Project

Joshua Owusu Ansah

11/3/2020

**Research Question: Is there an association between gender and the living condition of persons in a country**

Explanatory Variable: Gender(categorical).

Response Variable:Living condition of persons(categorical). ## 1. Load data set(s) and libraries

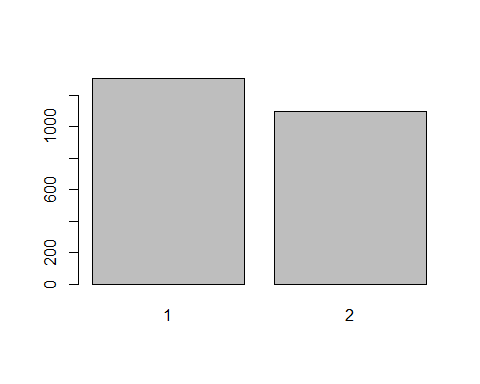
library(readr)  
load("gha\_r7\_2018.RData")  
library(descr)  
library(stats)

## 2. Create variable subset

vars=c("URBRUR","Q4B","Q101")  
my\_data=gha\_r7\_data[vars]

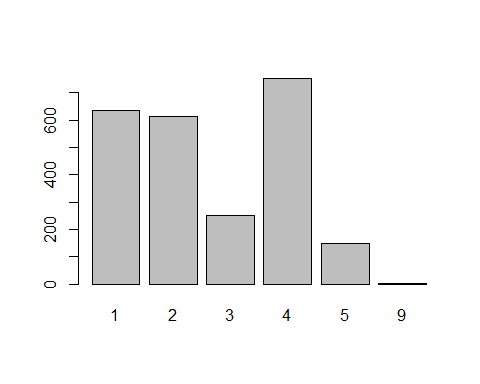
## 3. Data management I: check for and recode errors and NAs

freq(my\_data$URBRUR)



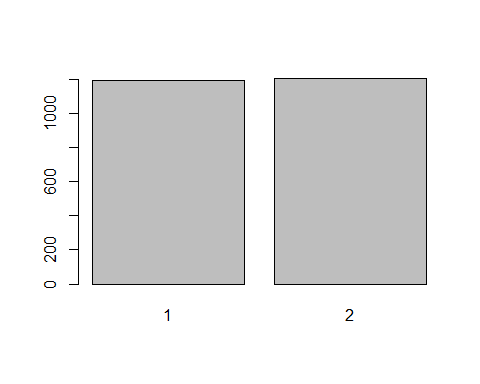
## Urban or Rural Primary Sampling Unit   
## Frequency Percent  
## 1 1304 54.33  
## 2 1096 45.67  
## Total 2400 100.00

freq(my\_data$Q4B)



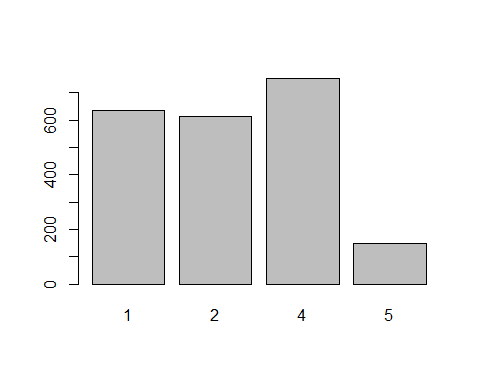
## Q4b. Your present living conditions   
## Frequency Percent  
## 1 635 26.458  
## 2 613 25.542  
## 3 251 10.458  
## 4 751 31.292  
## 5 147 6.125  
## 9 3 0.125  
## Total 2400 100.000

freq(my\_data$Q101)



## Q101. Gender of respondent   
## Frequency Percent  
## 1 1195 49.79  
## 2 1205 50.21  
## Total 2400 100.00

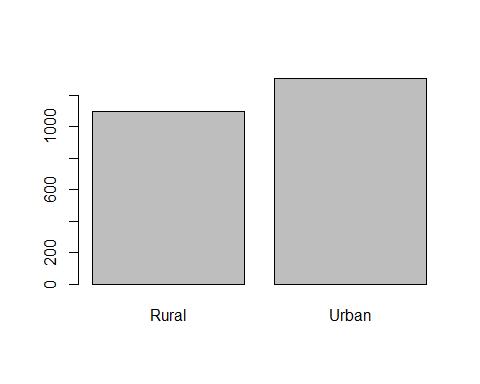
my\_data$Q4B[my\_data$Q4B==9 | my\_data$Q4B==-1 | my\_data$Q4B==8 | my\_data$Q4B==3]<-NA  
freq(my\_data$Q4B)



## Q4b. Your present living conditions   
## Frequency Percent Valid Percent  
## 1 635 26.458 29.59  
## 2 613 25.542 28.56  
## 4 751 31.292 35.00  
## 5 147 6.125 6.85  
## NA's 254 10.583   
## Total 2400 100.000 100.00

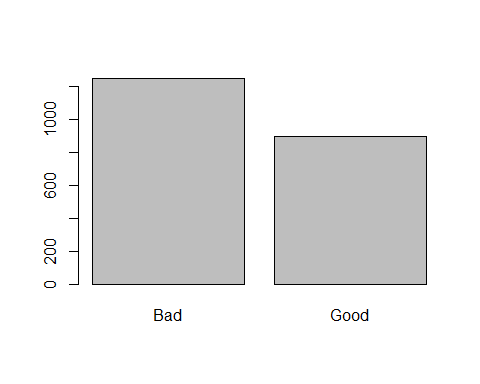
## 4. Data management II: further subset and create secondary variable

my\_data$URBRUR[my\_data$URBRUR==1]<- "Urban"  
my\_data$URBRUR[my\_data$URBRUR==2]<- "Rural"  
freq(my\_data$URBRUR)



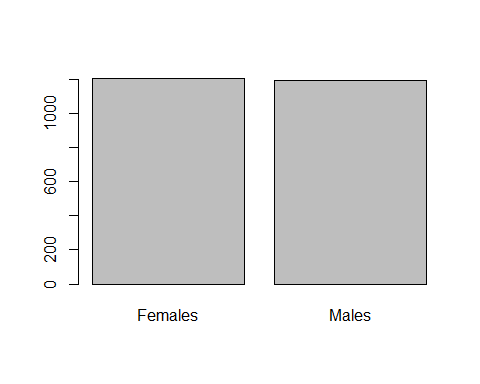
## Urban or Rural Primary Sampling Unit   
## Frequency Percent  
## Rural 1096 45.67  
## Urban 1304 54.33  
## Total 2400 100.00

my\_data$livingCondition<-rep(NA,nrow(my\_data))  
my\_data$livingCondition[my\_data$Q4B==1 | my\_data$Q4B==2]<-"Bad"  
my\_data$livingCondition[my\_data$Q4B==4 | my\_data$Q4B==5]<-"Good"  
  
freq(my\_data$livingCondition)



## my\_data$livingCondition   
## Frequency Percent Valid Percent  
## Bad 1248 52.00 58.15  
## Good 898 37.42 41.85  
## NA's 254 10.58   
## Total 2400 100.00 100.00

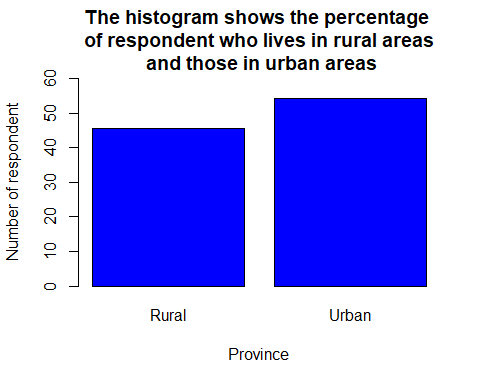
my\_data$Q101[my\_data$Q101==1]<-"Males"  
my\_data$Q101[my\_data$Q101==2]<-"Females"  
freq(my\_data$Q101)



## Q101. Gender of respondent   
## Frequency Percent  
## Females 1205 50.21  
## Males 1195 49.79  
## Total 2400 100.00

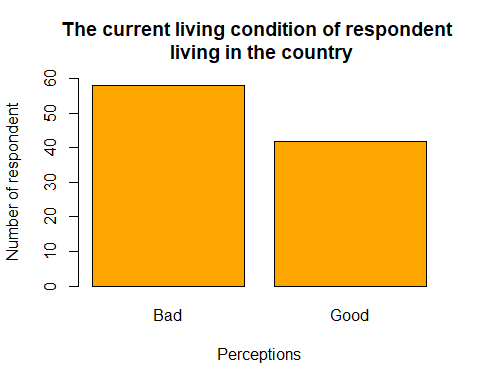
## 5. Descriptive statistics (sample means, standard deviations, proportions) and univariate displays

freq(my\_data$URBRUR,main="The histogram shows the percentage \n of respondent who lives in rural areas \n and those in urban areas",names=c("Rural","Urban"),ylab="Number of respondent",xlab="Province",col="blue",ylim=c(0,60),y.axis="percent")



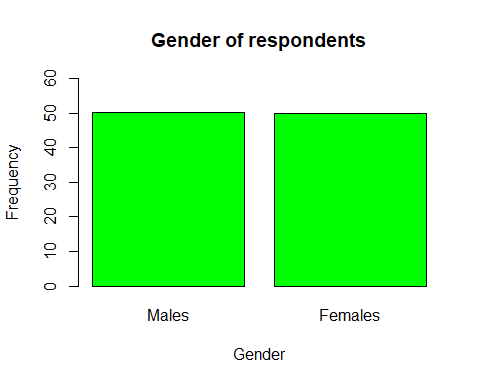
## Urban or Rural Primary Sampling Unit   
## Frequency Percent  
## Rural 1096 45.67  
## Urban 1304 54.33  
## Total 2400 100.00

freq(my\_data$livingCondition,main="The current living condition of respondent \n living in the country",names=c("Bad","Good"),ylab="Number of respondent",xlab="Perceptions",col="orange",ylim=c(0,60),y.axis="percent")



## my\_data$livingCondition   
## Frequency Percent Valid Percent  
## Bad 1248 52.00 58.15  
## Good 898 37.42 41.85  
## NA's 254 10.58   
## Total 2400 100.00 100.00

freq(my\_data$Q101,main="Gender of respondents",names=c("Males","Females"),ylim=c(0,60),ylab="Frequency",xlab="Gender",col="green",y.axis="percent")



## Q101. Gender of respondent   
## Frequency Percent  
## Females 1205 50.21  
## Males 1195 49.79  
## Total 2400 100.00

With regards to the living condition of the respondents, 52.07% of them said their living conditions is bad, 37.46% said theirs is good and 10.47% also said theirs is neutral/neither good nor bad. Notwithstanding among the respondent, 50.21% are females and 49.79% are males. This imply that, approximately there is an equal representation of men and women. 45.67% of the respondent live in the rural area and 54.33% are in the urban areas.

## 6. Bivariate tables and graphs

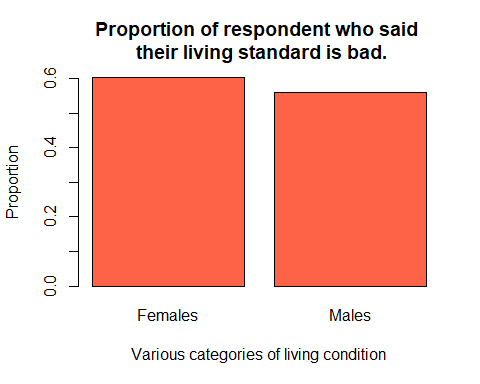
tab1=table(my\_data$livingCondition,my\_data$Q101);tab1

##   
## Females Males  
## Bad 647 601  
## Good 424 474

tab1\_colProp=prop.table(tab1,2)  
round(tab1\_colProp,4)

##   
## Females Males  
## Bad 0.6041 0.5591  
## Good 0.3959 0.4409

barplot(tab1\_colProp[1, ],ylim = c(0,0.6),main = "Proportion of respondent who said \n their living standard is bad.",ylab = "Proportion",xlab = "Various categories of living condition",col = "tomato")

 Among those who said their living standard is good, 35% are females and 40% are males.

## 7. Moderation by gender

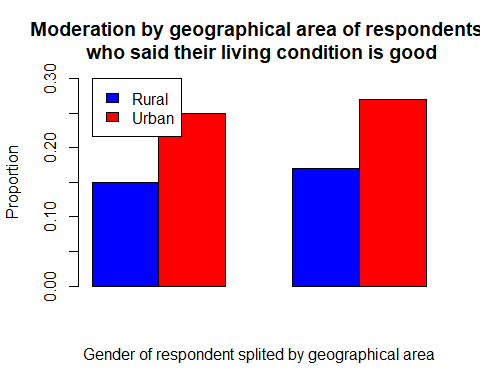
tbl=ftable(my\_data$Q101~my\_data$livingCondition+my\_data$URBRUR);tbl

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 336 325  
## Urban 311 276  
## Good Rural 156 188  
## Urban 268 286

proportion\_tab=round(prop.table(tbl,2),2);proportion\_tab

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 0.31 0.30  
## Urban 0.29 0.26  
## Good Rural 0.15 0.17  
## Urban 0.25 0.27

barplot(proportion\_tab[3:4, ],beside=TRUE,ylim = c(0,0.3),col = c("blue","red"),main = "Moderation by geographical area of respondents \n who said their living condition is good",xlab = "Gender of respondent splited by geographical area",ylab = "Proportion")  
legend(x=1,y=0.3,legend = c("Rural","Urban"),fill = c("blue","red"))

 Among those who said their living condition is good, for those in Rural areas, 13% are females and 16% are males. That notwithstanding, for those in Urban areas, 22% are females and 24% are males. This shows that in both rural areas and urban centers men turn to have a better living condition than women.

### Project Assign 4

## 6. Bivariate tables and graphs

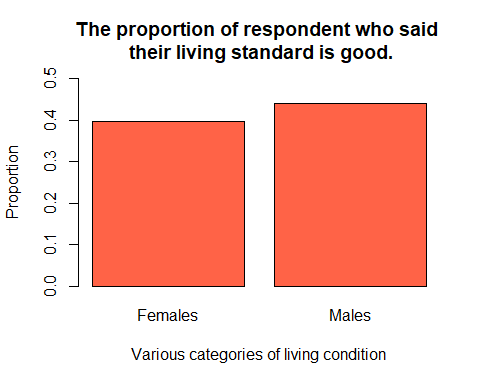
tab1=table(my\_data$livingCondition,my\_data$Q101)   
tab1\_colProp=prop.table(tab1,2)  
round(tab1\_colProp,4)

##   
## Females Males  
## Bad 0.6041 0.5591  
## Good 0.3959 0.4409

round(tab1,4)

##   
## Females Males  
## Bad 647 601  
## Good 424 474

barplot(tab1\_colProp[2, ],ylim = c(0,0.5),main = "The proportion of respondent who said \n their living standard is good.",ylab = "Proportion",xlab = "Various categories of living condition",col = "tomato")



### CHI SQUARE TEST

chisq=chisq.test(my\_data$livingCondition,my\_data$Q101)  
chisq

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: my\_data$livingCondition and my\_data$Q101  
## X-squared = 4.2889, df = 1, p-value = 0.03836

chisq$expected

## my\_data$Q101  
## my\_data$livingCondition Females Males  
## Bad 622.8369 625.1631  
## Good 448.1631 449.8369

chisq$observed

## my\_data$Q101  
## my\_data$livingCondition Females Males  
## Bad 647 601  
## Good 424 474

chisq$residuals

## my\_data$Q101  
## my\_data$livingCondition Females Males  
## Bad 0.9682007 -0.9663977  
## Good -1.1413911 1.1392657

### Section 8

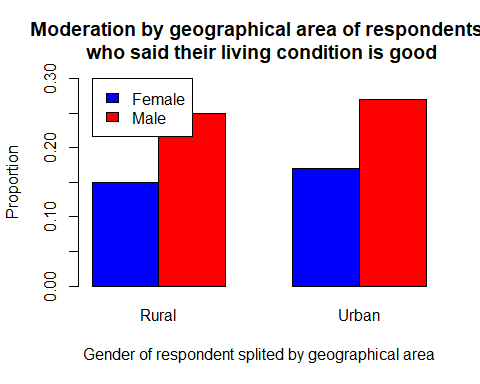
# For those who said their living condition is good.   
tbl=ftable(my\_data$Q101~my\_data$livingCondition+my\_data$URBRUR);tbl

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 336 325  
## Urban 311 276  
## Good Rural 156 188  
## Urban 268 286

proportion\_tab=round(prop.table(tbl,2),2);proportion\_tab

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 0.31 0.30  
## Urban 0.29 0.26  
## Good Rural 0.15 0.17  
## Urban 0.25 0.27

barplot(proportion\_tab[3:4, ],beside=TRUE,ylim = c(0,0.3),col = c("blue","red"),main = "Moderation by geographical area of respondents \n who said their living condition is good",xlab = "Gender of respondent splited by geographical area",ylab = "Proportion",names.arg = c("Rural","Urban"))  
legend(x=1,y=0.3,legend = c("Female","Male"),fill = c("blue","red"))



# For those who said their living condition is bad.   
tbl=ftable(my\_data$Q101~my\_data$livingCondition+my\_data$URBRUR);tbl

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 336 325  
## Urban 311 276  
## Good Rural 156 188  
## Urban 268 286

proportion\_tab=round(prop.table(tbl,2),2);proportion\_tab

## my\_data$Q101 Females Males  
## my\_data$livingCondition my\_data$URBRUR   
## Bad Rural 0.31 0.30  
## Urban 0.29 0.26  
## Good Rural 0.15 0.17  
## Urban 0.25 0.27

barplot(proportion\_tab[1:2, ],beside=TRUE,ylim = c(0,0.4),col = c("grey","yellow"),main = "Moderation by geographical area of respondents \n who said their living condition is bad",xlab = "Gender of respondent splited by geographical area",ylab = "Proportion",names.arg = c("Rural","Urban"))  
legend(x=1,y=0.4,legend = c("Female","Male"),fill = c("grey","yellow"))

