#Required packages

library(dplyr)

library(haven)

library(xtable)

library(plm)

library(pglm)

library(broom)

library(knitr)

library(tidyverse)

library(tidyr)

library(haven)

library(stargazer)

library(kableExtra)

library(knitr)

library(expss)

library(forcats)

library(devtools)

library(questionr)

```

```{r}

#Getting datasets

setwd("C:/Users/natim/Desktop/Thesis/Thesis2/Data/DHS")

women <- read\_dta("women.dta")

men <- read\_dta("men.dta")

household <- read\_dta("household.dta")

```

```{r}

#Household variables: relationship to head, age of the household member, school attendance status

household1 <- subset(household, select = c(hhid,

hv002,

hv101\_01,

hv101\_02,

hv101\_03,

hv101\_04,

hv101\_05,

hv101\_06,

hv101\_07,

hv101\_08,

hv101\_09,

hv101\_10,

hv101\_11,

hv101\_12,

hv101\_13,

hv101\_14,

hv101\_15,

hv101\_16,

hv101\_17,

hv101\_18,

hv101\_19,

hv101\_20,

hv101\_21,

hv105\_01,

hv105\_02,

hv105\_03,

hv105\_04,

hv105\_05,

hv105\_06,

hv105\_07,

hv105\_08,

hv105\_09,

hv105\_10,

hv105\_11,

hv105\_12,

hv105\_13,

hv105\_14,

hv105\_15,

hv105\_16,

hv105\_17,

hv105\_18,

hv105\_19,

hv105\_20,

hv105\_21,

hv129\_01,

hv129\_02,

hv129\_03,

hv129\_04,

hv129\_05,

hv129\_06,

hv129\_07,

hv129\_08,

hv129\_09,

hv129\_10,

hv129\_11,

hv129\_12,

hv129\_13,

hv129\_14,

hv129\_15,

hv129\_16,

hv129\_17,

hv129\_18,

hv129\_19,

hv129\_20,

hv129\_21,

hv106\_01,

hv106\_02,

hv106\_03,

hv106\_04,

hv106\_05,

hv106\_06,

hv106\_07,

hv106\_08,

hv106\_09,

hv106\_10,

hv106\_11,

hv106\_12,

hv106\_13,

hv106\_14,

hv106\_15,

hv106\_16,

hv106\_17,

hv106\_18,

hv106\_19,

hv106\_20,

hv106\_21,

hv104\_01,

hv104\_02,

hv104\_03,

hv104\_04,

hv104\_05,

hv104\_06,

hv104\_07,

hv104\_08,

hv104\_09,

hv104\_10,

hv104\_11,

hv104\_12,

hv104\_13,

hv104\_14,

hv104\_15,

hv104\_16,

hv104\_17,

hv104\_18,

hv104\_19,

hv104\_20,

hv104\_21))

#Getting age in long format

data01 <- gather(household1, Age1, age, hv105\_01:hv105\_21)

subset1 <- subset(data01, select = c(hhid,

age))

#Getting school status in long format

data02 <- gather(household1,School, school\_status, hv129\_01:hv129\_21)

subset2 <- subset(data02, select =c(school\_status))

#Getting relationship to head in long format

data03 <- gather(household1, Head, relationship\_to\_head, hv101\_01:hv101\_21)

subset3 <- subset(data03, select = c(relationship\_to\_head))

#Getting level of education in long format

data04 <- gather(household1, Level, level\_educ,hv106\_01:hv106\_21 )

subset4 <- subset(data04, select = c(level\_educ))

#Joining long format datasets

data\_long <- cbind(subset1, subset3, subset2, subset4)

#Subsetting the long format dataset to respondents age 5 to 24

subset00 <- subset(data\_long, age >= 18 & age <= 24)

crosstab <- prop.table(table(subset00$age, subset00$level\_educ),2)\*100

table(household$hv106\_01)

freq\_school <- subset00 %>%

count(level\_educ) %>%

mutate("%" = n / sum(n) \*100) %>%

mutate\_at(3, round, 2)

```

```{r}

#Getting the variables of interest from women´s dataset

women1 <- subset(women, select = c(caseid,

v002,

v012,

v131,

v101,

v106,

v107,

v133,

v149,

v218,

s106a,

s106b,

s109,

s110,

s111,

s112,

s113,

s115,

s116))

subset1 <- subset(women1, v012 >= 12 & v012 <= 16)

summary(women$s109)

table(subset1$v149)

levels(subset1$s109)

freq\_students <- subset1 %>%

count(s109) %>%

mutate("%" = n / sum(n) \*100) %>%

mutate\_at(3, round, 2)

age\_school <- prop.table(table(subset1$v012, subset1$s109),2)\*100

```

```{r}

subset2 <- subset(subset1, s109 == 0)

#Crosstab ethnicity and school status

ethni\_school <- prop.table(table(subset2$v131, subset2$s109),2)\*100