Homework 5

Reyna Orellana

4/12/2021

R Markdown

```
## R Markdown for Homework

Homework 5 1. Descriptions of the variables - how they were collected, any missing values, etc
2. How you cleaned and coded the data, including a before/after comparison as needed
3. summary descriptives of the recoded variables
4. appropriate visualizations (not required)2. Description of the relationship between the variables, i
3. Initial demonstration of the relationship, which could include correlation, visualization, or statis
4. (Optional Advanced) Try creating a function that will allow you to easily and accurately implement a
5. (Optional Advanced) If you are working on a model, play around with some of the visual diagnostic to
```

```
#Installing packages that I will need to use markdown
install.packages ("tidyverse")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages ("tinytex")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages ("dplyr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages ("rmarkdown")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
#loading tidyverse
library(tidyverse, tinytex)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rmarkdown)
library(dplyr)
```

I used the General Social Survey (GSS) dataset from 2018. Using GSS allows us to generalize findings to the broader United States population because of its random sampling methods. The GSS contains a standard core of demographic, behavioral, and attitudinal questions, plus other topics of special interest. Of interest to us, I kept 36 variables and constructed 10 varles to analyze.

```
#Add gss data from 2018
library(haven)
GSS2018 <- read_dta("GSS2018.dta")
# shortening dataset and creating new variables in the next chunk
```

My dataset started with 2,348 observations and 1065 variables. I will focus on the questions around abortion and only participants that were asked on ballot 1 and 2 of the surveys will be filtered for that reason. After filtering and selecting my variables which included abortion topic, demographics, and other social status markers that can be measured by factors such as race, educational attainment, and other similar variables, my dataset became smaller. I now have 1559 ovservations and 36 variables.

```
#making my datasetsmaller, keeping only variables that will be used for my analysis
ab18<- select(GSS2018, "abany":"abstate1","age","ballot","bible","born", "class",
              "cohort", "coldeg1", "degree", "depress", "occ10", "partyid", "race",
              "sex", "hispanic", "maeduc", "paeduc", "income", "relig")
ab18<- filter(ab18, ballot != 3)
```

```
#recoding variables from numeric to categorical.
#race
ab18<- ab18 %>%
  mutate(race_3cat = case_when(
    race == 1 ~ "White",
    race == 2 ~ "Black",
    race == 3 ~ "Other"
  ))
#checking
table(select(ab18, race, race_3cat ))
```

```
##     race_3cat
## race Black Other White
##    1     0     0  1120
##    2   252     0     0
##    3     0   187     0
```

```
#age
summary(ab18$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   35.00   50.00   50.12   64.00   89.00       7
```

```
#recoding age variable
ab18 <- ab18%>%
  mutate(age_4cat = case_when( age >= 18 & age <= 29 ~ "18-29",
                               age >= 30 & age <= 39 ~ "30-39",
                               age >= 40 & age <= 49 ~ "40-49",
                               age >= 50 & age <= 89 ~ "50+"))
table(select(ab18, age_4cat))
```

```
##
```

```
## 18-29 30-39 40-49   50+
##   235   275   255   787
```

```
#checking if recoded correctly
table(select(ab18, age_4cat, age))
```

```
##          age
## age_4cat 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##    18-29 13 16  6 16 18 14 23 24 20 28 28 29  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0 33 20 29 27 40 25 21 25 22 33  0
##    40-49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 28
##    50+    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##          age
## age_4cat 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
##    18-29  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    40-49 27 27 26 24 23 30 23 24 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    50+    0  0  0  0  0  0  0  0  0 24 27 22 38 24 32 33 30 28 27 28 28 16 27
##          age
## age_4cat 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
##    18-29  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    30-39  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    40-49  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    50+   25 27 28 17 22 20 33 27 15 18 18 17 13 22  7  8 11  8 10  6  5  8 10
##          age
## age_4cat 87 88 89
##    18-29  0  0  0
##    30-39  0  0  0
##    40-49  0  0  0
##    50+    2  5 21
```

```
#able to have an abortion legally or not. Merging abmlegal and abflegal
ab18 <- ab18 %>%
  mutate(legal = case_when(abfelegl == 1 | abmelegl == 1 ~ 1,
                           abfelegl == 2 | abmelegl == 2 ~ 2,
                           abfelegl == 3 | abmelegl == 3 ~ 3))
ab18 <- ab18 %>%
  mutate(legal_cat = case_when(abfelegl == 1 | abmelegl == 1 ~ "Should",
                               abfelegl == 2 | abmelegl == 2 ~ "Should not",
                               abfelegl == 3 | abmelegl == 3 ~ "Depends"))
#checking
table(select(ab18, legal))
```

```
##
##   1   2   3
## 601 222 689
```

```
#class 1- lower class 2- working class 3- middle class 4- upper class
table(select(ab18, class))
```

```
##
##   1   2   3   4
## 149 654 695  53
```

```
ab18 <- ab18 %>%
  mutate(classr = case_when(class == 1 ~"lower class",
                            class == 2 ~ "working class",
```

```
                            class == 3 ~ "middle class",
                            class == 4 ~ "upper class"))
table(select(ab18, classr))
```

```
##
##   lower class  middle class   upper class working class
##           149           695            53           654
```

```
#recoding help abotion questions. Will add 4 of the help variables abhelp1-abhelp4
ab18<- ab18 %>%
  mutate(abortion_help = case_when(
    abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==1 ~ "Strong Yes",
    abhelp1 == 1 & abhelp2 == 2 & abhelp3 == 1 & abhelp4 ==1 ~ "Weak Yes",
    abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 2 & abhelp4 ==1 ~ "Weak Yes",
    abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==2 ~ "Weak Yes",
    abhelp1 == 2 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==1 ~ "Weak Yes",
    abhelp1 == 1 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==2 ~ "Weak No",
    abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==1 ~ "Weak No",
    abhelp1 == 2 & abhelp2 == 1 & abhelp3 == 2 & abhelp4 ==2 ~ "Weak No",
    abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 1 & abhelp4 ==2 ~ "Weak No",
    abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==2 ~ "Strong No",
    TRUE ~ "Not Available"
  ))
#recoding help abotion questions. Will add 4 of the help variables abhelp1-abhelp4 on a scale of 10
ab18<- ab18 %>%
  mutate(abhelpr = case_when(abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==1 ~ 10,
                             abhelp1 == 1 & abhelp2 == 2 & abhelp3 == 1 & abhelp4 ==1 ~ 9,
                             abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 2 & abhelp4 ==1 ~ 8,
                             abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==2 ~ 7,
                             abhelp1 == 2 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==1 ~ 6,
                             abhelp1 == 1 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==2 ~ 5,
                             abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==1 ~ 4,
                             abhelp1 == 2 & abhelp2 == 1 & abhelp3 == 2 & abhelp4 ==2 ~ 3,
                             abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 1 & abhelp4 ==2 ~ 2,
                             abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==2 ~ 1))
#different way to code abortion help
table(select(ab18, abhelpr))
```

```
##
##   1   2   4   5   6   7   8   9  10
## 150   1 208  18   3   1  18 365 420
```

```
#different way to code abortion help
table(select(ab18, abortion_help))
```

```
##
## Not Available     Strong No    Strong Yes       Weak No      Weak Yes
##           375           150           420           227           387
```

```
ab18<- ab18 %>%
  mutate(abhelp = case_when(
    abhelp1 == 1 & abhelp2 == 1 & abhelp3 == 1 & abhelp4 ==1 ~ "Yes",
    abhelp1 == 2 & abhelp2 == 2 & abhelp3 == 2 & abhelp4 ==2 ~ "No"
  ))
table(select(ab18, abhelp))
```

```
##
##  No Yes
## 150 420
```

```
#sex (56% 868 female, 44% 691 male)
ab18<- ab18 %>%
  mutate(sexp = case_when(sex == 1 ~ "Male",
                          sex == 2 ~ "Female"
  ))

table(select(ab18, sexp))
```

```
##
## Female   Male
##    868    691
```

```
#religion 1- protestant 2- catholic 3- jewish 4- none 5- Other 6- buddhism 7- hinduism 8- other eastern
ab18<- ab18%>%
  mutate(religion = case_when(relig == 1 ~ "Protestant",
                              relig == 2 ~ "Catholic",
                              relig == 3 ~ "Jewish",
                              relig == 4 ~ "None",
                              relig == 5 ~ "Other",
                              relig == 6 ~ "Buddhism",
                              relig == 7 ~ "Hinduism",
                              relig == 8 ~ "Other Eastern religion",
                              relig == 9 ~ "Muslim/Islam",
                              relig == 10 ~ "Orthodox-christian",
                              relig == 11 ~ "Christian",
                              relig == 12 ~ "Native American",
                              relig == 13 ~ "Inter-nondenomentational"
                              ))
prop.table(table(select(ab18, religion)))
```

```
##
##                 Buddhism                Catholic               Christian
##             0.0071197411             0.2330097087             0.0168284790
##                 Hinduism Inter-nondenomentational                   Jewish
##             0.0038834951             0.0006472492             0.0168284790
##             Muslim/Islam          Native American                     None
##             0.0064724919             0.0006472492             0.2194174757
##       Orthodox-christian                    Other   Other Eastern religion
##             0.0019417476             0.0129449838             0.0006472492
##               Protestant
##             0.4796116505
```

My main topic of interest is abortion views, which for purposes of analysis we will make our dependent variable. ##hypothesis My main topic of interest is abortion options, which for purposes of analysis we will make our dependent variable. My null hypothesis is that moral opposition to abortion depends on race and gender. My alternative hypothesis is that moral opposition to abortion does not depends on race and gender

There was no significance difference for being morally opposed to abortion by racial categories or gender. So I reject my alternative hypothesis because moral opposition does not depend on race or gender. Although this might have been because abmoral is not the right type of dv since its categories are only 3 categories that cannot be added up and divided. However in the following model I used two different variables that were better suited. After creating a variable that varied from 10 being strong yes help to 1 being strong no

help for abortion and class. My hypothesis is that there will be no statistical significance. My alternative hypothesis is that there will be statistical significance. In this model there was a significance difference in the help that participants would give to a family member or friend that decided to have an abortion and this changes by class. In this model I reject my null hypothesis because there is statistical significance.

```
#chi square test
#inspect cross-tab with column proportions
prop.table(xtabs(~ abmoral + race_3cat, ab18))
```

```
##        race_3cat
## abmoral      Black      Other      White
##       1 0.03418803 0.03024326 0.21499014
##       2 0.04536489 0.03353057 0.19395135
##       3 0.08218277 0.05588429 0.30966469
```

```
#calculate chisq.test
chisq.test(ab18$abmoral,ab18$race_3cat)
```

```
##
##  Pearson's Chi-squared test
##
## data:  ab18$abmoral and ab18$race_3cat
## X-squared = 9.0347, df = 4, p-value = 0.06024
```

```
#There was no significance differnce for being morally opposed to abortion by racial or gender categori
prop.table(xtabs(~ abmoral + sexp, ab18))
```

```
##        sexp
## abmoral    Female      Male
##       1 0.1663379 0.1130835
##       2 0.1505588 0.1222880
##       3 0.2399737 0.2077581
```

```
#calculate chisq.test
chisq.test(ab18$abmoral,ab18$sexp)
```

```
##
##  Pearson's Chi-squared test
##
## data:  ab18$abmoral and ab18$sexp
## X-squared = 3.7906, df = 2, p-value = 0.1503
```

```
##calculate group means of DV for categorical IV
tapply(ab18$abhelpr,ab18$classr,mean, na.rm=TRUE)
```

```
##   lower class  middle class   upper class working class
##      6.747826      7.707182      8.024390      7.095634
```

```
anova(lm(ab18$abhelpr~ab18$classr))
```

```
## Analysis of Variance Table
##
## Response: ab18$abhelpr
##               Df Sum Sq Mean Sq F value   Pr(>F)
## ab18$classr    3    160  53.327  5.1191 0.001604 **
## Residuals   1176  12251  10.417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#estimate bivariate regression lm(DV~catIV)
summary(lm(ab18$abhelpr~ab18$classr))
```

```
##
## Call:
## lm(formula = ab18$abhelpr ~ ab18$classr)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.024 -3.096  1.904  2.293  3.252
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.7478     0.3010  22.420  < 2e-16 ***
## ab18$classrmiddle class     0.9594     0.3313   2.896  0.00385 **
## ab18$classrupper class      1.2766     0.5871   2.174  0.02987 *
## ab18$classrworking class    0.3478     0.3350   1.038  0.29941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.228 on 1176 degrees of freedom
##   (379 observations deleted due to missingness)
## Multiple R-squared:  0.01289,    Adjusted R-squared:  0.01037
## F-statistic: 5.119 on 3 and 1176 DF,  p-value: 0.001604
```