Exercise 9: Analysis

Ethan vanderWilden

11/12/2020

Review

3. Set up your files and folder structure.

```
getwd()
```

```
## [1] "/Users/ethan/Documents/UW-Madison/Fall 2020/PS811/PS811-Exercises"
```

```
#already in the right place because I am in my exercises R proj
```

4. Read the ANES .dta data into R using the `here` package.

```
anes2016 <- read_dta("anes_timeseries_2016.dta") #loading in data
```

5. Download the ANES 2016 codebook (available on the `ps811/data` repository). We will look at the full sample variables.

Done!

6. You want to know whether owning a house (pre-election) affects which party the respondent choose to contribute to (post-election). Identify these variables from the codebook and rename the variables to names that are easier to reference.

```
#table(anes2016$V161334) #home ownership variable
anes2016$homeowner <- anes2016$V161334 #variable now stored under "homeowner"
anes2016 <- mutate(anes2016,
                   homeowner = case_when(
                     homeowner == 1 ~ 0, #non-homeowner (paying rent)
                     homeowner == 4 ~ 0, #non-homeowner (other arrangement)
                     homeowner == 2 ~ 1, #homeowner (paying mortgage)
                     homeowner == 3 ~ 1)) #homeowner (no payments due)



#table(anes2016$V162014a) #which party respondent donated to
anes2016$party_donate <- anes2016$V162014a
anes2016 <- mutate(anes2016,
                   party_donate = case_when(
                     party_donate == 1 ~ 0, #Donates democrat
                     party_donate == 2 ~ 1)) #Donates republican
#NOTE: not including donations to both parties or to other parties because I think that deviates from t
```

7. Now identify pre-election demographic variables, such as age, gender, and race. Manipulate these variables in ways that you believe would best capture these demographics and explain why you manipulated these variables that way you did. Rename these variables to names that are easier to reference.

```r
#table(anes2016$V161267) #age
anes2016$age <- anes2016$V161267
anes2016$age <- na_if(anes2016$age, -9) #getting rid of NAs
anes2016$age <- na_if(anes2016$age, -8)

#table(anes2016$V161342) #gender
anes2016$gender <- anes2016$V161342
anes2016 <- mutate(anes2016,
                   gender = case_when(
                     gender == 1 ~ 0, #gender = MALE
                     gender == 2 ~ 1)) #gender = FEMALE


#EXPLANATION: This is definitely problematic to be excluding the data to only 2 binary
#genders. However, I'm not too sure of how we might run a regression on this data if we
#cannot make the gender variable numeric, and with the inclusion of more than two genders,
#I think the numeric nature of this data would no longer make sense. I recognize that if I
#were to actually be writing this paper, I would look more into how to treat this data
#without being exclusionary.

#table(anes2016$V161310x) #race
anes2016$race <- anes2016$V161310x
anes2016 <- mutate(anes2016,
                   race = case_when(
                     race == 1 ~ 0, #race = WHITE (non-Hispanic)
                     race == 2 ~ 1, #race = BLACK (non-Hispanic)
                     race == 3 ~ 1, #race = ASIAN/HAWAIIAN/PAC ISLANDER
                     race == 4 ~ 1, #race = NATIVE AMERICAN
                     race == 5 ~ 1, #race = HISPANIC
                     race == 6 ~ 1)) #race = OTHER
#Race variable as binary (0 = White non-Hispanic, 1 = Non-White)

#EXPLANATION: Again, I recognize that it is problematic to turn race into a binary.
#However, I think that if I want to run a regression, I need all of the variable to make
#sense as numeric. Including all of the mentioned races would make the order of the data
#make no sense.
```

8. Provide descriptive summaries for each variable.

```r
anes_test <- anes2016[1843:1847]
anes_test <- na.omit(anes_test)
datasummary((`Homeownership` = homeowner) + (`Party Donation` = party_donate) +
              (`Age` = age) + (`Gender` =gender) + (`Race` = race)
            ~ Mean + SD + Min + Max,
            data = anes_test,
            output = 'latex')
```

	Mean	SD	Min	Max
Homeownership	0.72	0.45	0.00	1.00
Party Donation	0.38	0.49	0.00	1.00
Age	56.98	18.13	18.00	90.00
Gender	0.49	0.50	0.00	1.00
Race	0.22	0.41	0.00	1.00

9. Run an appropriate regression analysis and insert the table into the R Markdown document.

```
model1 <- glm(anes_test$party_donate ~ anes_test$homeowner)
model2 <- glm(anes_test$party_donate ~ anes_test$homeowner + anes_test$age +
                        anes_test$gender + anes_test$race)

# Copying code from https://cimentadaj.github.io/blog/2016-08-22-producing-stargazer-tables-with-odds-r


stargazer2 <- function(model, odd.ratio = F, ...) {
  if(!("list" %in% class(model))) model <- list(model)

  if (odd.ratio) {
    coefOR2 <- lapply(model, function(x) exp(coef(x)))
    seOR2 <- lapply(model, function(x) exp(coef(x)) * summary(x)$coef[, 2])
    p2 <- lapply(model, function(x) summary(x)$coefficients[, 4])
    stargazer(model, coef = coefOR2, se = seOR2, p = p2, ...)

  } else {
    stargazer(model, ...)
  }
}

#Making table with odds ratios for logit regression
models <- list(model1, model2)
stargazer2(models, odd.ratio = T, type = "latex",
           title = "Odds Ratios: Effect of Homeownership on Partisan Donation",
          covariate.labels = c("Homeownership", "Age", "Gender", "Race"),
          dep.var.labels   = "Party Donated To",
          header = FALSE)
```

10. Create a coefficient plot based on the above table.

```
#Coefficient plot
vars <- c('anes_test$homeowner' = 'Homeowner?',
          'anes_test$age' = 'Age',
          'anes_test$gender' = 'Gender',
          'anes_test$race' = 'Race')
modelplot(model2, coef_map = vars)
```

Table 1: Odds Ratios: Effect of Homeownership on Partisan Donation

	Dependent variable:	
	Party Donated To	
	(1)	(2)
Homeownership	1.173***	1.028
	(0.063)	(0.059)
Age		1.006***
		(0.001)
Gender		0.897**
		(0.042)
Race		0.863**
		(0.050)
Constant	1.302***	1.082
	(0.060)	(0.090)
Observations	398	398
Log Likelihood	−273.563	−256.607
Akaike Inf. Crit.	551.126	523.215
Note:	*p<0.1; **p<0.05; ***p<0.01	

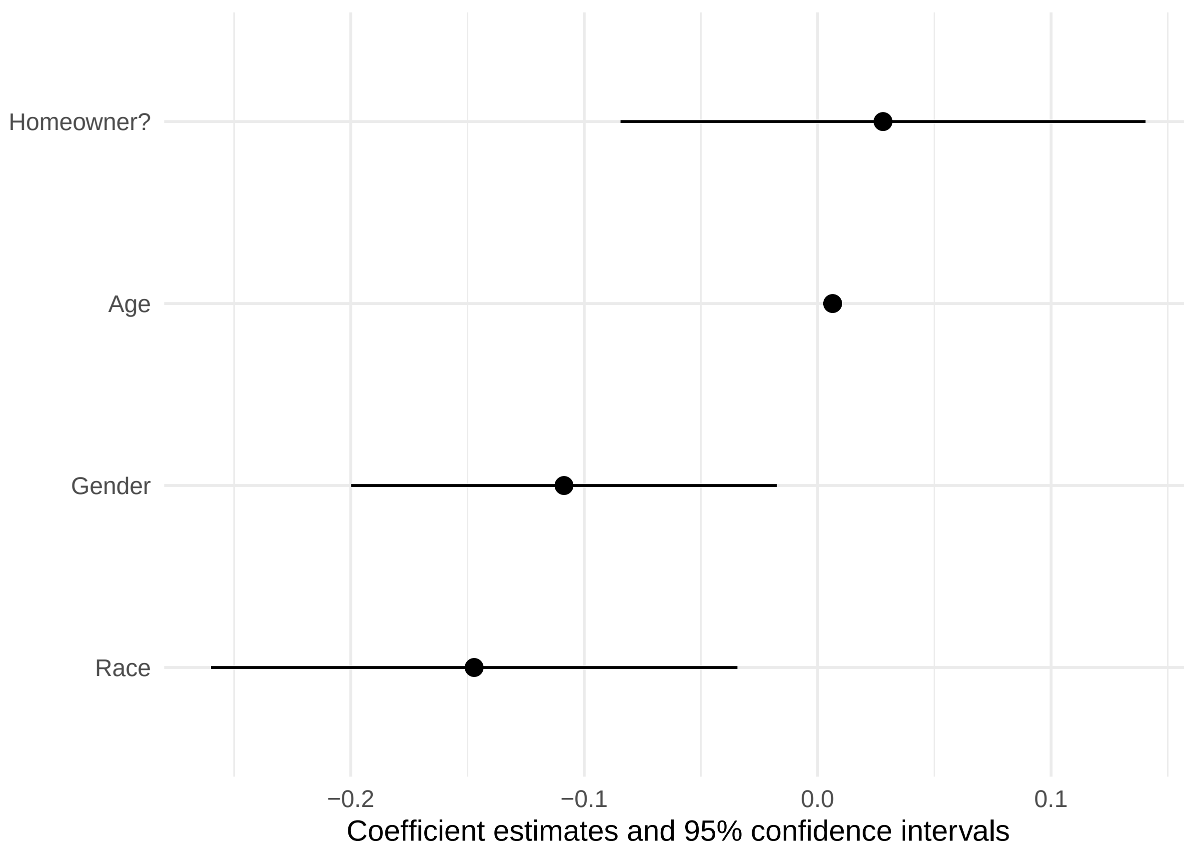

Table 2: Descriptive Statistics for Numeric Variables

	N	Mean	SD	Min	Max
Religious Attendance	3407.00	2.43	1.40	1.00	6.00
Voter Ideology	4041.00	4.76	2.00	1.00	10.00
Simplified Ideology	4041.00	0.73	0.76	0.00	2.00
Sex	4804.00	0.52	0.50	0.00	1.00
Age	4804.00	50.93	17.84	18.00	98.00

Table 3: Religion

data	N	%
Agnostic/Indifferent/Athiest	345	7.2
Non-Practicing Catholic	2199	45.8
Other Religion	154	3.2
Practicing Catholic	1061	22.1

Your project (Tables above included)

Now it's your turn. Use the tools you used today to conduct data analysis for one of your final seminar papers.

1. Create a descriptive statistics summary table for your main variables of interest. Note the number of observations.

```
datasummary((`Religious Attendance` = na.omit(rel_attendance)) +
              (`Voter Ideology` = na.omit(voter_ideology)) +
              (`Simplified Ideology` = na.omit(simple_ideology)) +
              (`Sex` = na.omit(sex)) + (`Age` = na.omit(age))
            ~ (`N` = length) + Mean + SD + Min + Max,
            data = mydata,
            output = 'latex',
            title = "Descriptive Statistics for Numeric Variables")
```

```
#Descriptive Stats for Categorical Religion Variable
datasummary_skim(mydata$religion, type='categorical',
                 title = "Religion", output = 'latex')
```

```
#Descriptive Stats for Categorical Party Vote Variable
datasummary_skim(mydata$partyvote, type='categorical',
                 title = "Party Supported", output = 'latex')
```

Table 4: Party Supported

data	N	%
Cs	248	5.2
Podemos	429	8.9
PP	664	13.8
PSOE	1163	24.2
Vox	362	7.5

2. If you are planning to run a regression, please write out the regression formula. Please take into consideration the dependent variable and its distribution. If you already have the data, you may go ahead and run it. If you do not have the data and is in the process of collecting it, write out the formula. Pre-analysis plans are becoming more common in the discipline, so being able to record what you *plan* to do is becoming increasingly more important.

I am still working out exactly what tests I am going to do. I think I will do a difference in proportions test to demonstrate that the radical right party does not dominate the 'religious vote' among parties on the right.

I am also going to do a regression to assess the relationship between religious attendance and ideology. I still definitely need to do more thinking on what exactly I want to do with my statistics. However, I could run a regression to assess this relationship:

$$RELIGIOSITY_i = \beta_0 + \beta_1 IDEOLOGY_i + \beta_2 AGE_i + \beta_3 SEX_i + \epsilon_i$$

```
no_control <- lm(mydata$voter_ideology ~ mydata$rel_attendance)
model <- lm(mydata$voter_ideology ~ mydata$rel_attendance + mydata$age + mydata$sex)

stargazer(no_control, model,  type = "latex",
           title = "Effect of Religious Attendance  on Voter Ideology",
          covariate.labels = c("Religious Attendance Level", "Age", "Sex"),
          dep.var.labels   = "Voter Ideology",
          header = FALSE)
```

Table 5: Effect of Religious Attendance on Voter Ideology

	Dependent variable:	
	Voter Ideology	
	(1)	(2)
Religious Attendance Level	0.309***	0.334***
	(0.026)	(0.027)
Age		−0.003
		(0.002)
Sex		−0.185**
		(0.072)
Constant	4.452***	4.674***
	(0.071)	(0.121)
Observations	2,823	2,823
R^2	0.049	0.052
Adjusted R^2	0.049	0.051
Residual Std. Error	1.880 (df = 2821)	1.877 (df = 2819)
F Statistic	146.448*** (df = 1; 2821)	51.948*** (df = 3; 2819)
Note:		*p<0.1; **p<0.05; ***p<0.01

Submit

Email me (mshieh2@wisc.edu) the link to your `ps811-exercises` repository when you are done.