

Module 4 - Instructions

Oliver Engist

04/04/2021

In the next assignment we want to replicate some plots from the paper “Female Socialization: How Daughters Affect Their Legislators’ 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 “R” and “tidyverse”
- Change the path of the working directory to your working directory.

```
library(R)
library(tidyverse)
```

```
#load knitr and magrittr
```

```
library(knitr)
library(magrittr)
```

```
#load stargazer for plotting
library(stargazer)
```

- 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 <- import("basic.dta")
genold <- import("genold108.dta")
basic108 <- basic %>% filter(basic$congress==108)
combineddata <- inner_join(basic108, genold, by=c("name", "district", "statenam"))
#I viewed both data sets and saw that name, district and statenam seemed to be
#identifying variables. If they contained the exact same combinations of name,
#district, and state, an inner-join would create a new data set also with 435
#observations, which it did. Also I double checked that the name variable had
#the same name in both data sets, which it did. We see that we have 62 variables,
#which is the genold variable added to the 61 variables in basic108.
```

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

```
#keep only relevant variables
combineddata %<>% select(party, statenam, ngirls, totchi,
                        rgroup, region, srvlng, female, white, age, genold)

#recode *genold* to factor. Observations not fitting the levels automatically become NA
combineddata$genold %<>% factor(levels=c("G", "B"))

#recode party
combineddata$party %<>% as.character()
combineddata$party %<>% recode("1" = "D", "2" = "R", "3" = "I")
combineddata$party %<>% factor(levels=c("D", "R", "I"))

#recode rgroup and region
combineddata$rgroup %<>% factor(levels=c("0", "1", "2", "3", "4"))
combineddata$region %<>% factor(levels=c("1", "2", "3", "4", "5", "6", "7", "8", "9"))

#create new variables
combineddata %<>% mutate(agesq = age^2)
combineddata %<>% mutate(srvlngsq = srvlng^2)
combineddata %<>% mutate(numchi = factor(totchi))
```

Replicating 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
- Save the main coefficient of interest (β_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*
- 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

```
options(width=80)
#Creating subsets for democrats and repub, and running the six regressions and saving
#coefficients and standard deviations.

Demssubset <- combineddata %>% filter(party=="D")
Repssubset <- combineddata %>% filter(party=="R")
```

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

```

####number of daughters, full congress
#numchi works as a fixed effect since it is a factor variable, created before. Will be in
#all regressions on number of daughters.
reg1 <- lm(ngirls~genold + white + female + party + age + agesq + srvlng + srvlngsq +
          rgroup + region + numchi, combineddata)
#save the beta1 from the reg1 list
beta1_dau_con <- -reg1[["coefficients"]][["genoldB"]]
#save the summary list and enter it to find the link to the standard deviation since I
#did not find the standard deviation in the reg1 list. The SD of beta1 is the second
#number in the second column in the "genoldB"-row.
sum_dau_con <- summary(reg1)
SD_dau_con <- sum_dau_con[["coefficients"]][["genoldB",2]]

####number of children, full congress
reg2 <- lm(totchi~genold + white + female + party + age + agesq + srvlng + srvlngsq +
          rgroup + region, combineddata)
#save the beta1 from the reg2 list:
beta1_chi_con <- -reg2[["coefficients"]][["genoldB"]]
#save summary list to save SD, as before.
sum_chi_con <- summary(reg2)
SD_chi_con <- sum_chi_con[["coefficients"]][["genoldB",2]]

####number of daughters, democrats
reg3 <- lm(ngirls~genold + white + female + age + agesq + srvlng + srvlngsq + rgroup +
          region + numchi, Demsubset)
#save the beta1 from the reg3 list:
beta1_dau_dem <- -reg3[["coefficients"]][["genoldB"]]
#save the summary list to save SD:
sum_dau_dem <- summary(reg3)
SD_dau_dem <- sum_dau_dem[["coefficients"]][["genoldB",2]]

####number of children, democrats
reg4 <- lm(totchi~genold + white + female + age + agesq + srvlng + srvlngsq + rgroup +
          region, Demsubset)
#save the beta1 from the reg4 list:
beta1_chi_dem <- -reg4[["coefficients"]][["genoldB"]]
#save the summary list to save SD:
sum_chi_dem <- summary(reg4)
SD_chi_dem <- sum_chi_dem[["coefficients"]][["genoldB",2]]

####number of daughters, republicans
reg5 <- lm(ngirls~genold + white + female + age + agesq + srvlng + srvlngsq + rgroup +
          region + numchi, Repsubset)
#save the beta1 from the reg5 list:
beta1_dau_rep <- -reg5[["coefficients"]][["genoldB"]]
#save the summary list to save SD:
sum_dau_rep <- summary(reg5)
SD_dau_rep <- sum_dau_rep[["coefficients"]][["genoldB",2]]

####number of children, republicans
reg6 <- lm(totchi~genold + white + female + age + agesq + srvlng + srvlngsq + rgroup +
          region, Repsubset)
#save the beta1 from the reg6 list:

```

```

betal_chi_rep <- -reg6[["coefficients"]][["genoldB"]]
#save the summary list to save SD:
sum_chi_rep <- summary(reg6)
SD_chi_rep <- sum_chi_rep[["coefficients"]][["genoldB",2]]

#Since the genold takes the value B as = 1, we attain the correct coefficients, but of
#the opposite sign. I solve this by setting a minus sign in front of the value when saving
#the coefficient.

##Create and print table
stargazer(reg1,reg2,reg3,reg4,reg5,reg6,type='text', title="Appendix Table 1: Evidence on Legislator Child Gender Mix Selection, 108th Congress",
          align=TRUE,column.labels = c("Congress","Democrats", "Republicans"),
          column.separate = c(2,2,2), digits=2, keep.stat="n",object.names = TRUE, keep = "genoldB", notes = "for whom gender of the first born could be established Number of children",
          "regressions include controls for legislator race, gender, party, age, ",
          "age squared, service length and its square, religion, and region Number of",
          "daughters regressions include the preceding covariates, as well as fixed",
          "effects for total number of children Standard errors in parentheses"),
          notes.align = "l")

```

```

##
## Appendix Table 1: Evidence on Legislator Child Gender Mix Selection, 108th Congress
## =====
##                               Dependent variable:
##                               -----
##               ngirls      totchi      ngirls      totchi      ngirls      totchi
##               Congress      Democrats      Republicans
##               (1)      (2)      (3)      (4)      (5)      (6)
##               reg1      reg2      reg3      reg4      reg5      reg6
## -----
## genoldB      -1.36***      0.08      -1.39***      -0.09      -1.23***      0.28
##               (0.08)      (0.15)      (0.11)      (0.18)      (0.11)      (0.23)
## -----
## Observations      227      227      104      104      122      122
## =====
## Note:      *p<0.1; **p<0.05; ***p<0.01
##            The sample includes the 227 of the 381 parent members of the 108th Congress,
##            for whom gender of the first born could be established Number of children
##            regressions include controls for legislator race, gender, party, age,
##            age squared, service length and its square, religion, and region Number of
##            daughters regressions include the preceding covariates, as well as fixed
##            effects for total number of children Standard errors in parentheses

```