# Assignment 1b- Zamima Islam Sabaa

1/17/22

**STEP 1: Setting my working directory and loading required packages in R**

```
rm(list = ls())         #clears environment
directory <- setwd("~/Documents/SFU courses/Spring 2022/Econ836 - A.
Econometrics/Assignm")              #sets working directory
library(haven)          #needed to load the .dta stata data file
library(tidyverse)      #needed for datacleaning

library(dplyr)
library(labelled)       #needed for labels
library(modelsummary)   #needed for tables
library(data.table)
```

**STEP 2: Reading the dataset**

```
d.s <- read_dta("~/Documents/SFU courses/Spring 2022/Econ836 - A.
Econometrics/Assignm/Census_2016_Hierarchial.dta") #loads the dataset and
names it "d.s"
```

**STEP 3: Cleaning the dataset**

Here I will clean three variables that are needed in my analysis which have missing data:

1. agegrp: this is the variable that defines the "agegroup" and has data that is not available is labelled as "88", so here I will replace those observations with NA
2. empin: this is the variable that defines the "income" and has data that is not available and not applicable is labelled as "88888888,99999999", so here I will replace those observations with NA
3. sex: this is the variable that defines observations "sex" and has data that is not available is labelled as "8", so here I will replace those observations with NA

```
#filtering out "Agegroup"
class(d.s$agegrp) #checking the class of the variable

[1] "haven_labelled" "vctrs_vctr"     "double"

d.s <- d.s %>%
  mutate(agegrp = case_when(
    agegrp == 88 ~ as.double(NA),
                          TRUE ~ as.double(agegrp)
```

```
  ))

#filtering out "Income"
class(d.s$empin) #checking the class of the variable
```

```
[1] "numeric"
```

```
d.s <- d.s %>%
  rename(income= 'empin') %>%
  mutate(income = case_when(
    income %in% c(88888888,99999999) ~ as.numeric(NA),
    TRUE ~ income
  ))

#filtering out "sex"
class(d.s$sex) #checking the class of the variable
```

```
[1] "haven_labelled" "vctrs_vctr"     "double"
```

```
d.s <- d.s %>%
  mutate(sex = case_when(
    sex == 8 ~ as.double(NA),
    TRUE ~ as.double(sex)
  ))
```

Now I will check how many observations I have for each person:

```
people <-
  d.s %>%
  group_by(PP_ID) %>%
  summarise(obs= n())

people %>% filter(obs != 1)
```

```
# A tibble: 0 × 2
# … with 2 variables: PP_ID <dbl>, obs <int>
```

It shows that there is one observation for each person, so I don't have to remove any rows of observations

**STEP 4: Constructing a dummy variable for homosexuals:**

The "cfstat" variable shows whether a person lives with a non-relative then it is coded as 7, so I create a dummy variable as follows:

```
d.s <-
  d.s %>%
  mutate(non_relative = ifelse(cfstat == 7, 1, 0)) #if cfstat==7, code it as
1 or 0 otherwise


# Now, the grouping at the household level, filtering out the people whose
```

```
age we're not interested in and those who were not in full-time work

h.h <-
  d.s %>%
  filter(agegrp %in% c(9,10,11) & wrkact==11) %>%
  group_by(HH_ID) %>%
  summarise(non_relative = sum(non_relative),
            people = n(),                 #the no. of people there are in the
HH_ID
            sexes = n_distinct(sex))


#focuses only on the agegroups '45 - 64' & those who were based on full-
time,full-year work
#the summarise gives me new variables
#counts whether the people living in the household have the same sex, if same
sex we get 1, if different sex we get 2
#the arrange function gives me the "desc" in descending order
```

Now I will determine if a household potentially has homosexual people in it (individuals unrelated through family) by marking households which have a greater number of people in the house compared to the number of genders. If there is, for instance, a house with 3 people in it where at least one of them is unrelated in family terms and only two genders, there is a potential homosexual couple living in the household. As I have no more than this, I have no choice to later "mark" all of the people living in the household as homosexual.

Now I make the "hom_sxl" indicator at the household level, and later left join this data to the individual level dataframe to mark each individual as presumed homosexuals.

```
# Creating the indicator

h.h <- h.h %>%
  mutate(hom_sxl = ifelse(non_relative >= 1 & people > sexes,
'Homosexual','Heterosexual')) %>%
  select(HH_ID, hom_sxl)

# Make it at the individual level

d.join <-
  d.s %>%
  filter(agegrp %in% c(9,10,11)) %>%
  left_join(h.h, by = 'HH_ID') %>%   #
  mutate(hmsxl = as.factor(hom_sxl))
```

**STEP 5: Creating the data.frame that will give me the final means table**

```
table<-
  d.join %>%
```

```
  mutate(agegrp = case_when(
         agegrp %in% c(9,10) ~ '45-54',
         agegrp == 11 ~ '55-64'),
         sex = ifelse(sex == 1, 'Female', 'Male'),
         hmsxl = as.factor(hmsxl),
         MarStH = as_factor(MarStH)) %>%
  rename(Income = 'income',
         Sexual_orientation = 'hom_sxl',
         Mar_status = 'MarStH') %>%
  select(Sexual_orientation, agegrp, Income, sex, Mar_status)
```

**STEP 6: Output of the table**

```
#using the datasummary to create final Table and saving it as an excel file

m_table<-
datasummary((Income * Sexual_orientation* Mar_status) ~ (Mean*sex*agegrp),
            data = table,
            output='data.frame')

write.csv(m_table, 'table.csv')
```

Finally I use excel to create the ratios and the resulting table gives us the final values and ratios

| Sexual_orientation | Mar_status | Male / 45-54 | Male / 55-64 | Female / 45-54 | Female / 55-64 |
|---|---|---|---|---|---|
| **Heterosexual** | Never legally married (and not living common law) | $62,436.44 | $56,581.53 | $61,611.80 | $61,146.74 |
| | Legally married (and not separated) | $90,990.62 | $81,711.76 | $55,695.37 | $50,313.53 |
| | Living common law | $72,884.28 | $69,739.80 | $52,267.84 | $47,482.99 |
| | Separated, divorced or widowed (and not living common law) | $78,049.26 | $74,523.11 | $59,301.69 | $53,780.94 |
| **Homosexual** | Never legally married (and not living common law) | $50,875.00 | $44,521.74 | $67,356.12 | $60,000.00 |
| | Legally married (and not separated) | $60,600.00 | $21,666.67 | $61,142.86 | $31,000.00 |
| | Living common law | $45,666.67 | $29,000.00 | $31,000.00 | |
| | Separated, divorced or widowed (and not living common law) | $48,615.38 | $38,625.00 | $57,923.08 | $43,947.37 |
| Ratios of Heterosexuals/Homosexuals | Never legally married (and not living common law) | 122.73% | 127.09% | 91.47% | 101.91% |
| | Legally married (and not separated) | 150.15% | 377.13% | 91.09% | 162.30% |
| | Living common law | 159.60% | 240.48% | 168.61% | |
| | Separated, divorced or widowed (and not living common law) | 160.54% | 192.94% | 102.38% | 122.38% |