

Module 4 - Instructions

Oliver Engist, edited by 24025

04/04/2021

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.

```
*{r, message=FALSE, warning=FALSE}
```

```
#To clear global environment
rm(list = ls())

#In the beginning I thought I did the set up the right way, now in the end I
#obviously realized that I wasn't connected to github at all during the whole time.
#I've learned now how I should do it and the work is all my own.
#I'm sorry about this.

#loading libraries and setting working directory
library(rio)
library(tidyverse)
library(knitr)
library(stargazer)
library(yaml)

setwd("C:/Users/erica/Documents/Skola/Handels/Master/7316 - 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

```
#Importing data
basic <- import("./basic.dta")
genold108 <- import("./genold108.dta")
```

```

#Filtering by 108
basic108 <- basic %>% filter(congress == "108")

#Joining on name, statename and district b/s they are the same over both datasets
data <- left_join(basic108, genold108, by=c("name", "statenam", "district"))

```

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

```

#Selecting the variables we need for recreating table 1:
#Totalnumber of children, number of girls, gender of oldest child, gender, age,
#service length, religion, region and color.
table1 <- select(data, totchi, ngirls, party, genold, female, age, srvlng, rgroup, region, white)

#Recode genold to factor and fix missing values
table1$genold <- table1$genold %>% na_if("") %>% as.factor ()

#Recode political party to factor and change 1->D, 2->R and 3->I
table1$party <- table1$party %>% as.factor() %>% recode("1"="D", "2"="R", "3"="I")

#Also, recode rgroup and region to factors
table1$rgroup <- table1$rgroup %>% as.factor()
table1$region <- table1$region %>% as.factor()

#Create age squared and service length squared
table1 <- table1 %>% mutate(agesq = age^2)
table1 <- table1 %>% mutate(srvlngsq = srvlng^2)

#Create total number of children as a factor
table1 <- table1 %>% mutate(totchifac = as.factor(totchi))

#Checking that it worked
str(table1)

## 'data.frame': 435 obs. of 13 variables:
## $ totchi : num 0 3 1 6 3 2 2 4 5 0 ...
## ..- attr(*, "label")= chr "Total children"
## ..- attr(*, "format.stata")= chr "%9.0g"
## $ ngirls : num 0 1 1 2 2 2 2 2 2 0 ...
## ..- attr(*, "format.stata")= chr "%8.0g"
## $ party : Factor w/ 3 levels "D","R","I": 1 1 2 2 1 1 1 1 2 1 ...
## $ genold : Factor w/ 2 levels "B","G": NA NA 2 1 NA 2 2 NA NA NA ...
## $ female : num 0 0 0 0 0 0 0 0 0 0 ...

```

```
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ age      : num  64 60 37 55 56 57 45 55 55 46 ...
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ srvlng   : num  13 21 7 3 1 7 13 3 11 5 ...
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ rgroup   : Factor w/ 5 levels "0","1","2","3",...: 1 5 2 2 2 2 3 2 2 ...
## $ region   : Factor w/ 9 levels "1","2","3","4",...: 9 2 6 4 7 1 2 9 6 9 ...
## $ white    : num   1 1 1 1 1 1 1 0 1 1 ...
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ agesq    : num  4096 3600 1369 3025 3136 ...
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ srvlngsq : num   169 441 49 9 1 49 169 9 121 25 ...
##   .- attr(*, "format.stata")= chr "%9.0g"
## $ totchifac: Factor w/ 12 levels "0","1","2","3",...: 1 4 2 7 4 3 3 5 6 1 ...
```

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

```
?lm

#Running the regression with total number of children as dependent,
#but I couldn't get the X to work...
#And saving the coefficient for genold
reg_1 <- lm(totchi~ genold+age+agesq+party+srvlng+srvlngsq+rgroup+region+white, table1)
reg_1_coeff <- reg_1[["coefficients"]][["genoldG"]]

#Same thing but depending on party
reg_1D <- lm(totchi~ genold+age+agesq+srvlng+srvlngsq+rgroup+region+white, table1, party=="D" )
reg_1R <- lm(totchi~ genold+age+agesq+srvlng+srvlngsq+rgroup+region+white, table1, party=="R" )

reg_1D_coeff <- reg_1D[["coefficients"]][["genoldG"]]
reg_1R_coeff <- reg_1R[["coefficients"]][["genoldG"]]

#Then we do it with number of girls as dependent variable
reg_2 <- lm(ngirls~ genold+age+agesq+party+srvlng+srvlngsq+rgroup+region+white, table1)
```

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.

```

reg_2D <- lm(ngirls~ genold+age+agesq+srvlng+srvlngsq+rgroup+region+white, table1, party=="D" )
reg_2R <- lm(ngirls~ genold+age+agesq+srvlng+srvlngsq+rgroup+region+white, table1, party=="R" )

reg_2_coeff <- reg_2[["coefficients"]][["genoldG"]]
reg_2D_coeff <- reg_2D[["coefficients"]][["genoldG"]]
reg_2R_coeff <- reg_2R[["coefficients"]][["genoldG"]]

#We need to find standard errors of all these regressions
#First we summarize the regression
reg_1_sum <- summary(reg_1)
#And then extract standard errors, it's the second number of the b1 coeff-row.
reg_1_SD <- reg_1_sum[["coefficients"]][["genoldG",2]]

#It works, so I'll do it with th rest of the regressions aswell
reg_1D_sum <- summary(reg_1D)
reg_1R_sum <- summary(reg_1R)
reg_2_sum <- summary(reg_2)
reg_2D_sum <- summary(reg_2D)
reg_2R_sum <- summary(reg_2R)

reg_1D_SD <- reg_1D_sum[["coefficients"]][["genoldG",2]]
reg_1R_SD <- reg_1R_sum[["coefficients"]][["genoldG",2]]
reg_2_SD <- reg_2_sum[["coefficients"]][["genoldG",2]]
reg_2D_SD <- reg_2D_sum[["coefficients"]][["genoldG",2]]
reg_2R_SD <- reg_2R_sum[["coefficients"]][["genoldG",2]]

#Now to the table
stargazer(reg_1,reg_1D,reg_1R,reg_2,reg_2D,reg_2R,type='text', title="Appendix Table 1: Evidence on Leg
          align=TRUE,column.labels = c("Congress","Democrats", "Republicans"),
          column.separate = c(2,2,2), digits=2, keep.stat="n",object.names = TRUE, keep = "genoldG", no
"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:
##                               -----
##                               totchi                               ngirls
##                               Congress                               Democrats                               Republicans
##                               (1)                               (2)                               (3)                               (4)                               (5)                               (6)
##                               reg                               reg                               reg                               reg                               reg                               reg
## -----
## genoldG                               -0.08                               0.12                               -0.26                               1.36***                               1.51***                               1.23***
##                               (0.15)                               (0.18)                               (0.23)                               (0.13)                               (0.18)                               (0.19)
##
## -----
## Observations                               227                               104                               122                               227                               104                               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
```