

Recreating Research: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues

NAME: _____

Empirical Analysis using Data from Washington (2008, AER)

This project 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.

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`, `lfe` 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 (removing the `#`) the following code. If you are working in R Studio Cloud these should load automatically or you will be prompted to load them.

Packages:

```
library(tinytex)
library(dplyr)
library(haven)
library(stargazer)
library(ggplot2)
library(lfe)
library(kableExtra)
library(RColorBrewer)
```

Load Data

```
df<-read_dta("C:\\Users\\roryq\\Downloads\\basic.dta")
```

How many observations are in the original dataset?

Code and Answer:

```
obs<-nrow(df); obs
```

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

there are 1740 observations

##Cleaning the data: 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.

Code:

```
df_105 <- df %>% filter(congress==105)
```

Reduce Data Frame:

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

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:

```
df_105<- df_105 %>% select(aauw,nowtot,totchi,ngirls,party,female,age)
```

Make sure your final dataset is a data frame.

Code:

```
df_105<-as.data.frame(df_105)
is(df_105)
```

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

Summary Statistics:

Report summary statistics for all the remaining variables in the dataset. Present these summary statistics in a formatted table.

Code:

```
stargazer(df_105,type="latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Oct 25, 2024 - 10:48:36 AM

Table 2:

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

Code:

```
df_105$repub <- ifelse(df_105$party==2,1,0)
```

Construct a variable called *age2*, where $\text{age2}=\text{age}^2$.

Code:

```
df_105$age2<-df_105$age^2
```

Analysis

Models from Page 315):

Estimate the following linear regression models. Report all three regression results in one formatted table.

$$\text{Model 1: } aaui = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi + \epsilon_i$$

$$\text{Model 2: } aaui = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi + \beta_3 female_i + \beta_4 repub_i + \epsilon_i$$

$$\text{Model 3: } aaui = \beta_0 + \beta_1 ngirls_i + \beta_2 totchi + \beta_3 female_i + \beta_4 repub_i + \beta_5 age_i + \beta_6 age_i^2 + \epsilon_i$$

Code:

```
model1<-felm(aaui~ngirls+totchi,data=df_105)

model2<- felm(aaui~ngirls+totchi+female+repub, data=df_105)

model3<- felm(aaui~ngirls+totchi+female+repub+age+age2, data=df_105)

stargazer(model1, model2,model3, se = list(model1$rse, model2$rse, model3$rse),type="latex",
title = "- Impact of Female Children on Legislator Voting on Women's Issues",
dep.var.caption = "",
covariate.labels = c("Number of Daughters", "Total Children","Female","Republican", "Age",
"Age squared","Constant"),
dep.var.labels = "AAUW")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Oct 25, 2024 - 10:48:37 AM

Table 3: - Impact of Female Children on Legislator Voting on Women's Issues

	AAUW		
	(1)	(2)	(3)
Number of Daughters	5.776** (2.714)	2.825** (1.306)	2.899** (1.289)
Total Children	-7.992*** (1.784)	-3.149*** (0.964)	-3.557*** (0.964)
Female		12.577*** (3.258)	12.064*** (3.205)
Republican		-71.783*** (2.100)	-71.286*** (2.176)
Age			0.814 (0.971)
Age squared			-0.006 (0.010)
Constant	59.982*** (3.520)	87.822*** (1.809)	63.184*** (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)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01	

Interpretation:

Interpret your estimate of β_1 from the first regression.

Answer:

From regression one, the coefficient implies that an increase in number of daughters by 1 (holding number of children constant) is associated with an increase in AAUW score of ~ 5.8 . This coefficient is statistically significant at a 95% confidence interval, this high significance indicates that the relationship is reliable and meaningful. Additionally the increase effect itself is large and even with robust standard errors, the lowest estimate would still see an increase of AAUW scores.

How does age relate to the aa uw score?

Answer:

The coefficient for age is 0.814, indicating that on average as age increases so does aa uw voting score. This relationship however is not statistically significant even at the 90% confidence level. The age squared term is negative, indicating that the effect of age on AAUW score decreases the older you get. The the effect ‘flips’ is ~ 67.8 years of age. Again the age squared term is not significant at the 90% confidence level.

Differences Between Male and Female Legislators):

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 of *female*, *female* \times *ngirls* and *female* \times *totchi*

Code and Answer:

```
df_mB<- df_105 %>% filter(female==1)
df_mC<- df_105 %>% filter(female==0)

model_A<- model1
model_B<- felm(aauw~ngirls+totchi,data=df_mB)
model_C<- felm(aauw~ngirls+totchi,data=df_mC)
model_D<-felm(aauw~ngirls+totchi+female+female*ngirls+female*totchi,data=df_105)

stargazer(model_A, model_B,model_C,model_D, se = list(model_A$rse, model_B$rse, model_C$rse
, model_D$rse),type="latex",
title = "- Impact of Female Children on Legislator Voting on Women's Issues",
dep.var.caption = "",
dep.var.labels = "AAUW")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Oct 25, 2024 - 10:48:37 AM

the interaction term with number of daughters is not significant indicating that there isn't a significantly different effect on voting behaviors between male and female legislators based on the number of daughters they have (holding all else constant). Additionally the interaction term of total children is not significant, indicating that there is not a significantly different effect on voting score between male and female legislators based on the total number of children (holding all else constant). In short no evidence that there is a difference, however legislators are a small population, so standard errors are quite high, maybe in a broader application there could be differences.

Table 4: - Impact of Female Children on Legislator Voting on Women's Issues

	AAUW			
	(1)	(2)	(3)	(4)
ngirls	5.776** (2.714)	3.043 (10.070)	5.071* (2.829)	5.071* (2.838)
totchi	-7.992*** (1.784)	-5.428 (6.360)	-7.525*** (1.845)	-7.525*** (1.850)
female				28.176*** (9.561)
ngirls:female				-2.029 (10.220)
totchi:female				2.097 (6.471)
Constant	59.982*** (3.520)	84.532*** (9.058)	56.356*** (3.650)	56.356*** (3.661)
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

Model Relationships:

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:

Model C and D have the same coefficients because there are no interaction terms. To find the coefficients for B from D, take the the number of daughters (ngirls) from model D and add the interaction term coefficient of female:ngirls to it. For example the coefficient for model B would be

$$\beta_{B1} = 5.071 + (-2.029) = 3.042$$

$$\beta_{B1} = ngirls_D + female : ngirls_D$$

. To find the coefficient for total children for model B you take the coefficient total children (totchi) from model D and add the interaction term totalchildren:Female to it. For example to find the model B coefficient,

$$\beta_{B2} = -7.525 + 2.097 = -5.428$$

$$\beta_{B2} = totchi_D + female : totchi_D$$

.

Graph Reproduction:

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.

Code:

```
# Filter the data
df_2 <- df_105 %>% filter(totchi == 2)
colors <- colorRampPalette( brewer.pal(n = 3, name = "Blues"))(15)[c(9,12,15)]

# Create the ggplot
ggplot(df_2, aes(x = ngirls, y = nowtot, fill = factor(ngirls))) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean", color = "black") +
  geom_text(stat = "summary", fun = "mean", aes(label = round(..y.., 1)),
           position = position_dodge(width = 0.9), vjust = 2, size = 7, col="white") +
  labs(
    title = "Average NOW Score by Number of Daughters",
    subtitle = "Of Legislators with 2 Total Children",
    x = "Number of Daughters",
    y = "NOW Score"
  ) +
  scale_fill_manual(values = colors) +
  scale_y_continuous(breaks = seq(0, max(df_2$nowtot) + 20, by = 20)) +
  theme_minimal(base_size = 15) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 26),
    plot.subtitle = element_text(hjust = 0.5, face = "italic", size = 24),
    axis.title.x = element_text(face = "italic", size = 23),
    axis.title.y = element_text(face = "italic", size = 23),
    panel.grid.major = element_line(color = "lightgrey"),
    panel.grid.minor = element_blank(),
    legend.position = "none",
    axis.text.x = element_text(size = 20),
    axis.text.y = element_text(size = 20)
  )
```

Average NOW Score by Number of Daughters
Of Legislators with 2 Total Children

