Exercise 9: Analysis

Evgeniya Mitrokhia

11/15/2020

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.

```
here::here("ps9")
```

```
## [1] "C:/Users/user/Documents/PS811_exercises/ps9"
```

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

```
data <- read_dta(here::here("ps9", "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.

```
#names(data)

#owning a house (pre-election) - V161334
#party the respondent choose to contribute to (post-election) -V162016a

names(data)[names(data) == "V161334"] <- "home_own"
names(data)[names(data) == "V162016a"] <- "party_contr"

table(data$home_own)
```

```
##
##   -9   -8    1    2    3    4
##   34    2 1286 1754  886  308
```

```
table(data$party_contr) #1 - dem, 2 - republican
```

```
##
##   -7   -6   -1    1    2    3    5
##   86  536 3364  154  118    5    7
```

```
newdata <- subset(data, home_own != -9 & home_own != -8)
newdata <- subset(data, party_contr == 1 | party_contr == 2)

table(newdata$party_contr) #I leave only party ientification
```

```
##
##   1   2
## 154 118
```

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.

```
#Gender - V161342
#Race - V161310x
#Age - V161267
#Education - V161270
#Employment - V161277

names(newdata)[names(newdata) == "V161342"] <- "gender"
names(newdata)[names(newdata) == "V161310x"] <- "race"
names(newdata)[names(newdata) == "V161267"] <- "age"
names(newdata)[names(newdata) == "V161270"] <- "educ"
names(newdata)[names(newdata) == "V161277"] <- "empl"

table(newdata$gender)
```

```
##
##  -9   1   2   3
##   4 136 131   1
```

```
newdata <- subset(newdata, gender == 1 | gender == 2)
newdata$gender[newdata$gender == 2] <- 0

table(newdata$gender) #0 - female, 1 - male
```

```
##
##   0   1
## 131 136
```

```
table(newdata$race)
```

```
##
##  -2   1   2   3   5   6
##   1 198  20   9  23  16
```

```
newdata <- subset(newdata, race != -2)

newdata$race[newdata$race == 2] <- 0
newdata$race[newdata$race == 3] <- 0
newdata$race[newdata$race == 5] <- 0
newdata$race[newdata$race == 6] <- 0

table(newdata$race) #1 - white, 0 - other
```

```
##
##   0   1
##  68 198
```

```
table(newdata$age)
```

```
##
```

```
## -9 18 19 20 21 23 25 26 27 28 31 32 33 34 35 36 37 39 42 43 44 45 46 47 48 49
##  5  2  1  3  1  4  4  3  2  1  5  2  5  1  6  2  2  3  2  1  3  2  4  4  2  3
## 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
##  3  3  6  2  5  5  4  4  4  4  6  5  4  8  7  4  7  9 10  7  6  3  9  3  8  6
## 76 77 78 79 80 81 82 83 84 85 87 88 89 90
##  6  5  6  4  4  6  3  3  1  4  2  1  1  5
```

```
newdata <- subset(newdata, age != -9)

table(newdata$educ)
```

```
##
##  4  5  7  8  9 10 11 12 13 14 15 16 90
##  1  2  3  3 38 41 22 14 59 59  9  9  1
```

```
table(newdata$empl)
```

```
##
##   1   4   5   6   7   8
## 113  13 114   9   9   3
```

8. Provide descriptive summaries for each variable.

```
myvars <- c("home_own", "party_contr", "gender", "race", "age", "educ", "empl")
use_data <- newdata[myvars]

summary(use_data)
```

```
##     home_own      party_contr        gender            race
##  Min.   :1.000   Min.   :1.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :2.000   Median :1.000   Median :1.0000   Median :1.000
##  Mean   :2.253   Mean   :1.437   Mean   :0.5134   Mean   :0.751
##  3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :4.000   Max.   :2.000   Max.   :1.0000   Max.   :1.000
##       age             educ           empl
##  Min.   :18.00   Min.   : 4.0   Min.   :1.000
##  1st Qu.:49.00   1st Qu.:10.0   1st Qu.:1.000
##  Median :63.00   Median :13.0   Median :5.000
##  Mean   :59.67   Mean   :12.2   Mean   :3.356
##  3rd Qu.:73.00   3rd Qu.:14.0   3rd Qu.:5.000
##  Max.   :90.00   Max.   :90.0   Max.   :8.000
```

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

```
table(use_data$party_contr) #1 - dem, 0 - republican
```

```
##
##   1   2
## 147 114
```

```
use_data$party_contr[use_data$party_contr == 2] <- 0

model1<- glm(party_contr ~ home_own + gender + race + age, data = use_data, family = "binomial")
summary(model1)
```

```
##
## Call:
## glm(formula = party_contr ~ home_own + gender + race + age, family = "binomial",
```

```
##     data = use_data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.055  -1.107   0.547   1.079   1.553
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.902003   0.600412   4.833 1.34e-06 ***
## home_own    -0.171434   0.171755  -0.998 0.318215
## gender      -0.436332   0.266466  -1.637 0.101530
## race        -1.242645   0.350561  -3.545 0.000393 ***
## age         -0.017506   0.008487  -2.063 0.039132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 357.64  on 260  degrees of freedom
## Residual deviance: 326.01  on 256  degrees of freedom
## AIC: 336.01
##
## Number of Fisher Scoring iterations: 3
```

```
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(model1 , type = 'html',
          title = "Effect of homeownership on party contributions",
          covariate.labels = c("Gender", "Age", "Race"),
          dep.var.labels = "Party Contribution (1=GOP)",
          header = FALSE
          )
```

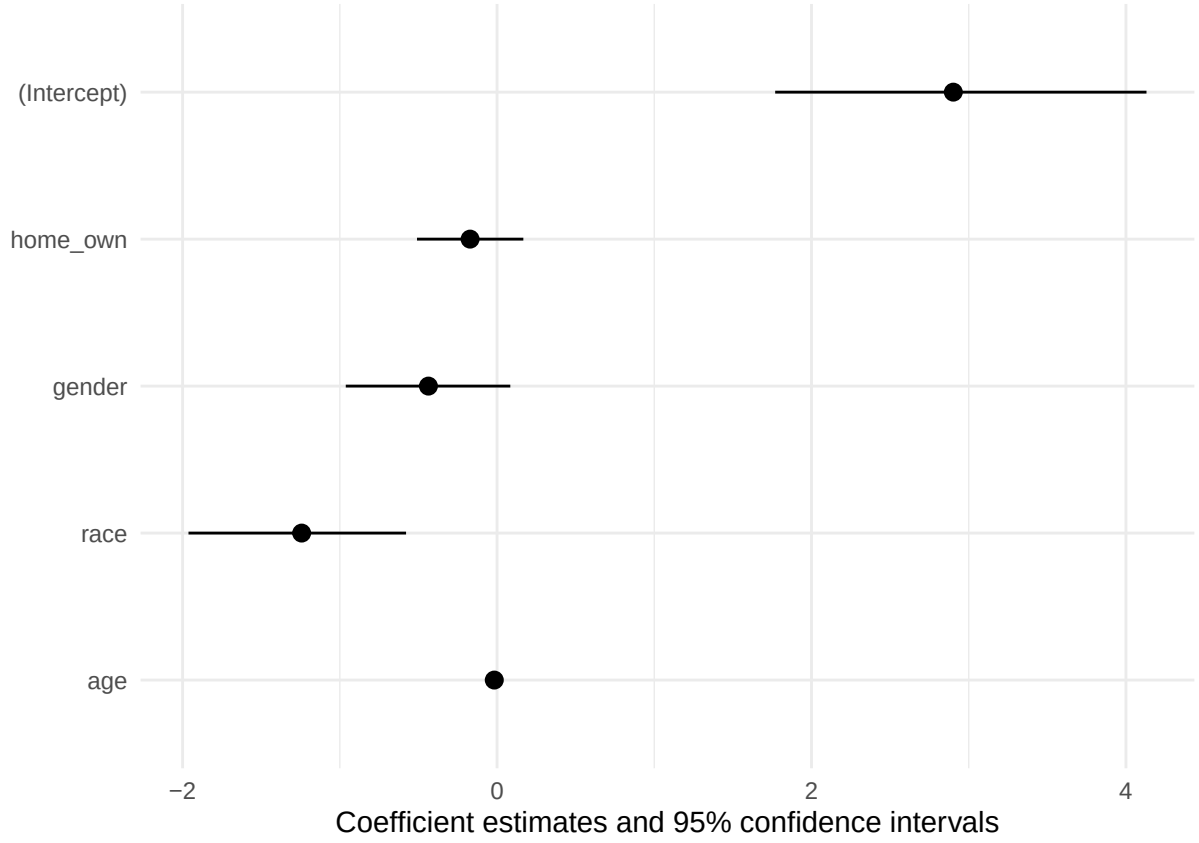
```
##
## <table style="text-align:center"><caption><strong>Effect of homeownership on party contributions</str
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td></td><td colspan="1" style="border-bottom: 1px solid black"></td></tr>
## <tr><td style="text-align:left"></td><td>Party Contribution (1=GOP)</td></tr>
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left"></td><td>(0.172)</td></tr>
## <tr><td style="text-align:left"></td><td></td></tr>
## <tr><td style="text-align:left">Age</td><td>-0.436</td></tr>
## <tr><td style="text-align:left"></td><td>(0.266)</td></tr>
## <tr><td style="text-align:left"></td><td></td></tr>
## <tr><td style="text-align:left">Race</td><td>-1.243<sup>***</sup></td></tr>
## <tr><td style="text-align:left"></td><td>(0.351)</td></tr>
## <tr><td style="text-align:left"></td><td></td></tr>
## <tr><td style="text-align:left">age</td><td>-0.018<sup>**</sup></td></tr>
## <tr><td style="text-align:left"></td><td>(0.008)</td></tr>
```

```
## <tr><td style="text-align:left"></td><td></td></tr>
## <tr><td style="text-align:left">Constant</td><td>2.902<sup>***</sup></td></tr>
## <tr><td style="text-align:left"></td><td>(0.600)</td></tr>
## <tr><td style="text-align:left"></td><td></td></tr>
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left">Log Likelihood</td><td>-163.006</td></tr>
## <tr><td style="text-align:left">Akaike Inf. Crit.</td><td>336.011</td></tr>
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## </table>
```

```
#the code creates latex code but in the pdf it is not shown as a table (just the code)
```

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

```
modelplot(model1)
```



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.