Lab2

% DS 201 (Lab 2)

% Sidak Yntiso sgy210@nyu.edu, Office Hours: CDS 663 floor, 10-12 pm

% February 10, 2020

LATEX

1. Math in Rmarkdown

- In-line math is placed between dollar-sign brackets: $f_{ij}x_i \times y_j^2$
- Greek symbols, binary relations and other symbols: $\alpha, \beta, \gamma, \chi, \pi, \neq, \geq, \leq, \sim, \implies, \in, \infty$.
- Displayed math can be placed within slash + squared brackets e.g.

$$E[\bar{x}] = E[\frac{1}{n}\sum_{i=1}^{n} x_i]$$

$$= \frac{1}{N}\sum_{i=1}^{n} E[x_i]$$

$$= \frac{N\bar{x}}{N}$$

$$= \bar{x}$$

2. Installation

- Editor: to write input code (e.g. TeXstudio) + TeX: to transform the code into document (e.g. MacTex)
- Online: Overleaf
- Guides: Learn LATEX in 30 minutes

Exercise 1

This question is reproduced from Question 1.5.2 of Imai's QSS text. You will need to load the relevant packages

```
#Packages to install
#install.packages("tidyverse")
#install.packages("devtools")


#Load packages and datasets
# library(tidyverse)
# library(devtools)
# install_github("kosukeimai/qss-package",
#                 build_vignettes = TRUE)
```

The question asks us to compute the age-specific crude date rate:

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

where $x, x + delta$ represent an age-range e.g. [20,25) years.

The datasets (called Kenya and Sweden) are structured as follows:

Variable	Description
country	abbreviated country name
period	period during which data are collected
age	age group
births	number of births (in thousands), i.e., the 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)

Part A

For each dataset, add a column to each dataset that corresponds to total population (total population$_{x,x+\delta}$ = py.men$_{x,x+\delta}$ + py.women$_{x,x+\delta}$). Create another column that corresponds to the age-specific population share (pop_prop$_{x,x+\delta}$ = total population$_{x,x+\delta} \sum(total_population_{x,x+\delta})$). When computing fractions, notice that the datasets contains two periods for each country - 1950-1955 and 2005-2010.

Part B

Write a function that computes the ASDR for each age group in Kenya and Sweden.

Part C

Use the ASDR function to compute the crude death rate for each country: CDR = $\sum_x ASDR[x, x + \delta) \times P[x, x + \delta)$, where $P[x, x + \delta)$ is the proportion of the population in the age range $x - x + \delta$.

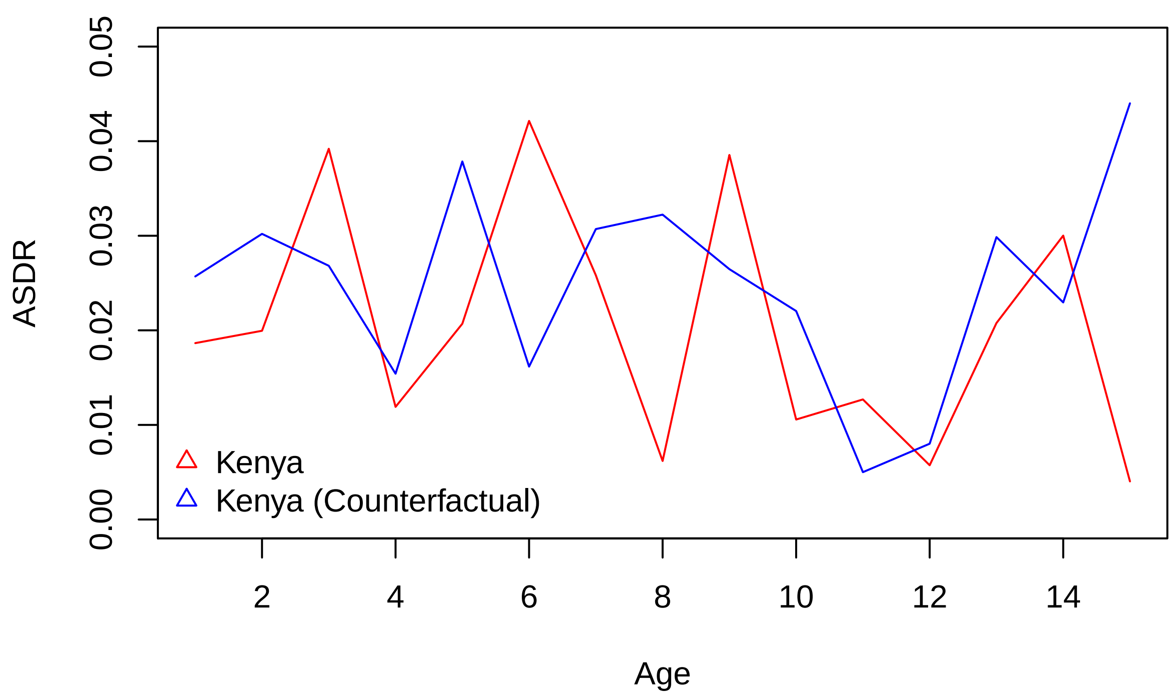
Part D

Imai notes the small difference in CDR between Kenya and Sweden, despite the 35 fold difference in GDP per capita. This is likely driven by differences in the underlying age distributions. To see how, conduct a counterfactual analysis where you use $ASDR[x, x + \delta)$ from Kenya and $P[x, x + \delta)$ from Sweden. Compute counterfactual CDR = $\sum_x ASDR_{Kenya}[x, x + \delta) \times P_{Sweden}[x, x + \delta)$ How does this counterfactual CDR compare with the original CDR of Kenya? Interpret the difference, if any.

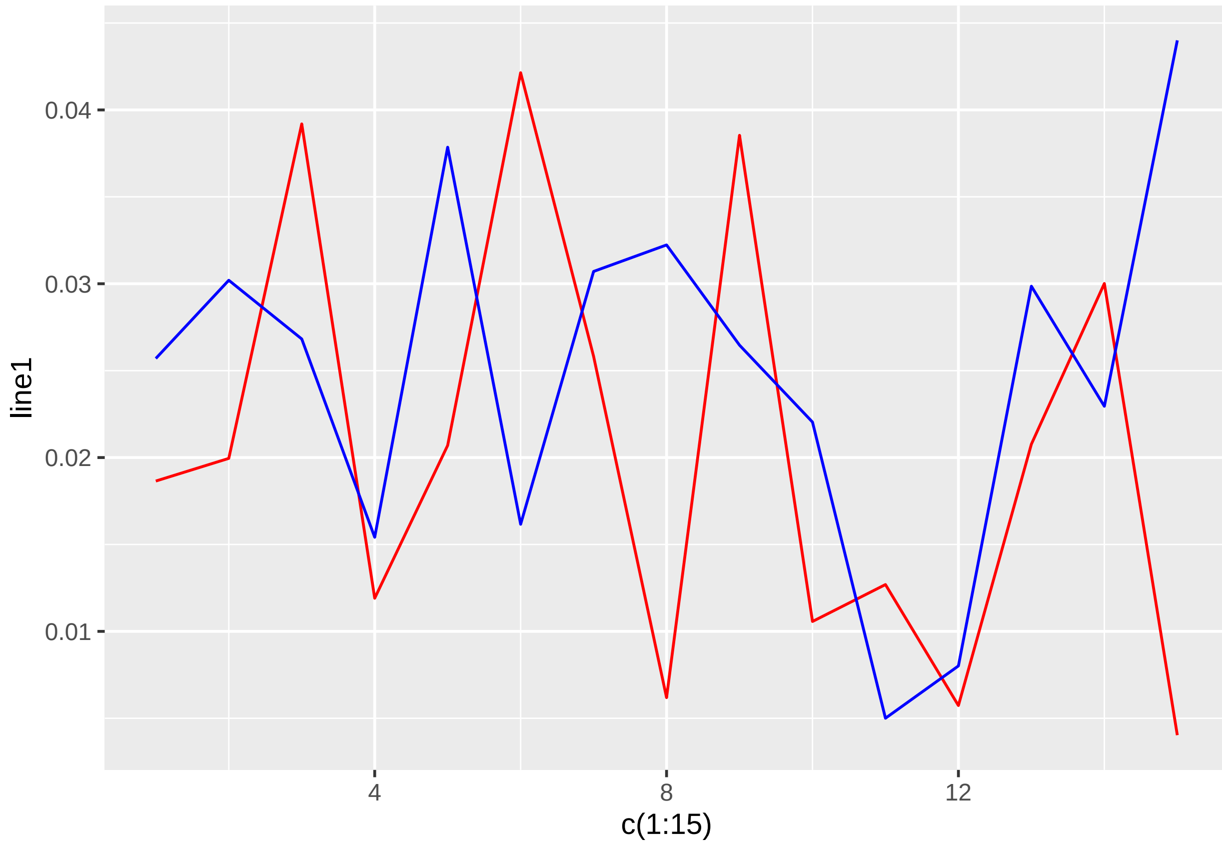
Sidenote: Figures in R

```
#plot has arguments (x,y), color, plot type ("p" = points,"l" = lines, "h" = "histogram", etc)
#overlay additional plots with ``lines''
line1 <- runif(15,min=0,max=0.05)
```

```r
line2 <- runif(15,min=0,max=0.05)

#adding line for line1 (in red), label figure
plot(x = c(1:15),y = line1,
     col="red",ylim=c(0,0.05),type="l",
     xlab="Age",ylab = "ASDR")
#adding line for line2 (in  blue)
lines(x = c(1:15),line2,col="blue")
#adding a legend to the bottom left
legend("bottomleft",
       legend = c("Kenya", "Kenya (Counterfactual)"),
       col= c("red","blue"),
       pch = c(2,2),
       bty = "n")

##you will probably want to use ggplot2
#install.packages("ggplot2")
dat <- data.frame(x = c(1:15),line1 = line1,line2=line2)

library(ggplot2)
```




```r
#plotting lines
ggplot(data=dat)+ #takes dataset
  geom_line(aes(x = c(1:15),y = line1),col="red")+ #plots lines
  geom_line(aes(x = c(1:15),y = line2),col="blue")
```




```
#include normal confidence intervals
#ggplot(data=dat)+
#  geom_smooth(aes(x = c(1:15),y = line1),col="red",method = "loess")+
#  geom_smooth(aes(x = c(1:15),y = line2),col="blue",method = "loess")
```

SOLUTIONS

```
#get working directory
getwd()
```

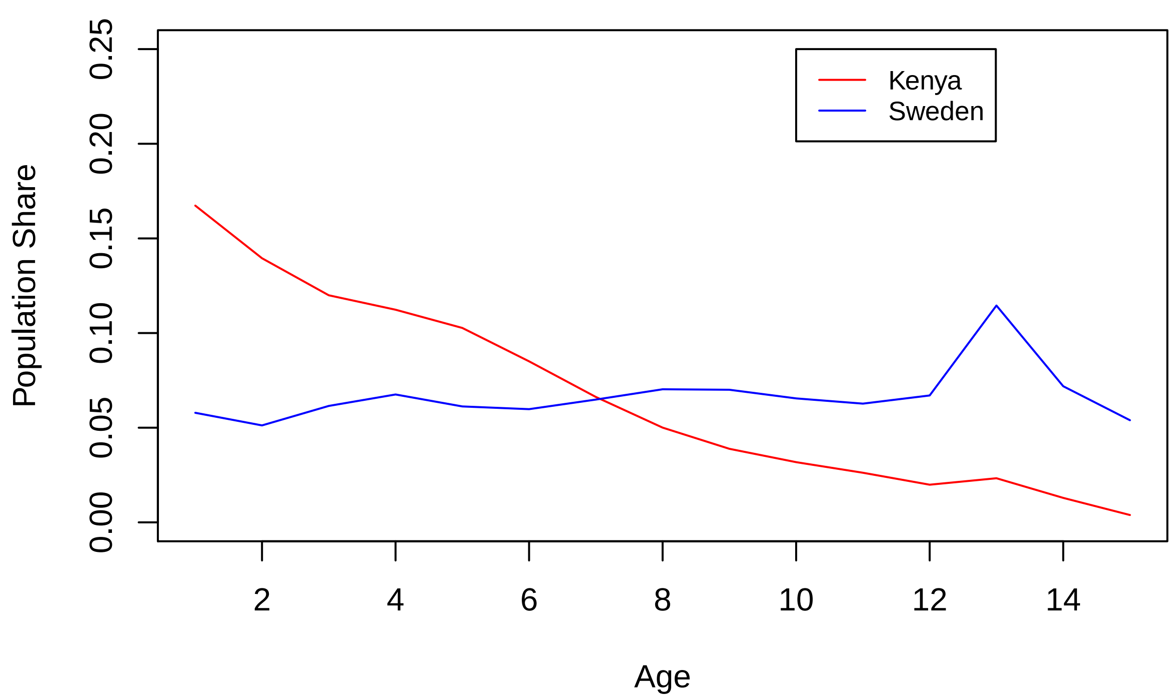
```
## [1] "C:/Users/Sidak Yntiso/Dropbox/CI/Week 2/Lab"
```

```
#Load packages and datasets
Kenya =read.csv("C:\\Users\\Sidak Yntiso\\Downloads\\Kenya.csv")
Sweden =read.csv("C:\\Users\\Sidak Yntiso\\Downloads\\Sweden.csv")
Kenya <- subset(Kenya, period=="2005-2010")
Sweden <- subset(Sweden, period=="2005-2010")

#compute total population
Kenya$total_pop <- Kenya$py.men+Kenya$py.women
Sweden$total_pop <- Sweden$py.men+Sweden$py.women

#compute proportions in each age group
Kenya$prop_pop <- Kenya$total_pop/sum(Kenya$total_pop)
Sweden$prop_pop <- Sweden$total_pop/sum(Sweden$total_pop)
```

```r
ASDR <- function(dat){
  dr <- as.numeric(dat$deaths)/as.numeric(dat$total_pop)
  return(dr)
}
ASDR(Kenya)
```

```
##  [1] 0.020920755 0.002911301 0.002918895 0.002942986 0.003885368 0.006558131
##  [7] 0.010603913 0.013881062 0.013474598 0.011288057 0.011152339 0.013898334
## [13] 0.025395531 0.061261551 0.158620510
```

```r
CDR_Kenya <- sum(ASDR(Kenya) * Kenya$prop_pop)
CDR_Sweden <- sum(ASDR(Sweden) * Sweden$prop_pop)

CDR_Kenya_counterfactual <-  sum(ASDR(Kenya) * Sweden$prop_pop)

#Interpretation: the ASDR is increasing in age and there are a higher share of older people in Sweden.
#If Kenya's population had the same age distribution as Sweden's, Kenya's CDR would be 2.3X

#population proportions
plot(c(1:15),as.numeric(Kenya$prop_pop),col="red",ylim=c(0,0.25),type="l",
     ylab="Population Share",xlab="Age")
lines(c(1:15),as.numeric(Sweden$prop_pop),col="blue",type="l")
legend(10,0.25, legend=c("Kenya", "Sweden"),
       col=c("red", "blue"),lty=1:1, cex=0.8)
```



```r
#ASDR
plot(c(1:15),as.numeric(ASDR(dat=Kenya)),col="red",ylim=c(0,0.05),type="l",
     ylab="ASDR",xlab="Age")
lines(c(1:15),as.numeric(ASDR(dat=Sweden)),col="blue")
legend(2,0.05, legend=c("Kenya", "Sweden"),
       col=c("red", "blue"),lty=1:1,cex=0.8)
```