POL501 - Answers to Problem Set 2

Student Name (REPLACE)

2024-10-14

Table of Contents

# Question 1

## Preliminary Steps.

Preliminaries. First, start by declaring the directory (folder) in which you have the data.

## Setting Up the Directory (REPLACE WITH YOUR FOLDER)  
setwd("C:/Users/Ignacio/Dropbox/PhD SBU/06\_Teaching/00\_POL-501/Problem Sets/PS-2/")  
  
## Confirm the working directory  
getwd()

## [1] "C:/Users/Ignacio/Dropbox/PhD SBU/06\_Teaching/00\_POL-501/Problem Sets/PS-2"

**Important**: Your RMD file must be placed in the same folder in which you have the data. Then, load the dataset to memory

# Load the dataset.  
load('dataframe-ps2.RData')  
# Check the objects loaded  
ls()

## [1] "autoprint" "df\_clean"

View some rows (using glimpse from dplyr, and check the variables using str from base R).

## View the first few rows of the dataset (glimpse is a function of the dyplr package)  
glimpse(df\_clean)

## Rows: 5,199  
## Columns: 8  
## $ RESPID <dbl> 2, 4, 8, 13, 16, 18, 20, 21, 22, 25, 26, 28, 30, 32, 33,…  
## $ PARTY <dbl+lbl> 2, 4, 2, 1, 1, 4, 3, 1, 1, 2, 4, 3, 2, 2, 4, 2, 2, 2…  
## $ INTFREQ <dbl+lbl> 1, 1, 1, 3, 1, 1, 4, 1, 2, 2, 5, 2, 2, 1, 2, 2, 1, 1…  
## $ RADIO <dbl+lbl> 2, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2…  
## $ ECON1MOD <dbl+lbl> 4, 4, 1, 4, 3, 4, 2, 4, 2, 2, 3, 3, 2, 3, 3, 2, 2, 2…  
## $ INFRASPEND <dbl+lbl> 1, 5, 1, 2, 2, 1, 1, 1, 3, 2, 3, 3, 3, 1, 3, 1, 1, 1…  
## $ MOREGUNIMPACT <dbl+lbl> 1, 3, 1, 2, 2, 3, 3, 2, 3, 1, 3, 2, 1, 1, 1, 1, 1, 1…  
## $ CRIMESAFE <dbl+lbl> 2, 2, 1, 5, 2, 4, 3, 2, 1, 1, 2, 2, 2, 3, 3, 2, 2, 2…

str(df\_clean)

## tibble [5,199 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ RESPID : num [1:5199] 2 4 8 13 16 18 20 21 22 25 ...  
## ..- attr(\*, "label")= chr "RESPID. Unique ID"  
## ..- attr(\*, "format.spss")= chr "F8.0"  
## ..- attr(\*, "display\_width")= int 10  
## $ PARTY : dbl+lbl [1:5199] 2, 4, 2, 1, 1, 4, 3, 1, 1, 2, 4, 3, 2, 2, 4, 2, 2, 2,...  
## ..@ label : chr "PARTY. In politics today, do you consider yourself a ..."  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:5] 1 2 3 4 99  
## .. ..- attr(\*, "names")= chr [1:5] "Republican" "Democrat" "Independent" "Something else" ...  
## $ INTFREQ : dbl+lbl [1:5199] 1, 1, 1, 3, 1, 1, 4, 1, 2, 2, 5, 2, 2, 1, 2, 2, 1, 1,...  
## ..@ label : chr "INTFREQ. About how often do you use the internet?"  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 99  
## .. ..- attr(\*, "names")= chr [1:7] "Almost constantly" "Several times a day" "About once a day" "Several times a week" ...  
## $ RADIO : dbl+lbl [1:5199] 2, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2,...  
## ..@ label : chr "RADIO. Do you listen to the radio?"  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:3] 1 2 99  
## .. ..- attr(\*, "names")= chr [1:3] "Yes" "No" "Don't know/Refused/Web blank"  
## $ ECON1MOD : dbl+lbl [1:5199] 4, 4, 1, 4, 3, 4, 2, 4, 2, 2, 3, 3, 2, 3, 3, 2, 2, 2,...  
## ..@ label : chr "ECON1MOD. How would you rate economic conditions in your community today?"  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:5] 1 2 3 4 99  
## .. ..- attr(\*, "names")= chr [1:5] "Excellent" "Good" "Only fair" "Poor" ...  
## $ INFRASPEND : dbl+lbl [1:5199] 1, 5, 1, 2, 2, 1, 1, 1, 3, 2, 3, 3, 3, 1, 3, 1, 1, 1,...  
## ..@ label : chr "INFRASPEND. Thinking about government spending on roads and bridges in the area where you live, do you think th"| \_\_truncated\_\_  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 99  
## .. ..- attr(\*, "names")= chr [1:6] "Increase a lot" "Increase a little" "Stay about the same" "Decrease a little" ...  
## $ MOREGUNIMPACT: dbl+lbl [1:5199] 1, 3, 1, 2, 2, 3, 3, 2, 3, 1, 3, 2, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "MOREGUNIMPACT. If more Americans owned guns, do you think there would be ..."  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:4] 1 2 3 99  
## .. ..- attr(\*, "names")= chr [1:4] "More crime" "Less crime" "No difference" "Don't know/Refused/Web blank"  
## $ CRIMESAFE : dbl+lbl [1:5199] 2, 2, 1, 5, 2, 4, 3, 2, 1, 1, 2, 2, 2, 3, 3, 2, 2, 2,...  
## ..@ label : chr "CRIMESAFE. How would you describe the area where you live, in terms of crime?"  
## ..@ format.spss: chr "F2.0"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 99  
## .. ..- attr(\*, "names")= chr [1:6] "Extremely safe" "Very safe" "Somewhat safe" "Not too safe" ...

Check that the values for each variable correspond to valid numeric values according to the questionnaire and the variables dictionary:

cat('Tabulation of PARTY: ')

## Tabulation of PARTY:

table(df\_clean$PARTY)

##   
## 1 2 3 4   
## 1514 1795 1345 545

cat('\n Tabulation of INTFREQ: ')

##   
## Tabulation of INTFREQ:

table(df\_clean$INTFREQ)

##   
## 1 2 3 4 5   
## 2071 2501 297 203 127

cat('\n Tabulation of RADIO: ')

##   
## Tabulation of RADIO:

table(df\_clean$RADIO)

##   
## 1 2   
## 4058 1141

cat('\n Tabulation of ECON1MOD: ')

##   
## Tabulation of ECON1MOD:

table(df\_clean$ECON1MOD)

##   
## 1 2 3 4   
## 261 1921 1977 1040

cat('\n Tabulation of INFRASPEND: ')

##   
## Tabulation of INFRASPEND:

table(df\_clean$INFRASPEND)

##   
## 1 2 3 4 5   
## 1480 1837 1602 174 106

cat('\n Tabulation of MOREGUNIMPACT: ')

##   
## Tabulation of MOREGUNIMPACT:

table(df\_clean$MOREGUNIMPACT)

##   
## 1 2 3   
## 2255 1320 1624

cat('\n Tabulation of CRIMESAFE: ')

##   
## Tabulation of CRIMESAFE:

table(df\_clean$CRIMESAFE)

##   
## 1 2 3 4 5   
## 450 1974 2218 448 109

Basic Summary Statistics using describe() from the psych package:

describe(df\_clean)

## vars n mean sd median trimmed mad min max range  
## RESPID 1 5199 9457.66 5483.22 9464 9473.45 7080.90 2 18838 18836  
## PARTY 2 5199 2.18 0.97 2 2.10 1.48 1 4 3  
## INTFREQ 3 5199 1.81 0.89 2 1.65 1.48 1 5 4  
## RADIO 4 5199 1.22 0.41 1 1.15 0.00 1 2 1  
## ECON1MOD 5 5199 2.73 0.84 3 2.73 1.48 1 4 3  
## INFRASPEND 6 5199 2.15 0.94 2 2.10 1.48 1 5 4  
## MOREGUNIMPACT 7 5199 1.88 0.86 2 1.85 1.48 1 3 2  
## CRIMESAFE 8 5199 2.58 0.85 3 2.55 1.48 1 5 4  
## skew kurtosis se  
## RESPID -0.02 -1.22 76.05  
## PARTY 0.33 -0.91 0.01  
## INTFREQ 1.54 2.87 0.01  
## RADIO 1.36 -0.16 0.01  
## ECON1MOD 0.02 -0.79 0.01  
## INFRASPEND 0.52 0.04 0.01  
## MOREGUNIMPACT 0.23 -1.59 0.01  
## CRIMESAFE 0.28 0.26 0.01

## Answer to (1.a)

# Step 1: Calculate the total number of valid respondents  
n\_total\_q1 <- nrow(df\_clean)  
# Explanation:  
# - `df\_clean` is assumed to be a cleaned dataset (data frame) that contains the survey data.  
# - `nrow(df\_clean)` calculates the total number of rows (i.e., respondents) in the `df\_clean` data frame.  
# - `n\_total\_q1` stores this total count. This gives us the total number of valid respondents who answered the relevant survey question.  
  
# Step 2: Filter for respondents who identify as either Democrat or Republican  
n\_rep\_dem <- df\_clean %>%  
 filter(PARTY == 1 | PARTY == 2) %>% # Logical OR condition to select Democrats (coded as 2) or Republicans (coded as 1)  
 nrow()  
# Explanation:  
# - `filter(PARTY == 1 | PARTY == 2)` is a filtering condition applied to the `df\_clean` data frame.  
# - `PARTY` is assumed to be the column in `df\_clean` that indicates party affiliation.  
# - PARTY == 1 corresponds to Republicans.  
# - PARTY == 2 corresponds to Democrats.  
# - The `|` symbol is a logical OR operator. It ensures that rows where PARTY is either 1 or 2 (i.e., either Republican or Democrat) are selected.  
# - `nrow()` is used again to count the number of rows (respondents) that satisfy this condition (i.e., either Republican or Democrat).  
# - The result is stored in `n\_rep\_dem`, which represents the number of respondents who identified as either Republican or Democrat.  
  
# Step 3: Calculate the probability of identifying as Democrat or Republican  
prob\_rep\_dem <- n\_rep\_dem / n\_total\_q1  
# Explanation:  
# - `prob\_rep\_dem` calculates the probability that a respondent identifies as either Democrat or Republican.  
# - The formula is straightforward: divide the number of respondents who are either Democrat or Republican (`n\_rep\_dem`) by the total number of respondents (`n\_total\_q1`).  
# - This gives the proportion of respondents who are either Republican or Democrat, which is the probability of interest.  
  
# Step 4: Display the result  
print(paste("Probability of identifying as Democrat or Republican:",   
 round(prob\_rep\_dem, 2)))

## [1] "Probability of identifying as Democrat or Republican: 0.64"

# Explanation:  
# - `print()` outputs the result to the console.  
# - `paste()` concatenates strings and variables together. Here, the text "Probability of identifying as Democrat or Republican:" is combined with the calculated probability `prob\_rep\_dem`.  
# - `round(prob\_rep\_dem, 2)` rounds the probability to two decimal places for cleaner display.  
# - The result is a message that shows the calculated probability, rounded to two decimal places.

The probability of identifying as Democrat or Republican is equal to 64 percent. This was computed by counting the respondents which classified as Democrat **OR** Republican. This is justified since Democrat and Republican are mutually exclusive categories. Then, we take the total of Democrats OR republicans divided by the total number of cases. In other words, **Probability of Democrat or Republican = (Democrats + Republicans)/(Total Respondents)**

## Answer to (1.b)

n\_total\_q2 <- nrow(df\_clean)  
  
n\_radio\_yes <- df\_clean %>%  
 filter(RADIO == 1) %>%   
 nrow()  
  
n\_inter\_constantly <- df\_clean %>%  
 filter(INTFREQ == 1) %>%   
 nrow()   
  
n\_radio\_yes\_AND\_inter\_constantly <- df\_clean %>%  
 filter(RADIO == 1 & INTFREQ == 1) %>% # the character '&' serves as the logical AND (in the dyplr package)  
 nrow()   
  
## Now using the previous calculations we can compute the following probabilities... (you take it from here)

**Explanation/Justification:**

## Answer to (1.c)

**Explanation/Justification:**

## Answer to (1.d)

**Explanation/Justification:**

## Answer to (1.e)

**Explanation/Justification:**

# Question 2

## Answer to (2.a)

**Explanation/Justification:**

## Answer to (2.b)

**Explanation/Justification:**

## Answer to (2.c)

**Explanation/Justification:**

## Answer to (2.d)

**Explanation/Justification:**

## Answer to (2.e)

**Explanation/Justification:**

## Answer to (2.f)

**Explanation/Justification:**