Module 4 - Instructions

Victoria Gruner

14/04/2021

#In the next assignment we want to replicate some plots from the paper "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues" (Washington, 2008). The paper explores whether having a daughter makes politicians more sensitive to women's rights issues and how this is reflected in their voting behavior. The main identifying assumption is that after controlling for the number of children, the gender composition is random. This might be violated if families that have a preference for girls keep having children until they have a girl. In this assignment we will prepare a dataset that allows us to test whether families engage in such a "female child stopping rule".

Setup

#* Load the libraries "Rio" and "tidyverse" #* Change the path of the working directory to your working directory.

```
library(rio)
library(tidyverse)
library(magrittr)
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.5
```

```
#install.packages("knitr")


setwd("C:/Users/Victoria/Desktop/INTRO R/Module_4")
```

#* import the data sets *basic.dta* and *genold108.dta* #* create a subset of the 108th congress from the *basic* dataset #* join this subset with the *genold* dataset

```
basic.data <- import("basic.dta")
genold.data <- import("genold108.dta")


basic108.data <- filter(basic.data, congress==108)


combined.data <- left_join(genold.data, basic108.data)
```

Data preparation

- check table 1 in the appendix of the paper and decide which variables are necessary for the analysis (check the footnote for control variables)
- drop all other variables. ——>#* Recode *genold* such that gender is a factor variable and missing values are coded as NAs. ——>#* Recode *party* as a factor with 3 levels (D, R, I)
- Recode *rgroup* and *region* as factors.
- generate variables for age squared and service length squared
- create an additional variable of the number of children as factor variable

```r
congress.data <- combined.data %>% select(name, ngirls, anygirls, genold, totchi, party, rgroup,region,

congress.data$genold %<>% na_if("") %>% as.factor()

#alternative approach:
#congress.data$genold <- as.factor(congress.data$genold)
#congress.data$genold[congress.data$genold != 'B'& congress.data$genold != 'G'] <- NA

congress.data$party <- as.factor(congress.data$party)

congress.data$party <- recode(congress.data$party, "1"="D", "2"="R", "3"="I")


congress.data$region <- as.factor(congress.data$region)
congress.data$rgroup <- as.factor(congress.data$rgroup)

congress.data <- congress.data %>% mutate(age_sq = age^2)
congress.data <- congress.data %>% mutate(srvlng_sq = srvlng^2)
congress.data <- congress.data %>% mutate(totchi_factor = as.factor(totchi))
```

Replicationg Table 1 from the Appendix

#We haven't covered regressions in R yet. Use the function *lm()*. The function takes the regression model
(formula) and the data as an input. The model is written as $y \sim x$, where x stands for any linear combination
of regressors (e.g. $y \sim x_1 + x_2 + female$). Use the help file to understand the function.

- Run the regression $total.children = \beta_0 + \beta_1 gender.oldest + \gamma'X$ where γ stands for a vector of coefficients
  and X is a matrix that contains all columns that are control variables.[1]

#total children

```r
lm.reg_tot <- lm(totchi ~ party + region + rgroup+age+age_sq+female+white+srvlng+srvlng_sq + genold, con

sd_1 <- summary(lm.reg_tot)$coefficients[22,2]
beta_1 <- summary(lm.reg_tot)$coefficients[22,1]

#number of girls
lm.reg_g <- lm(ngirls ~ party + region + rgroup+age+age_sq+female+white+srvlng+srvlng_sq + genold, cong

sd_1_g <- summary(lm.reg_g)$coefficients[22,2]
beta_1_g <- summary(lm.reg_g)$coefficients[22,1]
```

#* Run the same regression separately for Democrats and Republicans (assign the independent to one of the
parties). Save the coefficient and standard error of *genold*

```r
congress.rep <- filter(congress.data, party=='R')

lm.reg_rep <- lm(totchi ~ genold + region +rgroup+age+age_sq+female+white+srvlng+srvlng_sq, congress.re

sd_1_rep <- summary(lm.reg_rep)$coefficients[2,2]
beta_1_rep <- summary(lm.reg_rep)$coefficients[2,1]
```

---

[1]This is just a short notation instead of writing the full model with all control variables $totchi = \beta_0 + \beta_1 genold + \gamma_1 age + \gamma_2 age^2 + \gamma_3 Democrat + ... + \epsilon$ which quickly gets out of hand for large models.

```
#sd_rep <- summary(lm.reg_rep)$coefficients[2,2]

lm.reg_rep_g <- lm(ngirls ~ genold + region +rgroup+age+age_sq+female+white+srvlng+srvlng_sq , congress
sd_1_rep_g <- summary(lm.reg_rep_g)$coefficients[2,2]
beta_1_rep_g <- summary(lm.reg_rep_g)$coefficients[2,1]



congress.dem <- filter(congress.data, party=='D' | party=='I')

lm.reg_dem <- lm(totchi ~ genold + region +rgroup+age+age_sq+female+white+srvlng+srvlng_sq , congress.de
sd_1_dem <- summary(lm.reg_dem)$coefficients[2,2]
beta_1_dem <- summary(lm.reg_dem)$coefficients[2,1]


lm.reg_dem_g <- lm(ngirls ~ genold + region +rgroup+age+age_sq+female+white+srvlng+srvlng_sq , congress
sd_1_dem_g <- summary(lm.reg_dem_g)$coefficients[2,2]
beta_1_dem_g <- summary(lm.reg_dem_g)$coefficients[2,1]
```

#* Collect all the *genold* coefficients from the six regressions, including their standard errors and arrange
them in a table as in the paper. #* print the table

```
beta_vector <- matrix(c(beta_1, beta_1_g, beta_1_rep, beta_1_rep_g, beta_1_dem, beta_1_dem_g, sd_1, sd_
colnames(beta_vector) <- c('Full Congress Total Children', 'Full Congress Girls',' Democrats Total Chil
rownames(beta_vector) <- c('Beta', 'Standard Error', 'N')
beta_vector <- as.table(beta_vector)
print(beta_vector)
```

```
##                 Full Congress Total Children Full Congress Girls
## Beta                            -0.08388331          1.18608123
## Standard Error                   1.35753057          0.07291434
## N                               -0.28239327          1.45800412
##                 Democrats Total Children Democrats Girls
## Beta                          0.14911332      0.18896038
## Standard Error                0.12758694      0.18089377
## N                             0.23346451      0.17575881
##                 Republican Total Children Republican Girls
## Beta                         227.00000000     105.00000000
## Standard Error               227.00000000     122.00000000
## N                            105.00000000     122.00000000
```

export(beta_vector,"Congress.xls")

#stargazer(beta_vector, #header = FALSE, type = 'latex', #dep.var.labels = c('Full Congress','Democrats','Republicans'),
#column.labels = c('Total Children', 'Girls','Total Children', 'Girls','Total Children', 'Girls'), #column.separate = c(2,2,2)))