POL212 TA Session

Gento Kato

February 6, 2019

```
## Clear Workspace
rm(list = ls())

## Set Working Directory to the File location
## (If using RStudio, can be set automatically)
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()
```

```
## [1] "C:/GoogleDrive/Lectures/2019_01to03_UCD/POL212_TA/POL212_TA_resource"
```

```
## Required Package
library(foreign)
```

Practice of Analysis

1. Download 2016 General Social Survey Data in Stata Format from HERE. (See Individual Data Section)
2. Open Data in R

```
# Set Data Location (Set on your own)
dataloc <- "D:/BoxSync/Data/GSS/2016/data/GSS2016.sav"

# Load Data
d <- read.spss(dataloc, to.data.frame=TRUE,
               use.value.labels = FALSE, use.missings = FALSE)
```

3. Check the codebook HERE. Find following variables in data.

- Preference between Public Access to Government Information and Public Security (govtinfo)
- Safety in respondent's neighborhood
- Respondent's age
- Respondent's gender
- Respondent's marital status
- Number of children

Check the distribution by either table() or summary() function.

```
table(d$GOVTINFO)
```

```
##
##   -1    0    1    2    3    4    5    6    7    8    9   10   98   99
## 1477   39   31   51  102   66  311  139  175  186   58  200   19   13
```

```
table(d$NEISAFE)
```

```
##
##    1    2    3    4    8    9
## 1595  973  229   42    1   27
```

```
table(d$AGE)
```

```
##
```

```
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
##  7 33 26 33 44 49 35 56 42 58 42 56 54 57 42 54 49 56 52 58 44 42 46 36 50
## 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67
## 45 52 27 45 55 46 41 48 49 65 60 53 48 48 70 67 58 53 56 56 43 34 44 47 49
## 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 99
## 43 42 32 27 26 22 24 19 25 23 26 21 25 21 11 22 11 11 12  9  3 22 10
```

```
table(d$SEX)
```

```
##
##    1    2
## 1276 1591
```

```
table(d$MARITAL)
```

```
##
##    1    2    3    4    5    9
## 1212  251  495  102  806    1
```

```
table(d$CHILDS)
```

```
##
##   0   1   2   3   4   5   6   7   8   9
## 797 459 733 467 213  92  51  25  22   8
```

4. Build a new data.frame with transformed variables. Transformation includes:

- Drop Missing Values
- Combine the categories with too few N
- Define categorical variables as "factor" and set labels and reference category
- Recode binary variables to have values 1 and 0.

```
# Initiate Data.Frame with No Columns, but same number of rows as d
dnew <- d[,FALSE]


# Information Access Preference
?ifelse
dnew$GOVTINFO <- ifelse(d$GOVTINFO %in% c(-1,98,99),
                        NA,
                        d$GOVTINFO)
table(dnew$GOVTINFO)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10
##  39  31  51 102  66 311 139 175 186  58 200
```

```
# Neighborhood safety
dnew$NEISAFE <- ifelse(d$NEISAFE %in% c(8,9),
                       NA,
                       4-d$NEISAFE)
table(dnew$NEISAFE)
```

```
##
##    0    1    2    3
##   42  229  973 1595
```

```
dnew$NEISAFE[d$NEISAFE==4] <- 1
dnew$NEISAFE <- factor(dnew$NEISAFE,
```

```
                        labels=c("Unsafe","Sw Safe","Very Safe"))
table(dnew$NEISAFE)
```

```
##
##    Unsafe   Sw Safe Very Safe
##       271       973      1595
```

```
# Age
dnew$AGE <- ifelse(d$AGE==99,NA,d$AGE)
table(dnew$AGE)
```

```
##
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
##  7 33 26 33 44 49 35 56 42 58 42 56 54 57 42 54 49 56 52 58 44 42 46 36 50
## 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67
## 45 52 27 45 55 46 41 48 49 65 60 53 48 48 70 67 58 53 56 56 43 34 44 47 49
## 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
## 43 42 32 27 26 22 24 19 25 23 26 21 25 21 11 22 11 11 12  9  3 22
```

```
# Female
dnew$FEM <- d$SEX - 1
table(dnew$FEM)
```

```
##
##    0    1
## 1276 1591
```

```
# MARITAL STATUS
dnew$MARITAL <- ifelse(d$MARITAL==9,NA,d$MARITAL)
dnew$MARITAL <- factor(dnew$MARITAL,
                       labels=c("Married","Widowed","Divorced",
                                "Separated","Never Married"))
dnew$MARITAL <- relevel(dnew$MARITAL, ref="Never Married")
table(dnew$MARITAL)
```

```
##
## Never Married       Married       Widowed      Divorced     Separated
##           806          1212           251           495           102
```

```
# CHILDS
dnew$CHILDS <- ifelse(d$CHILDS==9,NA,d$CHILDS)
table(dnew$CHILDS)
```

```
##
##   0   1   2   3   4   5   6   7   8
## 797 459 733 467 213  92  51  25  22
```

5. Run OLS regression to test the hypothesis that the more safe the neghborhood is, more strongly the respondent support for public access to government information. Include age, gender, marital status, number of children as control variables. Show result summary.

```
# Conventional Way
m <- lm(GOVTINFO ~ NEISAFE + AGE + FEM + MARITAL + CHILDS, data=dnew)
# All remaining variables in data as right-hand-side variables
m <- lm(GOVTINFO~., data=dnew)
# Show Summary
summary(m)
```

```
##
```

```
## Call:
## lm(formula = GOVTINFO ~ ., data = dnew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0211 -1.4945 -0.0582  1.7488  4.9501
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.083029   0.310492  16.371  < 2e-16 ***
## NEISAFESw Safe   -0.348286   0.261331  -1.333  0.18285
## NEISAFEVery Safe -0.253912   0.253096  -1.003  0.31594
## AGE               0.011672   0.004885   2.389  0.01703 *
## FEM               0.431965   0.142210   3.038  0.00243 **
## MARITALMarried    0.621543   0.194853   3.190  0.00146 **
## MARITALWidowed   -0.097675   0.329634  -0.296  0.76704
## MARITALDivorced   0.466053   0.241461   1.930  0.05380 .
## MARITALSeparated -0.085109   0.373081  -0.228  0.81958
## CHILDS            0.088590   0.048266   1.835  0.06666 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.536 on 1330 degrees of freedom
##   (1527 observations deleted due to missingness)
## Multiple R-squared:  0.04091,    Adjusted R-squared:  0.03442
## F-statistic: 6.304 on 9 and 1330 DF,  p-value: 8.953e-09
```

```
# APSR TABLE
#install.packages("apsrtable")
library(apsrtable)
apsrtable(m)
```

```
## \begin{table}[!ht]
## \caption{}
## \label{}
## \begin{tabular}{ l D{.}{.}{2} }
## \hline
##   & \multicolumn{ 1 }{ c }{ Model 1 } \\ \hline
##  %                & Model 1\\
## (Intercept)      & 5.08 ^*\\
##                  & (0.31) \\
## NEISAFESw Safe   & -0.35  \\
##                  & (0.26) \\
## NEISAFEVery Safe & -0.25  \\
##                  & (0.25) \\
## AGE              & 0.01 ^*\\
##                  & (0.00) \\
## FEM              & 0.43 ^*\\
##                  & (0.14) \\
## MARITALMarried   & 0.62 ^*\\
##                  & (0.19) \\
## MARITALWidowed   & -0.10  \\
##                  & (0.33) \\
## MARITALDivorced  & 0.47   \\
##                  & (0.24) \\
```

```
## MARITALSeparated & -0.09  \\
##                   & (0.37) \\
## CHILDS            & 0.09   \\
##                   & (0.05)  \\
##  $N$               & 1340   \\
## $R^2$             & 0.04   \\
## adj. $R^2$        & 0.03   \\
## Resid. sd         & 2.54    \\ \hline
##  \multicolumn{2}{l}{\footnotesize{Standard errors in parentheses}}\\
## \multicolumn{2}{l}{\footnotesize{$^*$ indicates significance at $p< 0.05 $}}
## \end{tabular}
##  \end{table}
```

```
# I like Texreg
#install.packages("texreg")
library(texreg)
```

```
## Version:  1.36.23
## Date:     2017-03-03
## Author:   Philip Leifeld (University of Glasgow)
##
## Please cite the JSS article in your publications -- see citation("texreg").
```

```
m1 <- lm(GOVTINFO ~ NEISAFE, data=dnew)
m2 <- lm(GOVTINFO ~ ., data=dnew)
screenreg(list(m1,m2)) # in R console
```

```
##
## ========================================
##                    Model 1      Model 2
## ----------------------------------------
## (Intercept)          6.26 ***     5.08 ***
##                     (0.23)       (0.31)
## NEISAFESw Safe      -0.30        -0.35
##                     (0.26)       (0.26)
## NEISAFEVery Safe    -0.05        -0.25
##                     (0.25)       (0.25)
## AGE                               0.01 *
##                                  (0.00)
## FEM                               0.43 **
##                                  (0.14)
## MARITALMarried                    0.62 **
##                                  (0.19)
## MARITALWidowed                   -0.10
##                                  (0.33)
## MARITALDivorced                   0.47
##                                  (0.24)
## MARITALSeparated                 -0.09
##                                  (0.37)
## CHILDS                            0.09
##                                  (0.05)
## ----------------------------------------
## R^2                  0.00         0.04
## Adj. R^2             0.00         0.03
## Num. obs.         1345         1340
```

```
## RMSE                 2.58         2.54
## ========================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```
texreg(list(m1,m2)) # in tex
```

```
##
## \begin{table}
## \begin{center}
## \begin{tabular}{l c c }
## \hline
##  & Model 1 & Model 2 \\
## \hline
## (Intercept)      & $6.26^{***}$ & $5.08^{***}$ \\
##                  & $(0.23)$     & $(0.31)$     \\
## NEISAFESw Safe   & $-0.30$      & $-0.35$      \\
##                  & $(0.26)$     & $(0.26)$     \\
## NEISAFEVery Safe & $-0.05$      & $-0.25$      \\
##                  & $(0.25)$     & $(0.25)$     \\
## AGE              &              & $0.01^{*}$   \\
##                  &              & $(0.00)$     \\
## FEM              &              & $0.43^{**}$  \\
##                  &              & $(0.14)$     \\
## MARITALMarried   &              & $0.62^{**}$  \\
##                  &              & $(0.19)$     \\
## MARITALWidowed   &              & $-0.10$      \\
##                  &              & $(0.33)$     \\
## MARITALDivorced  &              & $0.47$       \\
##                  &              & $(0.24)$     \\
## MARITALSeparated &              & $-0.09$      \\
##                  &              & $(0.37)$     \\
## CHILDS           &              & $0.09$       \\
##                  &              & $(0.05)$     \\
## \hline
## R$^2$            & 0.00         & 0.04         \\
## Adj. R$^2$       & 0.00         & 0.03         \\
## Num. obs.        & 1345         & 1340         \\
## RMSE             & 2.58         & 2.54         \\
## \hline
## \multicolumn{3}{l}{\scriptsize{$^{***}p<0.001$, $^{**}p<0.01$, $^*p<0.05$}}
## \end{tabular}
## \caption{Statistical models}
## \label{table:coefficients}
## \end{center}
## \end{table}
```

6. Interpret the result. What is the direct interpretation of coefficients? Is hypothesis supported or not? What are the potential issues in the analysis?

7. Try using your own data to run simple regression.