ES 200: Exploring NSEE Survey Data

Week 10 - 2023

Instructions

Working in RMarkdown is a great habit to get into because your writing and analyses are easy to reproduce! No copy and paste errors, no lost scripts that create objects you forget the name of, and they are easy to share. BUT I will note there is one weakness. Sometimes troubleshooting code is easier in a script - but then make sure to move it into the RMarkdown right away.

Assignment Instructions: Create a new Rmarkdown file for yourself. To do this go to the `File` menu then go to `New File`. Select `R Markdown...` A new window will then open, make sure to select PDF for type. Give yourself a cool title if you want - but you can change this later. *Key*: Then copy two chunks from this file, the `setup` and `loadingdata` chunks. Then you are ready to go!

Some tips:

- Name your code chunks, this helps you navigate your document easily. But remember to make them **unique**! Code chunks start with this: ```` ```{r chunkname, echo=TRUE...} ```` Use the `echo=TRUE` for this assignment so I can see what your code looks like! But if you set to `FALSE` just the results are shown.
- Paths to data can be relative or absolute. I recommend relative for now, if you want to go to the pro-level use the `here` package!

For this week, make sure to download the data from Moodle and save it in a folder called "data". This is a good organizational habit. The Rmarkdown file can live in the root with your project file. But you are free to move that if you want to!

Conducting Our Analyses

Loading in and Inspecting data

First, lets load in the data. This uses the SPSS formatted data that has `.sav` as the file extension. What makes this format attractive in the social sciences is that the variables and data are labelled. Meaning that while you might see a `3` or a `4` in the dataframe, the human-readable version is stored there too, such as `Neither Agree nor Disagree` or `Somewhat Agree`. Variables capture the questions we ask people and often have a shorthand name but labels allows the full text to be saved too! You will see how this is handy later.

```
nsee2017 <- read_sav("data/nsee2017.sav")
```

Exploring the Data

The code below lets you peek at the variables, their levels, and the types of data stored. Explore the dataset and identify 3-4 that you are interested in.

```
look_for(nsee2017)[1:10, ] # This is a command from the labelled package, the number range returns the
```

```
##  pos variable             label                col_type values
##  1   RSPDNUM              Respondent Number    dbl
##  2   Sample               Sample Type          chr+lbl  [C] Cell
##                                                         [L] Land
```

```
##  3   version_weather_v1    Placement of weathe~ chr+lbl  [A] Weather questio~
##                                                          [B] Weather questio~
##  4   version_CDR           Version and placeme~ chr+lbl  [C] With extended i~
##                                                          [D] Without extende~
##  5   demog_age_year        Can you please tell~ dbl+lbl  [9999] Masked
##  6   AgeRecode             Age categories       dbl+lbl  [1] 18-29
##                                                          [2] 30-44
##                                                          [3] 45-64
##                                                          [4] Or 65 and older
##                                                          [98] Not Sure
##                                                          [99] Refused
##  7   demog_state           State of residence   chr
##  8   Region                States Recoded into~ dbl+lbl  [1] Northeast
##                                                          [2] South
##                                                          [3] Midwest
##                                                          [4] West
##  9   weather_pastsummer_v1 How would you descr~ dbl+lbl  [1] A lot warmer th~
##                                                          [2] Slightly warmer~
##                                                          [3] Slightly coller~
##                                                          [4] A lot cooler th~
##                                                          [5] About the same ~
##                                                          [98] Not sure
##                                                          [99] refused
##  10  gw_belief             Is there solid evid~ dbl+lbl  [1] Yes
##                                                          [2] No
##                                                          [98] Not sure
##                                                          [99] Refused
```

Step 1: Create Frequency Tables

A first step in survey analysis is to calculate basic frequency tables. Lets start here. Using the variables you discovered above, run frequency tables for each. Use the code below to build the tables.

The code below you can copy. I will step through what it means in class. But for your efforts, the second line with `select` is the key one to focus on. Here you select the variable to include.

Assignment: Create four of these frequency tables with variables you want to explore further. They can be things you want to explain (climate change attitudes) or things you think can help explain outcomes (demographics, political attitudes, etc.)

```
nsee2017 %>%
   select(worldviews_trump, Weight) %>% # Change the values here!
   mutate(across(where(is.labelled), ~as_factor(.x) %>% fct_drop)) %>% # Leave this as is
   as_survey_design(weight=Weight) %>% # Leave this as is
   tbl_svysummary(include=-Weight, missing = "no") %>% # Leave this as is
   as_kable_extra(
    booktabs = TRUE, # Makes it pretty!
    longtable = TRUE,
    linesep = "") %>%
   column_spec(1, width = "20em") # Changes the width of the first column
```

Characteristic	**N = 929**
Agreement with Donald Trump on important issues facing the country today	
All or nearly all issues	92 (10%)

Many, but not all issues	165 (18%)
A few issues	240 (26%)
No or almost no issues	396 (43%)
Not sure	23 (2.5%)

[1] n (%)

Step 2: Two-Way Frequency Tables

Next we will explore looking at how different type of respondents answered questions. Often we use demographic variables to explore potential influences on attitudes. In NSEE these variables start with `demog_`.

Or we might use another question to explore relationships between ideas. We might be interested in political views and policy preferences, such as agreement with Trump and disagreement with climate change policy.

Assignment: Your job is to create at least two tables that are two-way tables with two variables.

```
nsee2017 %>%
   select(worldviews_trump, AgeRecode, Weight) %>% # Select the Variables and include Weight!
   mutate(across(where(is.labelled), ~as_factor(.x) %>% fct_drop)) %>%
   as_survey_design(weight=Weight) %>%
   tbl_svysummary(include=-c(Weight), by=AgeRecode, missing = "no") %>%
   as_kable_extra(
    booktabs = TRUE,
    longtable = TRUE,
    linesep = "") %>%
   column_spec(1, width = "18em") %>% # Adjusted to fit the page
   column_spec(2:5, width = "5em") # Adjusted to fit the page
```

```
## 39 observations missing `AgeRecode` have been removed. To include these observations, use `forcats::
```

Characteristic	**18-29**, N = 222	**30-44**, N = 231	**45-64**, N = 252	**Or 65 and older**, N = 190
Agreement with Donald Trump on important issues facing the country today				
All or nearly all issues	6 (3.0%)	17 (7.3%)	33 (13%)	31 (17%)
Many, but not all issues	29 (13%)	41 (18%)	46 (19%)	41 (22%)
A few issues	69 (32%)	67 (29%)	53 (22%)	43 (23%)
No or almost no issues	108 (49%)	102 (44%)	105 (42%)	69 (37%)
Not sure	7 (3.1%)	4 (1.7%)	10 (4.0%)	2 (1.3%)

[1] n (%)

Step 3: Discussion

Assignment: Share with us your thoughts on how you selected the variables and what you found. What more might you want to know about in these data? This is a short reflection, 300-400 words.

```
nsee2017 %>%
   select(worldviews_trump, AgeRecode, Weight) %>% # Select the Variables and include Weight!
   rec(AgeRecode, rec="1,2=1 [Young]; 3,4=2 [Older]") %>%
   mutate(across(where(is.labelled), ~as_factor(.x) %>% fct_drop)) %>%
   mutate(noweight = 1) %>%
   as_survey_design(weight=noweight) %>%
```

```
  tbl_svysummary(include=-c(AgeRecode, Weight, noweight), by=AgeRecode_r, missing = "no") %>%
  add_p() %>%
  as_kable_extra(
    booktabs = TRUE,
    longtable = TRUE,
    linesep = "") %>%
  column_spec(1, width = "18em") %>% # Adjusted to fit the page
  column_spec(2:5, width = "5em") # Adjusted to fit the page
```

```
## 39 observations missing `AgeRecode_r` have been removed. To include these observations, use `forcats
```

Characteristic	**1**, N = 337	**2**, N = 553	**p-value**
Agreement with Donald Trump on important issues facing the country today			<0.001
All or nearly all issues	20 (6.0%)	82 (15%)	
Many, but not all issues	56 (17%)	118 (22%)	
A few issues	99 (30%)	121 (22%)	
No or almost no issues	150 (45%)	204 (38%)	
Not sure	9 (2.7%)	14 (2.6%)	

[1] n (%)

[2] chi-squared test with Rao & Scott's second-order correction