Understanding World Population Dynamics

Nicholas Archambault

12 February 2021

Introduction

Understanding population dynamics is essential to a range of objectives in the social sciences, including policy analysis, resource allocation, and mobilization studies. This project will examine elementary demographic quantities for two time periods: 1950 to 1955 and 2005 to 2010. The data files analyzed include population summaries of Kenya, Sweden, and the world, sourced from the United Nations Department of Economic and Social Affairs.

Variables and their descriptions can be found in the following table.

Variable	Description
country	abbreviated country name
period	period during which data were collected
age	age group
births	number of children born to women of that age group (in thousands)
deaths	number of deaths (in thousands)
py.men	person-years of men (in thousands)
py.women	person-years of women (in thousands)

Data are collected for a period of five years, where a *person-year* is the measure of the time contribution of each person during the period. A person who lives through the entire five-year period contributes five person-years, whereas a person who lives only through the first two years contributes two.

1. Crude Birth Rate

The crude birth rate (CBR) for a particular period is given as

$$CBR = \frac{births}{personyears}$$

.

We begin by computing the crude birth rate for each period, separately for Kenya, Sweden, and the world.

```r
# Load datasets
kenya <- read.csv("Kenya.csv", as.is = TRUE, stringsAsFactors = FALSE)
sweden <- read.csv("Sweden.csv", as.is = TRUE, stringsAsFactors = FALSE)
world <- read.csv("World.csv", as.is = TRUE, stringsAsFactors = FALSE)

# Sum total person-years
kenya$py.total <- kenya$py.men + kenya$py.women
sweden$py.total <- sweden$py.men + sweden$py.women
```

```r
world$py.total <- world$py.men + world$py.women

# Function to find CBR
cbr <- function(dataset) {
  # Sum births and person-years for each period
  births1950 <- sum(dataset$births[dataset$period == "1950-1955"])
  births2005 <- sum(dataset$births[dataset$period == "2005-2010"])
  py_total1950 <- sum(dataset$py.total[dataset$period == "1950-1955"])
  py_total2005 <- sum(dataset$py.total[dataset$period == "2005-2010"])

  # Get CPR for both periods
  cpr1950 <- births1950 / py_total1950
  cpr2005 <- births2005 / py_total2005
  cpr_both <- c(cpr1950, cpr2005)

  # Assign labels
  names(cpr_both) <- c("1950-1955", "2005-2010")
  return(cpr_both)
}

# Get CBR for countries and world
all_cbr <- rbind(cbr(kenya), cbr(sweden), cbr(world))
row.names(all_cbr) <- c("Kenya", "Sweden", "World")
knitr::kable(all_cbr, row.names = TRUE, digits = 3)
```

	1950-1955	2005-2010
Kenya	0.052	0.039
Sweden	0.015	0.012
World	0.037	0.020

Crude birth rates for each region have declined between the 1950's and 2000's. The world exhibits the greatest difference, with a 45.8% decrease across the latter half of the twentieth century. Kenya and Sweden follow, with 26.1% and 22.5% declines, respectively.

We also note that Kenya posts a CBR greater than the world average in both periods, while Sweden's is below average.

2. Age-Specific Fertility Rate

Unlike CBR, total fertility rate (TFR) adjusts for age compositions in the female population of a country. To derive TFR, we'll first calculate the age-specific fertility rate (ASFR), which represents the fertility rate for women within the reproductive age range: 15-50 years.

ASFR for a given age range $[x, x+\delta]$, where x is the starting age and δ the width of the range, is given as

$$ASFR_{[x,x+\delta]} = \frac{births_{[x,x+\delta]}}{personyears_{[x,x+\delta]}}$$

.

In this case, the range width δ is five years.

```r
# Get ASFR for all age groups
kenya$asfr <- kenya$births / kenya$py.women
```

```r
sweden$asfr <- sweden$births / sweden$py.women
world$asfr <- world$births / world$py.women

# Function to obtain ASFR for reproductive age groups within a period
asfr <- function(dataset, period) {
  # Subset to period
  dataset <- dataset[dataset$period == period, ]

  # Reproductive age range labels
  reproductive_window <- c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44",
                           "45-49")

  # Get rows corresponding to reproductive age range
  indicator <- dataset$age %in% reproductive_window

  # Get ASFR for reproductive ages
  asfr_age <- dataset$asfr[indicator]
  names(asfr_age) <- dataset$age[indicator]
  return(asfr_age)
}

# Get ASFR for countries and world
all_asfr <- rbind(asfr(kenya, "1950-1955"), asfr(kenya, "2005-2010"),
                  asfr(sweden, "1950-1955"), asfr(sweden, "2005-2010"),
                  asfr(world, "1950-1955"), asfr(world, "2005-2010"))

row.names(all_asfr) <- c("Kenya (1950-1955)", "Kenya (2005-2010)",
                         "Sweden (1950-1955)", "Sweden (2005-2010)",
                         "World (1950-1955)", "World (2005-2010)")

knitr::kable(all_asfr, format = "pipe", row.names = TRUE, digits = 3)
```

	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Kenya (1950-1955)	0.169	0.356	0.347	0.289	0.206	0.112	0.039
Kenya (2005-2010)	0.101	0.236	0.233	0.181	0.131	0.056	0.038
Sweden (1950-1955)	0.039	0.128	0.125	0.087	0.049	0.016	0.001
Sweden (2005-2010)	0.006	0.051	0.116	0.132	0.063	0.012	0.001
World (1950-1955)	0.090	0.238	0.252	0.204	0.138	0.064	0.015
World (2005-2010)	0.048	0.152	0.147	0.094	0.047	0.016	0.005

ASFR for all age groups has dropped noticeably in Kenya between the 1950's and the 2000's, sometimes by as much as 38%, such as for age group [30-35). The disparities are less pronounced for the upper tail of the age distribution: age group [45-50), for example, barely changed between time periods. This corroborates the previous finding that the overall birth rate is falling, since the declines in births by younger age brackets are not regained by older ones.

Things are different for Sweden, where we immediately notice that ASFR across all age groups is drastically lower than for Kenya, a reflection of Sweden's lower CBR. ASFR for younger Swedish age groups has declined between the 1950's and 2000's, but this is not true across all age groups the way it is in Kenya. We observe that the Swedish ASFR in age group [30-35) has jumped by 51.5%, while in age group [35-40) it has risen by 28.6%.

These differences between countries could be reflective of cultural and societal truths. In wealthier Sweden, where the cost of living is higher and women have better access to contraceptives, ASFR for younger age groups is low and declining across the decades, indicating that young Swedish women are choosing to wait to have children. In the 1950's, Swedish age group [20-25) featured the greatest ASFR, whereas in the 2000's that distinction belonged to age group [30-35).

In Kenya, where the standard of living is lower, medical access is not as prevalent, and more children are often needed to supplement family income, we observe that all women — and young women in particular — have much higher ASFR values. Though overall birth rates have declined across the board, Kenyan women are still choosing to have children mostly around the same (younger) ages.

3. Total Fertility Rate

Having calculated age-specific fertility rate, we can now find total fertility rate, the average number of children to which women will give birth if they live through their entire reproductive window.

$$TFR = ASFR_{[15,20)]} \times 5 + ASFR_{[20,25)]} \times 5 + ... + ASFR_{[45,50)]} \times 5$$

.

Each age-specific fertility rate is multiplied by five, which is the duration in years of each age range.

```r
# Function to calculate TFR
tfr <- function(dataset) {
  # Find ASFR by period
  asfr1950 <- asfr(dataset, "1950-1955")
  asfr2005 <- asfr(dataset, "2005-2010")

  # Find TFR according to formula
  tfr1950 <- sum(5 * asfr1950)
  tfr2005 <- sum(5 * asfr2005)

  tfr_both <- c(tfr1950, tfr2005)
  names(tfr_both) <- c("1950-1955", "2005-2010")

  return(tfr_both)
}

# Get TFR for countries and world
all_tfr <- rbind(tfr(kenya), tfr(sweden), tfr(world))
row.names(all_tfr) <- c("Kenya", "Sweden", "World")
knitr::kable(all_tfr, format = "pipe", row.names = TRUE, digits = 3)
```

	1950-1955	2005-2010
Kenya	7.591	4.880
Sweden	2.227	1.903
World	5.007	2.544

The average number of children birthed per woman across the world has been cut in half between the 1950's and 2000's, from 5 children per woman to roughly 2.5. Kenya's fertility rate is high compared to the rest of the world in both 1950-1955 and 2005-2010, while Sweden's rate is lower than the world average in both periods.

4. Crude Death Rate

We'll next examine crude death rate (CDR), a quantity analogous to crude birth rate that measures the ratio of deaths to person-years.

```r
# Function to calculate CDR
cdr <- function(dataset) {
  cdr_both <- c(sum(dataset$deaths[dataset$period == "1950-1955"]) /
                  sum(dataset$py.total[dataset$period == "1950-1955"]),
                sum(dataset$deaths[dataset$period == "2005-2010"]) /
                  sum(dataset$py.total[dataset$period == "2005-2010"]))
  names(cdr_both) <- c("1950-1955", "2005-2010")
  return(cdr_both)
}

# Get CDR for countries and world
all_cdr <- rbind(cdr(kenya), cdr(sweden), cdr(world))
row.names(all_cdr) <- c("Kenya", "Sweden", "World")
knitr::kable(all_cdr, format = "pipe", row.names = TRUE, digits = 3)
```

	1950-1955	2005-2010
Kenya	0.024	0.010
Sweden	0.010	0.010
World	0.019	0.008

Crude death rates have fallen in Kenya and across the world between the 1950's and 2000's by 56.6% and 57.7%, respectively, while remaining static in Sweden. This makes sense, as we know that advancements in medicine, sanitation, agriculture, and environmental regulations have contributed to a higher standard of living in the modern age.

5. Age-Specific Death Rate

Notably, despite disparities in the wealth of the two nations, CDR is roughly equal in Sweden and Kenya in the modern age. We would expect the Swedish value to be lower than that of a developing country like Kenya.

CDR, however, does not take into account the age composition of a population. In order to obtain a more accurate picture of deaths, we can calculate the age-specific death rate for an age range $[x, x+\delta]$. This value is given by

$$ASDR_{[x,x+\delta]} = \frac{deaths_{[x,x+\delta]}}{personyears_{[x,x+\delta]}}$$

.

```r
# Calculate ASDR by country
kenya$asdr <- kenya$deaths / kenya$py.total
sweden$asdr <- sweden$deaths / sweden$py.total
world$asdr <- world$deaths / world$py.total

# Function to obtain ASDR by period
asdr <- function(dataset, period){
  dataset <- dataset[dataset$period == period, ]
  asdr_vector <- dataset$asdr
```

```r
  names(asdr_vector) <- dataset$age

  return(asdr_vector)
}

# Get ASDR for countries and world
all_asdr <- cbind(asdr(kenya, "1950-1955"), asdr(kenya, "2005-2010"),
                  asdr(sweden, "1950-1955"), asdr(sweden, "2005-2010"),
                  asdr(world, "1950-1955"), asdr(world, "2005-2010"))

colnames(all_asdr) <- c("Kenya (1950-1955)", "Kenya (2005-2010)",
                         "Sweden (1950-1955)", "Sweden (2005-2010)",
                         "World (1950-1955)", "World (2005-2010)")

knitr::kable(all_asdr, format = "pipe", row.names = TRUE, digits = 3)
```

	Kenya (1950-1955)	Kenya (2005-2010)	Sweden (1950-1955)	Sweden (2005-2010)	World (1950-1955)	World (2005-2010)
0-4	0.067	0.021	0.005	0.001	0.055	0.013
5-9	0.009	0.003	0.000	0.000	0.006	0.001
10-14	0.006	0.003	0.000	0.000	0.004	0.001
15-19	0.006	0.003	0.001	0.000	0.005	0.001
20-24	0.008	0.004	0.001	0.000	0.006	0.002
25-29	0.009	0.007	0.001	0.000	0.006	0.002
30-34	0.010	0.011	0.001	0.001	0.007	0.003
35-39	0.011	0.014	0.002	0.001	0.009	0.003
40-44	0.013	0.013	0.003	0.001	0.011	0.004
45-49	0.015	0.011	0.004	0.002	0.013	0.005
50-54	0.018	0.011	0.006	0.003	0.017	0.007
55-59	0.024	0.014	0.010	0.005	0.024	0.010
60-69	0.042	0.025	0.021	0.010	0.042	0.020
70-79	0.094	0.061	0.060	0.028	0.087	0.047
80+	0.200	0.159	0.168	0.110	0.184	0.121

Despite relatively even CDR values for Kenya and Sweden in the period 2005-2010, we see that ASDR for the two nations tells a different story.

In the table below, the `ratio` column represents, for each age group, the ASDR of Kenya divided by that of Sweden. The resulting value is Kenya's ASDR for that age group as a multiple of the corresponding value for Sweden.

```r
# Calculate ratio between Kenya and Sweden to understand disparity more easily
mm <- as.data.frame(matrix(NA, 15, 2))
mm[, 1] <- names(asdr(sweden, "2005-2010"))
mm[, 2] <- unname(asdr(kenya, "2005-2010")) / unname(asdr(sweden, "2005-2010"))
colnames(mm) <- c("age", "ratio")
mean(mm$ratio)
```

```
## [1] 13.24919
```

```r
mm
```

```
##      age     ratio
## 1    0-4 30.807897
## 2    5-9 35.773751
## 3  10-14 25.705904
## 4  15-19 10.949524
## 5  20-24  8.271415
## 6  25-29 13.271700
## 7  30-34 20.968508
## 8  35-39 20.750281
## 9  40-44 12.965617
## 10 45-49  6.378798
## 11 50-54  3.731483
## 12 55-59  2.950868
## 13 60-69  2.583795
## 14 70-79  2.184820
## 15   80+  1.443458
```

Kenya's ASDR for certain age groups is up to 35 times greater than Sweden's. For no age group is Kenya's ASDR less than Sweden's, despite similar CDR's. Though the ratios between the two countries' ASDR values gradually diminish toward the upper end of the age spectrum, Kenya's ASDR values are typically double or many times greater than the ASDR values of the same age groups in Sweden. Averaged across all age groups, the ratio between Kenya's and Sweden's ASDR values is 13.24.

In both countries, the ASDR values are highest for three oldest age groups, though Kenya also exhibits a large ASDR for children in the [0-5) age bracket.

For no age range is the ASDR of Kenya less than that of Sweden. The table confirms our intuition that simply examining CDR does not capture the full story of death rates within the two countries. Death rates are substantially greater for Kenyan age groups, especially younger ones, indicating that the two nations' CDR values are similar due to the high proportion of older adults living in Sweden.

6. Counterfactual CDR for Kenya Using Sweden's Population

Another method for understanding the difference in CDR between the two countries is to compute the counterfactual: to find the CDR for Kenya using the population of Sweden, or vice versa. The formula to obtain the counterfactual CDR value is given by

$$CDR = ASDR_{[0,5)]} \times P_{[0,5)]} + ASDR_{[5,10)]} \times P_{[5,10)]} + ...$$

,

where $P_{[x,x+\delta)}$ is the proportion of the population within a given age range $[x, x+\delta)$. We can find the ratios of person-years within each age range relative to the total person-years across all age ranges in order to compute this counterfactual case.

```r
# Populations proportions for Kenya, 2005-2010
kenya_2005_prop <- kenya$py.total[kenya$period == "2005-2010"]
kenya_2005_prop <- kenya_2005_prop / sum(kenya_2005_prop)
names(kenya_2005_prop) <- kenya$age[kenya$period == "2005-2010"]

# Populations proportions for Sweden, 2005-2010
sweden_2005_prop <- sweden$py.total[sweden$period == "2005-2010"]
sweden_2005_prop <- sweden_2005_prop / sum(sweden_2005_prop)
names(sweden_2005_prop) <- sweden$age[sweden$period == "2005-2010"]

# Find Kenya ASDR, 2005-2010
kenya_2005_asdr <- kenya$asdr[kenya$period == "2005-2010"]

# Find Kenya CDR, 2005-2010
sum(kenya_2005_asdr * kenya_2005_prop)
```

```
## [1] 0.01038914
```

```r
# Counterfactual CDR for Kenya, 2005-2010
sum(kenya_2005_asdr * sweden_2005_prop)
```

```
## [1] 0.02321646
```

```r
# Compare Kenya and Sweden across age distributions
counterfactual <- cbind(kenya_2005_prop, sweden_2005_prop)
colnames(counterfactual) <- c("Kenya", "Sweden")
knitr::kable(counterfactual, format = "pipe", row.names = TRUE, digits = 3)
```

	Kenya	Sweden
0-4	0.167	0.058
5-9	0.140	0.051
10-14	0.120	0.061
15-19	0.112	0.068
20-24	0.103	0.061
25-29	0.085	0.060
30-34	0.066	0.065
35-39	0.050	0.070
40-44	0.039	0.070
45-49	0.032	0.065
50-54	0.026	0.063
55-59	0.020	0.067
60-69	0.023	0.115
70-79	0.013	0.072
80+	0.004	0.054

The counterfactual CDR for Kenya using Sweden's population proportions is over double the true Kenyan CDR. This indicates that if Kenya had the population proportions within each age group that Sweden has, Kenya's CDR would be much higher than both its own and Sweden's true CDR values. It appears Kenya's current population proportions are actually advantageous in keeping its CDR relatively low. It also confirms our assumptions about higher death rates in poorer, developing nations. Given similar population circumstances to Sweden, some factor in Kenyan society — whether healthcare, government, nutrition, climate, or some combination — prevents Kenya from achieving lower, Swedish-level death rates when controlling for differences in population proportions.

Conclusion

In this project, we examined a number of fundamental demographic quantities pertaining to births and deaths that provide a better sense of population dynamics in Kenya, Sweden, and across the world. Probing these parameters can yield a deeper understanding of the true differences in demographic disparities between wealthier and poorer nations, as well as the factors potentially responsible for them.