Problem Set 1

Rohan Krishnan

2024-01-23

Note: I created this homework document on my local machine by trying to replicate the homework pdf as closely as I could. Apologies if there are any slight discrepancies!

Empirical Analysis using Data from Washington (2008, AER)

This exercise uses data from Ebonya Washington's paper, "Female Socialization: How Daughters Affect their Legislator Father's voting on Women's Issues," published in the American Economic Review in 2008. This paper studies whether having a daughter affects legislator's voting on women's issues.

Finding the data

I have downloaded Washington's basic.dta file and made it available in the RCloud assignment workspace. I downloaded this data from the AER's website which links you to the ICPSR's data repository. Anyone can sign in to get access to the replication data files. These include the typical files in a replication folder: several datasets, several .do files (which is a STATA command file), and text files with the data descriptions which tell you about the different variables included in the dataset.

Set up and opening the data

Because this is a .dta file, you will need to open it with the read.dta function that is included in the haven packages.

Other packages you will need: dplyr, ggplot2 and stargazer.

If you are working on a desktop version of R (i.e not in the cloud workspace) and have not used a package before you will need to install the packages by un-commenting the following code. If you are working in R Studio Cloud these should load automatically or you will be prompted to load them.

```
#install.packages('haven',repos = "http://cran.us.r-project.org")
#install.packages("dplyr",repos = "http://cran.us.r-project.org")
#install.packages("stargazer",repos = "http://cran.us.r-project.org")
#install.packages("ggplot2",repos = "http://cran.us.r-project.org")
```

Hint: Once you have run these once, on your machine, you may want to comment them out with a # so that your code runs faster.

Question 1.1:

In the following chunk, call all the packages you will be using with the library function

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
```

Question 1.2:

Below, create a code chunk in which you load your data. Remember, since basic.dta is a .dta file, you will use the read.dta() function to load it.

```
basic <- read_dta("~/Downloads/Basic.dta")
```

Question 1.3:

How many observations are in the original dataset? Hint: use the nrow() function

Code and Answer:

```
nrow(basic) #There are 1740 observations in the original dataset
```

```
## [1] 1740
```

Cleaning the data

Question 2.1:

The original dataset contains data from the 105th to 108th U.S. Congress reported in the variable congress. We only want to keep the observations from the 105th congress.

Hint: Use the filter function in the dplyr package.

Code:

```
basic2.1 <- basic %>%
  filter(congress == 105)
```

Question 2.2:

The dataset contains many variables, some of which are not used in this exercise. Keep the following variables in the final dataset:

Hint: use the select function in dplyr.

Name	Description
aauw	AAUW score
nowtot	NOW score
totchi	Total number of children
ngirls	Number of daughters
party	Political party. Democrats if 1, Republicans if 2, and Independent if 3.
female	Female dummy variable
age	Age

You can find the detailed description of each variable in the original paper. The main variable in this analysis is AAUW, a score created by the American Association of University Women (AAUW). For each congress, AAUW selects pieces of legislation in the areas of education, equality, and reproductive rights. The AAUW keeps track of how each legislator voted on these pieces of legislation and whether their vote aligned with the AAUW's position. The legislator's score is equal to the proportion of these votes made in agreement with the AAUW.

Code:

```
basic2.2 <- basic2.1 %>%
  select(aauw, nowtot, totchi, ngirls, party, female, age)
```

Question 2.3:

Make sure your final dataset is a data frame. You can check your data's format with the command is(). If the first element of the returned vector is not "data.frame", convert your dataset with the function as.data.frame().

Code:

```
#Check if 'basic2.2' is a data frame
is(basic2.2) #First vector is "tbl_df"
```

```
## [1] "tbl_df"     "tbl"        "data.frame" "list"       "oldClass"
## [6] "vector"
```

```
#Convert to data frame
basic2.2 <- as.data.frame(basic2.2)

#Check if 'basic2.2' is a data frame
is(basic2.2) #First vector is "data.frame"
```

```
## [1] "data.frame" "list"       "oldClass"   "vector"
```

Summary Statistics

Question 3.1:

Report summary statistics for all the remaining variables in the dataset. Present these summary statistics in a formatted table, you can use stargazer or other packages. Make this table as communicative as possible.

Hints: If you want RMarkdown to display your outputted table, include the code results = "asis" in the chunk header. This is true for all chunks that output a formatted table. In the stargazer command, you will want to specify the format of the table by including the code type="latex" for PDF output. If you have trouble knitting to PDF, try installing MikTex (https://miktex.org/download)

Code:

```
stargazer(basic2.2, type = "latex", title = "Question 3.1 Summary Statistics"
          , header = FALSE, no.space = TRUE)
```

Table 2: Question 3.1 Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
aauw	435	47.308	42.021	0	100
nowtot	431	41.311	36.534	0	100
totchi	434	2.493	1.648	0	10
ngirls	434	1.274	1.125	0	7
party	435	1.529	0.504	1	3
female	435	0.110	0.314	0	1
age	435	51.671	9.618	26	87

Generate Variables

Question 4.1:

Construct a variable called *repub*, a binary set to 1 if the observation is for a republican, 0 otherwise

Code:

```
basic2.2$repub <- NA
basic2.2$repub[basic2.2$party == 2] <- 1
basic2.2$repub[basic2.2$party != 2] <- 0
```

Question 4.2:

Construct a variable called *age2*, where age2 = age^2.

Code

```
basic2.2$age2 <- NA
basic2.2$age2 <- (basic2.2$age)^2
```

Analysis

Question 5.1 (2 pages):

Estimate the following linear regression models using the felm command (part of the lfe package). Report your regression results in a formatted table using stargazer. Report robust standard errors in your table.

Hints:

- in stargazer specify se = list(model1$rse, model2$rse, model3$rse) and type = "latex".
- your estimates of β_1 should be similar, but not exactly the same, as the estimate in the first row, as the estimate in the first row, second column of table 2 in Washington (2008).

  Model 1 : $aauw_i = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi_i + \epsilon_i$

  Model 2 : $aauw_i = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi_i + \beta_3 female_i + \beta_4 repub_i + \epsilon_i$

  Model 3 : $aauw_i = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi_i + \beta_3 female_i + \beta_4 repub_i + \beta_5 age_i + /beta_6 age2_i + \epsilon_i$

Code:

```
#install.packages("lfe")
library(lfe)
```

```
## Loading required package: Matrix
```

```
model1 <- felm(aauw ~ ngirls + totchi, data = basic2.2)
model2 <- felm(aauw ~ ngirls + totchi + female + repub, data = basic2.2)
model3 <- felm(aauw ~ ngirls + totchi + female + repub + age + age2, data = basic2.2)

stargazer(model1, model2, model3,
          type = "latex",
          title = "Question 5.1 Model Summary Table",
          se = list(model1$rse, model2$rse, model3$rse),
          header = FALSE
          )
```

Table 3: Question 5.1 Model Summary Table

	Dependent variable:		
	aauw		
	(1)	(2)	(3)
ngirls	5.776**	2.825**	2.899**
	(2.714)	(1.306)	(1.289)
totchi	−7.992***	−3.149***	−3.557***
	(1.784)	(0.964)	(0.964)
female		12.577***	12.064***
		(3.258)	(3.205)
repub		−71.783***	−71.286***
		(2.100)	(2.176)
age			0.814
			(0.971)
age2			−0.006
			(0.010)
Constant	59.982***	87.822***	63.184***
	(3.520)	(1.809)	(23.987)
Observations	434	434	434
R^2	0.051	0.796	0.798
Adjusted R^2	0.047	0.794	0.795
Residual Std. Error	41.010 (df = 431)	19.055 (df = 429)	19.023 (df = 427)
Note:			*p<0.1; **p<0.05; ***p<0.01

Question 5.2:

Interpret your estimate of β_1 from the first regression. Be sure to touch upon Sign, Size and Significance

Answer: The relationship between the number of daughters a congressperson has and their AAUW score appears to be statistically significant. Specifically, each additional daughter a congressperson has is correlated to a 5.776 point increase in their AAUW score (indicating a higher voting preference towards bills protecting AAUW interests).

Question 5.3:

How does age relate to the aauw score? At what age does the relationship between the aauw score and age "flip"? Is this relationship statistically significant?

Answer: The age of a congressperson is statistically significantly correlated with their AAUW score. Specifically, each one year increase in age is correlated with a $0.814 + (-0.006) = 0.808$ point increase in their AAUW score up until the age of 67.83 years when the relationship flips.

Question 5.4 (2 pages):

It is possible that the effects of having daughters might be different for female and male legislators. Estimate four different models to think about this question:

- Model A: Model 1
- Model B: Model 1 on women only
- Model C: Model 1 on men only
- Model D: Model 1 with the addition of female, female x ngirls, and female x totchi

Present these four regressions in a stargazer table. Is there evidence that the effect of a daughter differs for male and female legislators?

Code and Answers:

```
modelA <- model1
modelB <- felm(aauw ~ ngirls + totchi, data = basic2.2[basic2.2$female == 1,])
modelC <- felm(aauw ~ ngirls + totchi, data = basic2.2[basic2.2$female == 0,])
modelD <- felm(aauw ~ ngirls  + totchi + female + female*ngirls + female*totchi, data = basic2.2)


stargazer(modelA, modelB, modelC, modelD, type = "latex",
          title = "Question 5.4 Model Summary Table",
          header = FALSE, no.space = TRUE)
```

Table 4: Question 5.4 Model Summary Table

	Dependent variable:			
	aauw			
	(1)	(2)	(3)	(4)
ngirls	5.776**	3.043	5.071*	5.071*
	(2.567)	(8.119)	(2.648)	(2.635)
totchi	−7.992***	−5.428	−7.525***	−7.525***
	(1.753)	(6.607)	(1.781)	(1.773)
female				28.176**
				(12.748)
ngirls:female				−2.029
				(8.874)
totchi:female				2.097
				(7.120)
Constant	59.982***	84.532***	56.356***	56.356***
	(3.573)	(11.703)	(3.670)	(3.653)
Observations	434	48	386	434
R^2	0.051	0.018	0.052	0.103
Adjusted R^2	0.047	−0.026	0.047	0.092
Residual Std. Error	41.010 (df = 431)	38.347 (df = 45)	40.213 (df = 383)	40.021 (df = 428)
Note:				*p<0.1; **p<0.05; ***p<0.01

The magnitude of the coefficients in model B (female members only) was smaller than those in model C (male members only). Neither variable coefficient in model B are not significant at the 10% level whereas the ngirls variable coefficient is significant at the 10% level and the totchi variable coefficient is significant at the 1% level. Additionally, the variable coefficients for model C and model A are very similar. Together, these indicate that male congress members' voting decisions can be predicted to be more swayed by having an additional daughter than female congress members. We cannot make any direct assumptions about the

mechanism of effect but we can say that it appears that there is a larger effect of having a daughter for male congressmen than female congressmen.

Question 5.4:

How do the coefficients in models B and C relate to those in model D? Specifically, how can I calculate β_1 and β_2 from models B and C using the results in model D?

Answer: The coefficients for model B are smaller in magnitude than those in model D. They were also not significant at the 10% level. The coefficients in model C are relatively similar to those in model D. With the female variable, model D effectively can replicate both model B (only female members) and model C (only male members) by setting the value of the female variable. You can calculate β_1 and β_2 from model B in model D by putting 1 for female. You can calculate β_1 and β_2 from model C using the results in model D by putting a 0 for male.

Question 5.5

Lets reproduce the first set of columns in the top chart of figure 1:

- Filter your data so that it only includes representatives with two children
- use ggplot, with geom_bar to generate this plot with the NOW score on the vertical axis and the number of daughters on the horizontal axis.
- Hint: geom_bar(position = "dodge", stat = "summary", fun = "mean")

Make your graph as nice as possible!

Code:

```
bar_data <- basic %>%
  filter(totchi == 2, congress == 105) %>%
  select(party, nowtot, ngirls)

bar_data %>%
  ggplot(aes(x = ngirls, y = nowtot))+
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
           col = "black", fill = c("white", "grey", "black")) +
  theme_bw() +
  labs(title = "Figure 1: Mean NOW Score, by Number of Female Children, 105th Congress",
       subtitle = "Replication for ECON 1190") +
  xlab("Number of Female Children") +
  ylab("Mean NOW Score") +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.minor.y = element_line(),
        plot.title = element_text(face = "bold"),
        plot.subtitle = element_text(face = "italic"),
        panel.border = element_blank(),
        axis.line = element_line())
```

Figure 1: Mean NOW Score, by Number of Female Children, 105th Cong

Replication for ECON 1190