

Exercice 1: Comprendre la dynamique des populations

Miranda Melson

07 June 2021

Cet exercice est tiré du livre de Imai Kosuke.

Comprendre la dynamique des populations est important pour de nombreux domaines des sciences sociales. Nous allons calculer certaines quantités démographiques à base des naissances et des décès pour deux pays et la population mondiale pour deux périodes: 1950 à 1955 et 2005 à 2010. Nous analyserons les fichiers de données CSV suivants: `Kenya.csv`, `Sweden.csv` et `World.csv`. Chaque fichier contient des données démographiques pour le Kenya, la Suède (Sweden) et le monde (World), respectivement. Le tableau ci-dessous présente les noms et les descriptions des variables dans chaque ensemble de données.

Nom	Description
<code>country</code>	Nom du pays abrégé
<code>period</code>	Période au cours de laquelle les données sont collectées
<code>age</code>	Groupe d'âge
<code>births</code>	Nombre de naissance en milliers (i.e., nombre d'enfants nés des femmes du groupe d'âge)
<code>deaths</code>	Nombre de décès en millier
<code>py.men</code>	Personne-années pour les hommes en milliers
<code>py.women</code>	Personne-années pour les femmes en milliers

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

Les données sont collectées pour une période de 5 ans, où *personne-années* est une mesure de la contribution temporelle de chaque personne au cours de la période. Par exemple, une personne qui vit pendant toute la période de cinq ans contribue pour cinq personnes-années alors que celle qui ne vit que pendant la première moitié de la période contribue à 2,5 personnes-années. Avant de commencer cet exercice, il serait judicieux d'inspecter directement chaque ensemble de données. Dans R, cela peut être fait avec la fonction `View`, qui prend comme argument le nom d'un fichier (`data.frame`) à examiner. Alternativement, dans RStudio, double-cliquer sur le nom du fichier que vous venez d'ouvrir dans l'onglet **Environnement** vous permettra de visualiser les données dans une vue de type tableur. Vous pouvez directement visualiser les données dans RMarkdown avec la commande `head`.

Remarque 1: Prenez toujours soins d'enregistrer les **données à utiliser dans le même fichier** que votre RMarkdown pour vous simplifier la vie.

Remarque 2: Tout au long de l'exercice, pensez à écrire de petite fonctions pour simplifier votre tâche.

Mise en route

Ce code vous permet de charger directement les données. Vous devez télécharger les données de Moodle et les placer avec ce fichier Rmarkdown dans le même dossier (voir remarque en haut).

```
## Telecharger les données
setwd("C:/Users/mme1s/OneDrive/Desktop/sicss_montreal/Cours_SICSS_Montreal/Semaine1_Formation_Intensive

Sweden <- read.csv("Sweden.csv")
Kenya <- read.csv("Kenya.csv")
World <- read.csv("World.csv")
```

```
head(Sweden)
```

```
##   country   period   age  births deaths   py.men py.women      l_x
## 1    SWE 1950-1955  0-4   0.000 13.765 1490.037 1410.502 100000.00
## 2    SWE 1950-1955  5-9   0.000  1.302 1542.698 1470.816  97575.29
## 3    SWE 1950-1955 10-14  0.000  1.216 1266.321 1217.133  97308.11
## 4    SWE 1950-1955 15-19 40.823  1.581 1078.133 1049.193  97093.74
## 5    SWE 1950-1955 20-24 141.137  2.264 1119.421 1105.129  96733.76
## 6    SWE 1950-1955 25-29 160.882  2.885 1305.003 1284.552  96223.56
```

```
head(Kenya)
```

```
##   country   period   age  births deaths   py.men py.women      l_x
## 1    KEN 1950-1955  0-4   0.000 398.314 2982.589 2977.828 100000.00
## 2    KEN 1950-1955  5-9   0.000  36.799 1977.935 1969.698  75157.42
## 3    KEN 1950-1955 10-14  0.000  19.520 1634.561 1633.975  71364.50
## 4    KEN 1950-1955 15-19 264.185  18.508 1588.554 1564.652  69469.47
## 5    KEN 1950-1955 20-24 485.564  21.361 1427.824 1364.061  67421.15
## 6    KEN 1950-1955 25-29 383.252  20.424 1204.917 1105.817  64854.80
```

```
head(World)
```

```
##   country   period   age  births deaths   py.men py.women
## 1  WORLD 1950-1955  0-4     0.00 101090.393 946910.6 904909.3
## 2  WORLD 1950-1955  5-9     0.00  7960.861 726903.3 694574.5
## 3  WORLD 1950-1955 10-14    0.00  5550.176 666794.7 635492.3
## 4  WORLD 1950-1955 15-19 54238.34  5831.195 626191.1 600677.9
## 5  WORLD 1950-1955 20-24 131939.23  6649.425 573518.2 555221.0
## 6  WORLD 1950-1955 25-29 128063.23  6425.219 508500.4 507277.0
```

QUESTION 1

Nous commençons par calculer *le taux brut de natalité* (TBN) pour une période donnée. Le TBN est défini comme suit:

$$\text{TBN} = \frac{\text{Nombre total de naissances}}{\text{Nombre total de personnes-années vécues}}$$

Calculez le TBN pour chaque période, séparément pour le Kenya, la Suède et le Monde. Pour faire ceci, il vous faut créer peut-être plusieurs objets: 1. NTNKP1 signifie : nombre total de naissance du Kenya pour la période 1. 2. NTAVHKP1 signifie Nombre total des années vécues par les hommes au Kenya pour la période 1 et ainsi de suite. Je vous donne ces résultats dans la commande.

Réponse 1

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
kenya_p1 <- Kenya %>% filter(period == "1950-1955")  
kenya_p2 <- Kenya %>% filter(period == "2005-2010")  
  
sweden_p1 <- Sweden %>% filter(period == "1950-1955")  
sweden_p2 <- Sweden %>% filter(period == "2005-2010")  
  
world_p1 <- World %>% filter(period == "1950-1955")  
world_p2 <- World %>% filter(period == "2005-2010")
```

```
tbn <- function(x){  
  b <- x[,4]  
  p <- x[,6] + x[,7]  
  b/p  
}
```

```
tbn_kenya_p1 <- tbn(kenya_p1)  
tbn_kenya_p2 <- tbn(kenya_p2)
```

```
tbn_sweden_p1 <- tbn(sweden_p1)  
tbn_sweden_p2 <- tbn(sweden_p2)
```

```
tbn_world_p1 <- tbn(world_p1)  
tbn_world_p2 <- tbn(world_p2)
```

```
compare_tbn <- rbind(tbn_kenya_p1, tbn_kenya_p2, tbn_sweden_p1, tbn_sweden_p2, tbn_world_p1, tbn_world_p2)  
colnames(compare_tbn) <- c("0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85-89", "90-94", "95-99")
```

```
compare_tbn
```

```
##           0-4 5-9 10-14      15-19      20-24      25-29      30-34  
## tbn_kenya_p1    0  0    0 0.083782982 0.17391977 0.16585726 0.13698443  
## tbn_kenya_p2    0  0    0 0.050131247 0.11782266 0.11628648 0.08945652  
## tbn_sweden_p1   0  0    0 0.019189819 0.06344519 0.06212728 0.04321264  
## tbn_sweden_p2   0  0    0 0.002907449 0.02477011 0.05684919 0.06499995  
## tbn_world_p1    0  0    0 0.044208743 0.11689080 0.12607411 0.10252691
```

```
## tbn_world_p2      0      0      0 0.023585119 0.07442527 0.07231265 0.04625872
##                35-39      40-44      45-49 50-54 55-59 60-69 70-79 80+
## tbn_kenya_p1  0.09655516 0.052091529 0.0189787273      0      0      0      0      0
## tbn_kenya_p2  0.06468294 0.028345065 0.0197793629      0      0      0      0      0
## tbn_sweden_p1 0.02400816 0.008043547 0.0006741510      0      0      0      0      0
## tbn_sweden_p2 0.03072767 0.005957032 0.0003022686      0      0      0      0      0
## tbn_world_p1  0.06950242 0.032168891 0.0076828791      0      0      0      0      0
## tbn_world_p2  0.02302596 0.008045888 0.0022410755      0      0      0      0      0
```

Votre interprétation

I notice that the highest crude birth rate (tbn : le taux brut de natalité) for Kenya period 1 is for 20-24 year-olds. This is the same age group for Kenya period 2, but a lower rate than in period 1. Kenya has the highest TBN of the countries. Sweden has a lower rate compared to the world and Kenya. In general, most of the peaks are seen at ages 20-24 but during the second period there are also higher rates among other age groups like 30-34 year-olds for sweden period 2 or 25-29 year-olds for the world period 1. They follow the normal “bell-curve” trend.

QUESTION 2

Prenez un moment pour réfléchir à la manière d’organiser les résultats précédents.

Stockez les résultats du TBN sous forme de matrice de dimension (2, 3), 2 étant le nombre de période et 3 le nombre de pays/région. Décrivez brièvement les tendances observées dans les TBN calculés.

Réponse 2

```
tbn_matrice <- matrix(c(sum(tbn_kenya_p1), sum(tbn_kenya_p2), sum(tbn_sweden_p1), sum(tbn_sweden_p2), s
rownames(tbn_matrice) <- c("1950-1955", "2005-2010")
colnames(tbn_matrice) <- c("Kenya", "Sweden", "World")
tbn_matrice
```

```
##                Kenya    Sweden    World
## 1950-1955 0.7281699 0.2207008 0.4990547
## 2005-2010 0.4865043 0.1865137 0.2498947
```

Votre interprétation

This is much more clear than earlier that period one had a higher crude birth rate than period 2 for all countries. Kenya has a higher crude birth rate compared to the world and Sweden, while Sweden has a lower crude birth rate than the world (and Kenya).

QUESTION 3

Le TBN est facile à comprendre mais pose des problèmes de comparaisons. Pourquoi le taux brut de natalité n’est pas un bon indicateur de comparaison du niveau de fécondité entre pays?

Réponse 3

NB: J'ai remarqué que l'exercice 1 en anglais n'a pas cette question. Je regarde les deux et je ne comprends pas pourquoi ils sont différents.

This way of displaying the TBN does not explain it by age group or by gender.

QUESTION 4

Une solution intermédiaire au TBN est le calcul du *taux général de fécondité (TGF)*. Celui-ci se calcule par:

$$\text{TGF} = \frac{\text{Nombre total de naissances}}{\text{Nombre total de personnes-années vécues par les femmes de 15-49 ans}}$$

Calculer ce taux pour le Kenya, la Suède et le Monde. Présenter les résultats aussi sous forme de matrice. Pourquoi cela vous paraît déjà meilleur que le TBN? Interpréter les résultats. Quel est le problème du taux général de fécondité.

Réponse 4

```
kenya_p1_tgf <- kenya_p1 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")
kenya_p2_tgf <- kenya_p2 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")

sweden_p1_tgf <- sweden_p1 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")
sweden_p2_tgf <- sweden_p2 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")

world_p1_tgf <- world_p1 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")
world_p2_tgf <- world_p2 %>% filter(age == "15-19"|age == "20-24"|age == "25-29"|age == "30-34"|age == "35-39"|age == "40-44")

tgf <- function(y){
  b <- y[,4]
  f <- y[,7]
  b/f
}

tgf_kenya_p1 <- tgf(kenya_p1_tgf)
tgf_kenya_p2 <- tgf(kenya_p2_tgf)

tgf_sweden_p1 <- tgf(sweden_p1_tgf)
tgf_sweden_p2 <- tgf(sweden_p2_tgf)

tgf_world_p1 <- tgf(world_p1_tgf)
tgf_world_p2 <- tgf(world_p2_tgf)

tgf_matrice <- matrix(c(sum(tgf_kenya_p1), sum(tgf_kenya_p2), sum(tgf_sweden_p1), sum(tgf_sweden_p2), sum(tgf_world_p1), sum(tgf_world_p2)),
  rownames(tgf_matrice) <- c("1950-1955", "2005-2010")
  colnames(tgf_matrice) <- c("Kenya", "Sweden", "World")
  tgf_matrice

##              Kenya    Sweden    World
## 1950-1955 1.5182820 0.4453834 1.0014496
## 2005-2010 0.9759137 0.3805528 0.5087247
```

Votre interprétation

While TGF results are similar to TBN results, they are less skewed and both Kenya and Sweden are more similar to the world now.

QUESTION 5

Pour remédier à ces problèmes, nous calculons l'*indice synthétique de fécondité* (ISF). Contrairement au TBN, l'ISF s'ajuste en fonction de la composition par âge de la population féminine en âge de procréer. Pour ce faire, nous devons d'abord calculer le *taux de fécondité par âge* (isf), qui représente le taux de fécondité chez les femmes en âge de procréer [15, 50]. Le isf pour la tranche d'âge $[x, x + \delta)$, où x est l'âge de départ et δ la largeur de la tranche d'âge (mesurée en années), est défini comme suit:

$$\text{isf}_{[x, x+\delta)} = \frac{\text{Nombre de naissances des femmes âgées de } [x, x + \delta)}{\text{Nombre de personnes-années vécues par les femmes âgées } [x, x + \delta)}$$

Notez que les crochets, $[,]$, incluent la limite alors que les parenthèses, $(,)$, l'excluent. Par exemple, $[20, 25)$ représente la tranche d'âge supérieure ou égale à 20 ans et inférieure à 25 ans. Dans les données démographiques typiques, la tranche d'âge δ est fixée à 5 ans.

Calculez le taux de fécondité par âge pour la Suède et le Kenya ainsi que pour le monde entier pour chacune des deux périodes. Stockez les isf résultants séparément pour chaque région. Interpréter ces résultats. Comment devrez-vous présenter ces résultats pour une interprétation optimale?

Réponse 5

```
isf <- function(a){
  x <- min(a[,4])
  d <- 5
  tfa <- a[x, x+d]
  birth <- a[,4]
  women.py <- a[,6]
  birth[tfa]/women.py[tfa]
}

isf_kenya_p1 <- isf(kenya_p1_tgf)
isf_kenya_p2 <- isf(kenya_p2_tgf)

isf_sweden_p1 <- isf(sweden_p1_tgf)
isf_sweden_p2 <- isf(sweden_p2_tgf)

isf_world_p1 <- isf(world_p1_tgf)
isf_world_p2 <- isf(world_p2_tgf)

isf_matrice <- matrix(c(sum(isf_kenya_p1), sum(isf_kenya_p2), sum(isf_sweden_p1), sum(isf_sweden_p2), sum(isf_world_p1), sum(isf_world_p2)),
  rownames(isf_matrice) <- c("1950-1955", "2005-2010")
  colnames(isf_matrice) <- c("Kenya", "Sweden", "World")
  isf_matrice
```

```
##           Kenya Sweden World
## 1950-1955      0      NA      0
## 2005-2010      0       0      0
```

Votre interprétation

I attempted to create this function but was unable to do so unfortunately.

QUESTION 6

En utilisant le isf , nous pouvons définir l'ISF comme le nombre moyen d'enfants que les femmes mettent au monde si elles vivent tout au long de leur vie reproductive.

$$ISF = isf_{[15, 20)} \times 5 + isf_{[20, 25)} \times 5 + \dots + isf_{[45, 50)} \times 5$$

Nous multiplions chaque taux de fécondité par âge par 5 car la tranche d'âge est de 5 ans. Calculez l'ISF pour la Suède et le Kenya ainsi que le monde entier pour chacune des deux périodes. Comme dans la question précédente, continuer à supposer que la tranche d'âge en âge de procréer des femmes est de [15, 50). Stockez les deux ISF résultants pour chaque pays et le monde en tant que matrice de dimension (2, 3). En général, comment est-ce que l'indice synthétique de fécondité a-t-il évolué au cours de la période de 1950 à 2000 dans le monde? Commentez les résultats de la Suède et du Kenya.

Réponse 6

```
isft <- function(z){
  sum(z[15,50] * 5)
}

isft_kenya_p1 <- isft(kenya_p1_tgf)
isft_kenya_p2 <- isft(kenya_p2_tgf)

isft_sweden_p1 <- isft(sweden_p1_tgf)
isft_sweden_p2 <- isft(sweden_p2_tgf)

isft_world_p1 <- isft(world_p1_tgf)
isft_world_p2 <- isft(world_p2_tgf)

isft_matrice <- matrix(c(sum(isft_kenya_p1), sum(isft_kenya_p2), sum(isft_sweden_p1), sum(isft_sweden_p2)),
  rownames(isft_matrice) <- c("1950-1955", "2005-2010")
  colnames(isft_matrice) <- c("Kenya", "Sweden", "World")
  isft_matrice
```

```
##           Kenya Sweden World
## 1950-1955      0      0      0
## 2005-2010      0      0      0
```

Votre interprétation

Same issue as the last problem.

ANNEXES

1. Glossaire

Fécondité: C'est l'étude des naissances

Fécondité = Fertility

Fécondité \neq Fertilité

Taux brut de natalité = Crude birth rate (CBR)

Taux global de fécondité générale

Indice synthétique de fécondité (ISF) = Total fertility rate (TFR)

Taux de fécondité par âge = Age specific fertility rate (ASFR)