Exercise 9: Analysis (Answer Key)

Marcy Shieh

11/12/2020

Review

1. Download the .Rmd version of this file from GitHub and change the author to your name.
2. Load packages.
3. Set up your files and folder structure.

```
# head of the folder
# some noticed that here() doesn't work but here::here() does
# the reason is that the plyr package (which I think is loaded with tidyverse) contains a here() comman
# so you want to call the here() command from the here package, thus specifying it
# to see whether the here command is in more than one package you've loaded and which ones
# run ?here
here::here()
```

```
## [1] "/Users/marcyshieh/ps811"
```

```
# a folder that contains files you need for this assignment (might not be the case for you!)
here::here("data")
```

```
## [1] "/Users/marcyshieh/ps811/data"
```

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

```
# anes_timeseries_2016.dta file is in my "data" folder...yours might not be
anes2016 <- read_dta(here::here("data", "anes_timeseries_2016.dta"))
```

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

```
# recode homeownership variable
anes2016$homeowner <- anes2016$V161334

# create a new variable that recodes everything
# I prefer keeping the homeowner variable for reference (even though V161334 is still in there)
anes2016 <- mutate(anes2016,
                   homeowner_adj = 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)
```

```
# recode party that respondent donated (in post-election survey)
anes2016$party_donate <- anes2016$V162016a

anes2016 <- mutate(anes2016,
                   party_donate = case_when(
                   party_donate == 1 ~ 0, # contributed to dem party
                   party_donate == 2 ~ 1)) # contributed to rep party
```

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.

```
# rename age
anes2016$age <- anes2016$V161267
table(anes2016$age) # check age variable
```

```
##
##  -9  -8  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
## 120   1  28  39  54  53  43  56  55  63  70  70  60  60  69  78  86  73  81  79
##  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55
##  69  85  66  75  51  66  67  51  46  70  83  66  53  67  59  72  59  81  78  78
##  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75
##  88  76  94  96  89  80  69  72  75  69  88  78  94  55  54  49  53  38  45  34
##  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##  41  26  29  21  24  24  24  12  13  12  14   9  12   6  27
```

```
# I would also group the ages...
# you can consider different ways of grouping ages
# but I'm going to follow Pew Research
# https://www.pewresearch.org/fact-tank/2020/07/09/younger-adults-differ-from-older-ones-in-perceptions
anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age >= 18 & age <= 29, "18-29", NA))

anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age >= 30 & age <= 49, "30-49", age_grouped))

anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age >= 50 & age <= 64, "50-64", age_grouped))

anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age >= 65, "65+", age_grouped))

anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age >= 65, "65+", age_grouped))

anes2016 <- mutate(anes2016,
                   age_grouped = ifelse(
                    age == -9 | age == -8, NA, age_grouped))
```

```r
table(anes2016$age_grouped)
```

```
##
## 18-29 30-49 50-64   65+
##   651  1381  1166   951
```

```r
# rename gender
anes2016$gender <- anes2016$V161342

# one way that many researchers have tried to deal with the issue of binary gender variables
# is by coding "male", "female" and "other"
# in general, there are not many "other" respondents...
table(anes2016$gender)
```

```
##
##   -9    1    2    3
##   41 1987 2231   11
```

```r
# so it depends on what you are looking at
# if you are most interested in the behavior of respondents who identify as female
# you will recode it to female...
# instead of thinking of gender as "male" or "female," you can think of it as "female" or "not female"
anes2016 <- mutate(anes2016,
                   female = case_when(
                   gender == 2 ~ 1, # gender is female = 1
                   gender != 2 ~ 0)) # gender is not female = 0

# you can include the others as well...
# of course you can do this in one big tidy chain, but personally i like to do this one at a time...
# if you are interested in the behavior of respondents who identify as male then you can look at the ma
anes2016 <- mutate(anes2016,
                   male = case_when(
                   gender == 1 ~ 1, # gender is male = 1
                   gender != 1 ~ 0)) # gender is not male = 0

anes2016 <- mutate(anes2016,
                   gender_other = case_when(
                   gender == 3 ~ 1, # gender is other = 1
                   gender != 3 ~ 0)) # gender is not other = 0

# you will find that there are not very many respondents in the dataset that identify as "other"
table(anes2016$gender_other)
```

```
##
##    0    1
## 4259   11
```

```r
# many researchers will argue that there are not enough respondents who identify as "other" for us to l
# one way that people have tried to get around this is by "weighing" certain variables
# https://bookdown.org/jespasareig/Book_How_to_weight_a_survey/introduction.html
# however, it seems like 2020 pollsters made sure to weigh their variables this time around
# and still messed up :'(

# alternatively, you can also do this...

anes2016 <- mutate(anes2016,
```

```
                gender_recode = ifelse(
                      gender == 2, "male", NA)
                )

anes2016 <- mutate(anes2016,
                gender_recode = ifelse(
                      gender == 1, "female", gender_recode)
                )

anes2016 <- mutate(anes2016,
                gender_recode = ifelse(
                      gender == 3, "other", gender_recode)
                )

anes2016 <- mutate(anes2016,
                gender_recode = ifelse(
                      gender == -9, NA, gender_recode)
                )

table(anes2016$gender_recode)
```

```
##
## female   male  other
##   1987   2231     11
```

```
# rename race
anes2016$race <- anes2016$V161310x

# so you can do white or non-white like you're a researcher from the 1980s...
anes2016 <- mutate(anes2016,
                      white = case_when(
                      race == 1 ~ 1, # race is white
                      gender != 1 ~ 0)) # grace is not white
table(anes2016$white)
```

```
##
##    0    1
##  666 3038
```

```
# as you can see, ANES over-samples white people
```

8. Provide descriptive summaries for each variable.

```
anes_summary <- dplyr::select(anes2016, homeowner_adj, party_donate, age_grouped, age, gender, gender_r
datasummary((`Homeownership` = homeowner_adj) + (`Party Donation` = party_donate) +
              (`Age` = age) + (`Gender` = gender) + (`Race` = race) ~
            Mean + SD + Min + Max,
            data = anes_summary,
            output = "latex")
```

	Mean	SD	Min	Max
Homeownership	0.62	0.48	0.00	1.00
Party Donation	0.43	0.50	0.00	1.00
Age	47.92	19.87	-9.00	90.00
Gender	1.43	1.14	-9.00	3.00
Race	1.79	1.57	-2.00	6.00

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

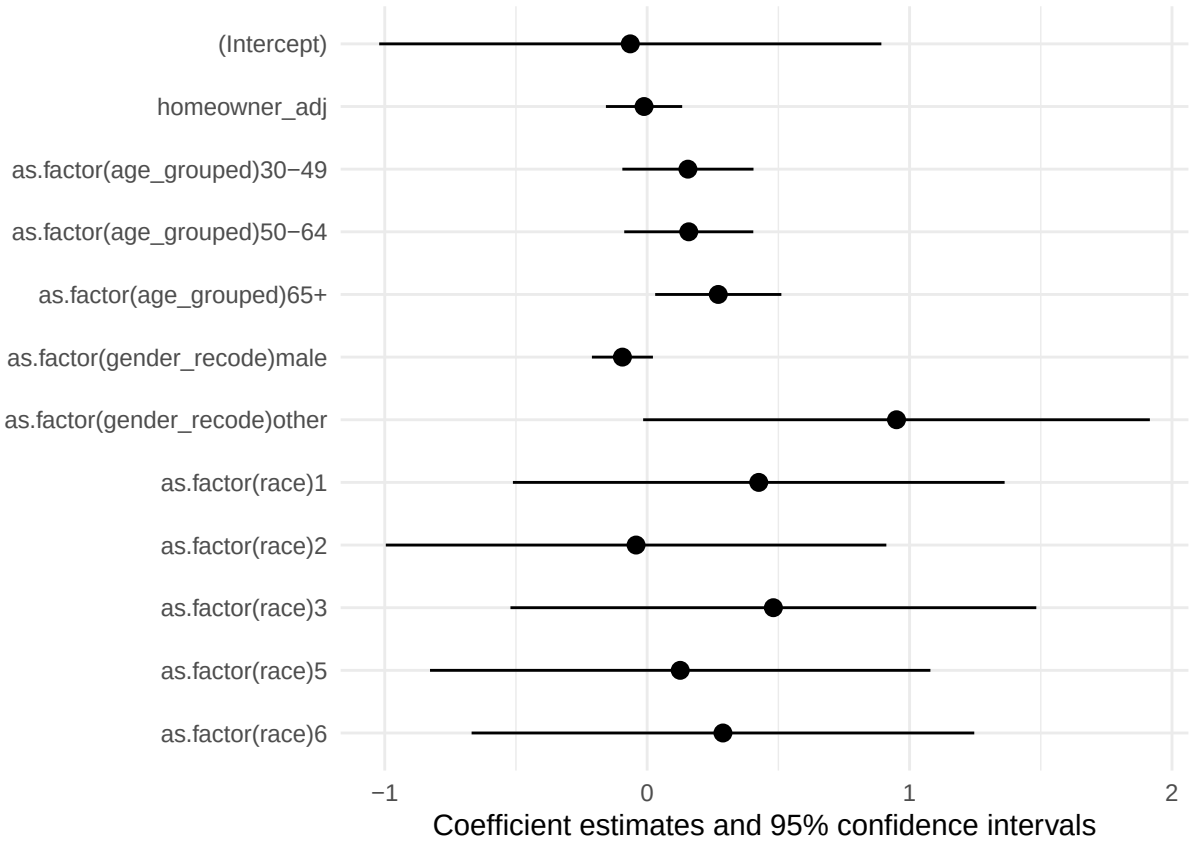
```
#Making table with odds ratios for logit regression
models <- list(model1, model2)
stargazer2(models, odd.ratio = T, type = "latex",
           title = "Effect of Homeownership on Partisan Donation",
          covariate.labels = c("Homeownership",
                               "Age:30-49",
                               "Age:50-64",
                               "Age:65+",
                               "Gender:Male", # reference gender is female
                               "Gender:Other",
                               "Race:White", # reference race is white
                               "Race:Black",
                               "Race:Asian",
                               "Race:Hispanic",
                               "Race:Other"),
          dep.var.labels   = "Donated to Republican Party",
          header = FALSE)
```

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

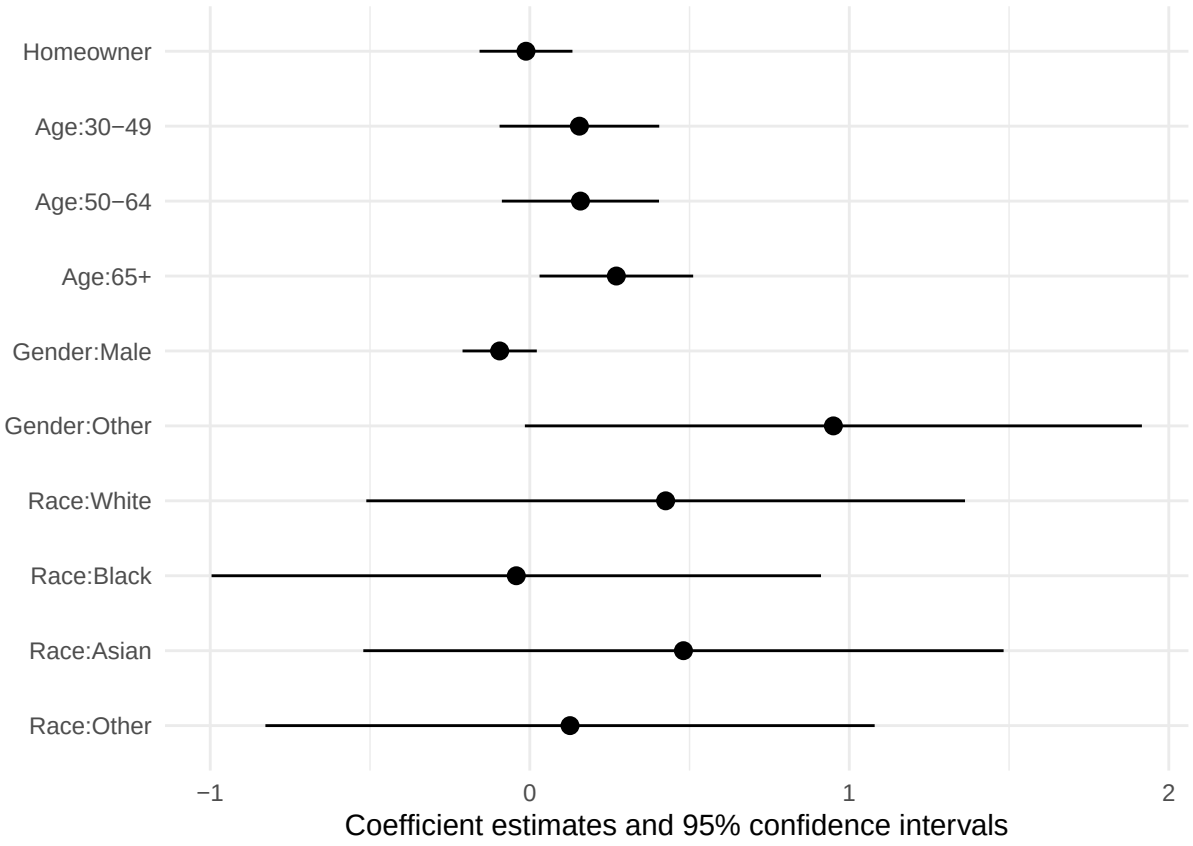
```
# you need to do a plain modelplot first so you know what to rename...
modelplot(model2)
```

Table 1: Effect of Homeownership on Partisan Donation

	Dependent variable:	
	Donated to Republican Party	
	(1)	(2)
Homeownership	1.162**	0.988
	(0.078)	(0.073)
Age:30-49		1.168
		(0.149)
Age:50-64		1.172
		(0.147)
Age:65+		1.311**
		(0.161)
Gender:Male		0.910
		(0.054)
Gender:Other		2.586*
		(1.274)
Race:White		1.530
		(0.731)
Race:Black		0.959
		(0.466)
Race:Asian		1.618
		(0.827)
Race:Hispanic		1.134
		(0.552)
Race:Other		1.335
		(0.652)
Constant	1.383***	0.938
	(0.079)	(0.458)
Observations	272	263
Log Likelihood	−193.508	−169.372
Akaike Inf. Crit.	391.015	362.744
Note:	*p<0.1; **p<0.05; ***p<0.01	




```
# rename
vars <- c('homeowner_adj' = 'Homeowner',
          'as.factor(age_grouped)30-49' = 'Age:30-49',
          'as.factor(age_grouped)50-64' = 'Age:50-64',
          'as.factor(age_grouped)65+' = 'Age:65+',
          'as.factor(gender_recode)male' = 'Gender:Male',
          'as.factor(gender_recode)other' = 'Gender:Other',
          'as.factor(race)1' = 'Race:White',
          'as.factor(race)2' = 'Race:Black',
          'as.factor(race)3' = 'Race:Asian',
          'as.factor(race)4' = 'Race:Hispanic',
          'as.factor(race)5' = 'Race:Other')
modelplot(model2, coef_map = vars)
```



Your project

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.
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.

Submit

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