>
1 1950-1955             2.98
2 2005-2010             2.91
```

```r
sum_asfrs_sweden
```

```
# A tibble: 2 x 2
  period    total_asfr_sweden
  <chr>                 <dbl>
1 1950-1955             0.874
2 2005-2010             0.864
```

Result Interpretation and comparison

In the period of 1950-1955, the total ASFR for Kenya was approximately 2.98, which remained almost the same 2.90 in the period of 2005-2010.Similarly in sweden for the period of 1950-1955 ASFR was approximately 0.87 and was also remained almost same 0.86 in the period 2005-2010.

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

creating a function to calculate a TRF

```
# Function to calculate the total fertility rate (TFR) -------

compute_tfr <- function(data) {
  data %>%
    group_by(period) %>%
    summarise(tfr = 5 * sum(asfr, na.rm = TRUE)) %>%
    pull(tfr)
}
```

Generating a TFR for each data set using the function to the data set.

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

```
[1] 8.789007 8.418928
```

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

```
[1] 14.92345 14.54248
```

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

```
[1] 4.368103 4.319080
```

Interpreatation

Total fertility rate (TFR) for the world was approximately 8.78 for the 1950-1955 period and 8.41 for the 2005-2010. For Kenya, the TFR was relatively higher, at 14.92 for the period of 1950-1955 and 14.54 for the 2005-2010. In contrast, Sweden's TFR was notably lower than the world as well as kenya, with values of 4.36 for the earlier period and 4.31 for the 2005-2010 period.

Calculating the total women and birth in the world by period and comparing the change between the period of 2 period of interest.

```r
# Totals 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
# Comparing 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
```

###Result

The total number of women in the world increased significantly, more than doubling from approximately 6.56 million in 1950-1955 to around 16.55 million in 2005-2010. Similarly, the total number of births rose from about 489,000 to approximately 675,000, reflecting a growth of 38% over the same period. Notably, the female population increased by about 152%, while the number of births saw a rise of 37.9% from 1950-1955 to 2005-2010.

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.

Creating a function to compute a crude death rate.

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

This code will compute a CDR for each data set and provide us with the result for two period 1950-1955 and 2005-2010.

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

```
[1] 0.019318929 0.008166083
```

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

```
[1] 0.02396254 0.01038914
```

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

```
[1] 0.009844842 0.009968455
```

Result

Globally, the CDR decreased from approximately 0.0193 in the 1950-1955 period to 0.0082 in the 2005-2010 period, Kenya's CDR declined from about 0.0240 to 0.0104 respectively. Whereas the CBR for Sweden remained relatively stable 0.0098 in the 1950-1955 period and 0.0100 for 2005-2010.

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.

writing a function to compute age specific death rate.

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

This code will compute the ASDR for each data set.

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

```
# A tibble: 30 x 12
# Groups:   period [2]
   country period   age   births deaths py.men py.women py_1950_1955 py_2005_2010
   <chr>   <chr>    <chr>  <dbl>  <dbl>  <dbl>    <dbl>        <dbl>        <dbl>
 1 WORLD   1950-1~ 0-4    0      1.01e5 9.47e5  904909.     1851820.           NA
 2 WORLD   1950-1~ 5-9    0      7.96e3 7.27e5  694575.     1421478.           NA
 3 WORLD   1950-1~ 10-14 0      5.55e3 6.67e5  635492.     1302287.           NA
 4 WORLD   1950-1~ 15-19 5.42e4 5.83e3 6.26e5  600678.     1226869.           NA
 5 WORLD   1950-1~ 20-24 1.32e5 6.65e3 5.74e5  555221.     1128739.           NA
 6 WORLD   1950-1~ 25-29 1.28e5 6.43e3 5.09e5  507277.     1015777.           NA
 7 WORLD   1950-1~ 30-34 8.92e4 6.21e3 4.33e5  437132.      870470.           NA
 8 WORLD   1950-1~ 35-39 5.60e4 6.88e3 4.00e5  405528.      805807.           NA
 9 WORLD   1950-1~ 40-44 2.43e4 8.00e3 3.74e5  382444.      756215.           NA
10 WORLD   1950-1~ 45-49 5.07e3 8.89e3 3.26e5  333889.      660167.           NA
# i 20 more rows
# i 3 more variables: py <dbl>, asfr <dbl>, asdr <dbl>
```

```r
(kenya_data <- compute_asdr(kenya_data))
```

```
# A tibble: 30 x 13
# Groups:   period [2]
   country period    age   births deaths py.men py.women     l_x py_1950_1955
   <chr>   <chr>     <chr>  <dbl>  <dbl>  <dbl>    <dbl>   <dbl>        <dbl>
 1 KEN     1950-1955 0-4      0    398.   2983.    2978. 100000         5960.
 2 KEN     1950-1955 5-9      0     36.8  1978.    1970.  75157.        3948.
 3 KEN     1950-1955 10-14    0     19.5  1635.    1634.  71365.        3269.
 4 KEN     1950-1955 15-19  264.    18.5  1589.    1565.  69469.        3153.
 5 KEN     1950-1955 20-24  486.    21.4  1428.    1364.  67421.        2792.
 6 KEN     1950-1955 25-29  383.    20.4  1205.    1106.  64855.        2311.
 7 KEN     1950-1955 30-34  269.    19.0  1033.     928.  62127.        1961.
 8 KEN     1950-1955 35-39  166.    18.9   913.     803.  59239.        1716.
 9 KEN     1950-1955 40-44   79.6   19.3   817.     711.  56114.        1528.
10 KEN     1950-1955 45-49   25.6   19.9   693.     655.  52678.        1347.
# i 20 more rows
# i 4 more variables: py_2005_2010 <dbl>, py <dbl>, asfr <dbl>, asdr <dbl>
```

```r
(sweden_data <- compute_asdr(sweden_data))
```

```
# A tibble: 30 x 13
# Groups:   period [2]
   country period    age   births deaths py.men py.women     l_x py_1950_1955
   <chr>   <chr>     <chr>  <dbl>  <dbl>  <dbl>    <dbl>   <dbl>        <dbl>
 1 SWE     1950-1955 0-4      0    13.8   1490.    1411. 100000         2901.
 2 SWE     1950-1955 5-9      0     1.30  1543.    1471.  97575.        3014.
 3 SWE     1950-1955 10-14    0     1.22  1266.    1217.  97308.        2483.
 4 SWE     1950-1955 15-19   40.8   1.58  1078.    1049.  97094.        2127.
 5 SWE     1950-1955 20-24 141.     2.26  1119.    1105.  96734.        2225.
 6 SWE     1950-1955 25-29 161.     2.88  1305.    1285.  96224.        2590.
 7 SWE     1950-1955 30-34 117.     3.61  1367.    1338.  95692.        2705.
 8 SWE     1950-1955 35-39   64.8   4.70  1366.    1333.  95042.        2699.
 9 SWE     1950-1955 40-44   21.8   6.81  1367.    1346.  94214.        2713.
10 SWE     1950-1955 45-49    1.70 10.0   1256.    1268.  93038.        2525.
# i 20 more rows
# i 4 more variables: py_2005_2010 <dbl>, py <dbl>, asfr <dbl>, asdr <dbl>
```

constructing the table to compare Sweden and Kenya

```r
# creating a table to compare asdr for the 2005-2010 period
sweden_asdr_2005_2010 <- sweden_data %>%
  filter(period == "2005-2010") %>% select(age, asdr)
```

```
Adding missing grouping variables: `period`
```

```r
kenya_asdr_2005_2010 <- kenya_data %>%
  filter(period == "2005-2010") %>% select(age, asdr)
```

```
Adding missing grouping variables: `period`
```

```r
#joint table for ASDR
join_table_asdr <- left_join(sweden_asdr_2005_2010, kenya_asdr_2005_2010, by = c("age", "peri
  "_sweden", "_kenya"
))

print(join_table_asdr)
```

```
# A tibble: 15 x 4
# Groups:   period [1]
   period    age   asdr_sweden asdr_kenya
   <chr>     <chr>       <dbl>      <dbl>
 1 2005-2010 0-4     0.000679     0.0209
 2 2005-2010 5-9     0.0000814    0.00291
 3 2005-2010 10-14   0.000114     0.00292
 4 2005-2010 15-19   0.000269     0.00294
 5 2005-2010 20-24   0.000470     0.00389
 6 2005-2010 25-29   0.000494     0.00656
 7 2005-2010 30-34   0.000506     0.0106
 8 2005-2010 35-39   0.000669     0.0139
 9 2005-2010 40-44   0.00104      0.0135
10 2005-2010 45-49   0.00177      0.0113
11 2005-2010 50-54   0.00299      0.0112
12 2005-2010 55-59   0.00471      0.0139
13 2005-2010 60-69   0.00983      0.0254
14 2005-2010 70-79   0.0280       0.0613
15 2005-2010 80+     0.110        0.159
```

Result and patterns observed

The age-specific death rates (ASDR) for Kenya are consistently higher than those for Sweden. Both countries show an increasing ASDR with age; however, the rate of increase is steeper in Kenya. For example, in the youngest age group (0-4 years), the ASDR for Kenya is 0.0209 , compared to Sweden's 0.000679. This trend continues into older age groups, as seen in the 70-79 age category, where the ASDR in Kenya is 0.0613, while in Sweden, it is only 0.0280. In the oldest age group (80+ years), Kenya's ASDR rises to 0.159, whereas Sweden's is 0.110.

Question 6

One way to understand the difference in the CDR between Kenya and Sweden is to compute the counter factual 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.

Creating a function to compute population proportion by period .

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

This code will execute the above function to compute the population porportion by period for each dataset.

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

```
# A tibble: 30 x 13
   country period   age   births deaths py.men py.women py_1950_1955 py_2005_2010
   <chr>   <chr>    <chr>  <dbl>  <dbl>  <dbl>    <dbl>        <dbl>        <dbl>
 1 WORLD   1950-1~ 0-4    0      1.01e5 9.47e5  904909.      1851820.           NA
 2 WORLD   1950-1~ 5-9    0      7.96e3 7.27e5  694575.      1421478.           NA
 3 WORLD   1950-1~ 10-14 0      5.55e3 6.67e5  635492.      1302287.           NA
 4 WORLD   1950-1~ 15-19 5.42e4 5.83e3 6.26e5  600678.      1226869.           NA
 5 WORLD   1950-1~ 20-24 1.32e5 6.65e3 5.74e5  555221.      1128739.           NA
 6 WORLD   1950-1~ 25-29 1.28e5 6.43e3 5.09e5  507277.      1015777.           NA
 7 WORLD   1950-1~ 30-34 8.92e4 6.21e3 4.33e5  437132.       870470.           NA
 8 WORLD   1950-1~ 35-39 5.60e4 6.88e3 4.00e5  405528.       805807.           NA
 9 WORLD   1950-1~ 40-44 2.43e4 8.00e3 3.74e5  382444.       756215.           NA
10 WORLD   1950-1~ 45-49 5.07e3 8.89e3 3.26e5  333889.       660167.           NA
# i 20 more rows
# i 4 more variables: py <dbl>, asfr <dbl>, asdr <dbl>, popP <dbl>
```

```
(kenya_data <- compute_pop_prop(kenya_data))
```

```
# A tibble: 30 x 14
   country period    age   births deaths py.men py.women    l_x py_1950_1955
   <chr>   <chr>     <chr>  <dbl>  <dbl>  <dbl>    <dbl>  <dbl>        <dbl>
 1 KEN     1950-1955 0-4        0   398.   2983.    2978. 100000        5960.
```

```
 2 KEN     1950-1955 5-9      0      36.8  1978.    1970.  75157.         3948.
 3 KEN     1950-1955 10-14    0      19.5  1635.    1634.  71365.         3269.
 4 KEN     1950-1955 15-19  264.     18.5  1589.    1565.  69469.         3153.
 5 KEN     1950-1955 20-24  486.     21.4  1428.    1364.  67421.         2792.
 6 KEN     1950-1955 25-29  383.     20.4  1205.    1106.  64855.         2311.
 7 KEN     1950-1955 30-34  269.     19.0  1033.     928.  62127.         1961.
 8 KEN     1950-1955 35-39  166.     18.9   913.     803.  59239.         1716.
 9 KEN     1950-1955 40-44   79.6    19.3   817.     711.  56114.         1528.
10 KEN     1950-1955 45-49   25.6    19.9   693.     655.  52678.         1347.
# i 20 more rows
# i 5 more variables: py_2005_2010 <dbl>, py <dbl>, asfr <dbl>, asdr <dbl>,
#   popP <dbl>
```

```
(sweden_data <- compute_pop_prop(sweden_data))
```

```
# A tibble: 30 x 14
   country period    age   births deaths py.men py.women     l_x py_1950_1955
   <chr>   <chr>     <chr>  <dbl>  <dbl>  <dbl>    <dbl>   <dbl>        <dbl>
 1 SWE     1950-1955 0-4      0    13.8   1490.    1411. 100000         2901.
 2 SWE     1950-1955 5-9      0     1.30  1543.    1471.  97575.        3014.
 3 SWE     1950-1955 10-14    0     1.22  1266.    1217.  97308.        2483.
 4 SWE     1950-1955 15-19   40.8   1.58  1078.    1049.  97094.        2127.
 5 SWE     1950-1955 20-24  141.    2.26  1119.    1105.  96734.        2225.
 6 SWE     1950-1955 25-29  161.    2.88  1305.    1285.  96224.        2590.
 7 SWE     1950-1955 30-34  117.    3.61  1367.    1338.  95692.        2705.
 8 SWE     1950-1955 35-39   64.8   4.70  1366.    1333.  95042.        2699.
 9 SWE     1950-1955 40-44   21.8   6.81  1367.    1346.  94214.        2713.
10 SWE     1950-1955 45-49    1.70 10.0   1256.    1268.  93038.        2525.
# i 20 more rows
# i 5 more variables: py_2005_2010 <dbl>, py <dbl>, asfr <dbl>, asdr <dbl>,
#   popP <dbl>
```

Creating a function to compute kenyas CDR if Kenya had swedens population distribution.

```
# 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
```

Result

The calculated CDR for Kenya, assuming it had Sweden's population distribution, is 0.0257 for the period 1950-1955 and 0.0232 for 2005-2010. In comparison, the actual CDR for Kenya during these periods is 0.02396254 for 1950-1955 and 0.01038914 for 2005-2010. This difference indicates that the actual mortality rates in Kenya are influenced significantly by its age distribution of population. The counterfactual CDRs suggest that if Kenya had a population structure similar to Sweden's, the mortality rates would be relatively higher.