Assignment 3 - Due Friday, October 25 at 11:59 PM

EEP/IAS 118

Background

The data for this exercise was used in two papers, one paper by Ebonya Washington, "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues", published in the American Economic Review in 2008 that asks whether having daughters influences the voting behavior of members of the US Congress. The hypothesis is that having (more) daughters makes legislators more likely to vote liberally (in terms of political alignment, and in contrast to conservatively) on issues concerning women. The second paper by Donald Green, Oliver Hyman-Metzger, Gaurav Sood and Michelle Zee, "Revisiting a Natural Experiment: Do Legislators with Daughters Vote More Liberally on Women's Issues?" published in Journal of Political Economy: Microeconomics in 2023 revisits this argument, expanding the analysis to Congresses before and after those studied by Washington.

For this exercise, we will focus on votes that took place in the 96, 108th, 116th Congresses, which held session in 1983/1985, 2003/04 and 2019/2021. As a measure of a liberal voting record, we use scores assigned by the American Association of University Women (AAUW), a liberal group that concerns itself with issues of interest to women. For all Congresses, the AAUW selects a different number of pieces of legislation in the areas of education, economic equality, civil rights and reproductive rights. The AAUW then assigned a score to each member of Congress. The scores range from 0 to 100 and measure the percentage of times the legislator voted in favor of the position held by the AAUW.

The dataset `legislators_2024.dta` contains the following characteristics for all members for all Congresses between 1985 and 2021:

- $ngirls$ number of daughters
- $nboys$ number of sons
- $nchildren$ total number of children
- $born$ birth year
- $female$ indicator for being female
- $party$ party represented in Congress (D, R or NA)
- $aauw\_voting$ AAUW score
- $congress$ the Congress session for which scoring was assigned

- $year$ year the Congress started

1. (a) Prepare the data

For the purposes of this exercise, we assume all members of Congress included in the data were either Democrats or Republicans. They are also thought to be either male or female. Please drop members with no party affiliation (party marked as 'NA').

We also want to focus our analysis on three different Congress sessions (98th Congress, 108th Congress and 116th Congress). Please only keep data from these three sessions.

Load in the data set `legislators_2024.csv` .

```
In [1]:  # Write your code here
```

Does the sample seem to be split equally (or approximately equally) between men and women? Between Democrats and Republicans? What is the average age of the members in 2024?

Hint: use the `table()` command to see basic summary counts for variables in your data

```
In [2]:  # Write your code here
```

Type your answer here

1. (b) Estimate and report results for the following regression models:

Generate a variable `ngirls2` $= \text{ngirls}^2$

Generate a dummy variable for `repub` that takes the value of 1 if the party is 'R' (Republican) and 0 otherwise.

Generate an interaction variable `repubngirls` $= \text{repub} * \text{ngirls}$

Generate an interaction variable `repubngirls2` $= \text{repub} * \text{ngirls2}$

Estimate the following three regression models, save the output as reg1, reg2, and reg3, and show the results of each using `summary()` :

$$aauw\_voting = \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + u \tag{1}$$
$$aauw\_voting = \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + \beta_4 ngirls2 + \\ \beta_5 nchildren + u \tag{2}$$
$$aauw\_voting = \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + \beta_4 ngirls2 + \\ \beta_5 repubngirls + \beta_6 repubngirls2 + \beta_7 nchildren + u \tag{3}$$

(Note: this method of generating interaction variables (multiplying them together) is appropriate when one of the interacted variables is a dummy variable, but may not be appropriate in all cases.)

In [3]: `# Write your code here`

1. (c) Interpret the coefficients:

i. Focusing on reg1 only, what seems to be the effect of being a female Congress member on AAUW score? Do female Congress members lean more conservative than male Congress members?

ii. Looking at the first model, what is the effect of having a daughter on predicted AAUW score? Without further information and any computation, can we uncover the effect of having daughters on AAUW score from models 2 and 3? (The answer is no but tell us in a couple sentences why we cannot at this stage give a numeric answer to the effect of more daughters on AAUW score for these two models)

i. Write your answer here

ii. Write your answer here

2. (a) Suggest which model is the best fit to the data. How did you determine this?

Write your answer here

2. (b) Interpret the marginal effect at the mean of the number of daughters on AAUW score in each model.

(Hint: Calculate the total marginal effect, using the coefficients for all terms including the number of daughters. Calculating a marginal effect at the mean involves plugging in the mean number of daughters in the sample into any estimate of the total marginal effect of the number of daughters where this effect varies by the number of daughters.)

(Hint: Instead of typing in numbers from the regressions manually, you can call regression coefficients using summary(reg1)$coefficients[#,1] for coefficient

number # starting with the intercept as number 1.)

(Hint: In model 3, the marginal effect will differ by particular subgroups. Interpret the different effects for each subgroup.)

In [4]:
```
# Write your code here
```

Write your answer here

3. (a) Test whether there is an effect of the number of daughters on AAUW scores using the second model. Be sure to describe carefully the null and alternative hypothesis.

For this, do not use any canned package, we want you to compute the test statistic yourself.

Hint: You can access the residuals from a regression you have saved as reg by calling `reg$residuals` , or you can access the r-squared by calling `summary(reg)$r.squared` .

Write your hypotheses here

In [10]:
```
# Write your code here
```

Write your answer here

3. (b) Predict the AAUW score for male Democrats who have 3 daughters and 0 son, and who are 60 years old. Suggest a 95% CI for that predicted value.

Hint: See part 3-A of Section Notes 7. To answer this question, you will need to re-run a new regression similar to reg3 but include age as one of the explanatory variables

In [9]:
```
# Write your code here
```

Write your answer here

3. (c) Suppose a particular 60-year old male Democrat has 3 daughters and no son. Generate a 95% CI for his *particular* AAUW score, continuing to use the modified third model.

(Hint: See part 3-B of Section notes 7. Note that you can use most the values you already calculated in part 3 (b) to answer this question.)

```
# Write code here
```

Write your answer here

3. (d) What difference do you notice between the two CI asked in parts 3. (b) and 3. (c)? What could drive these differences? Please explain in no more than 2-3 sentences

Write your answer here

4. (a) Suppose you think Republicans and non-Republicans may have different gender patterns in voting with respect to the AAUW score. That is, Republican men may vote differently than Republican women, who may vote differently than Democratic women who may vote differently than Democratic men.

Write down an estimation equation you could use to test whether Republican women, Democratic women, and Democratic men each vote differently than Republican men. Specify what your null and alternative hypotheses would be.

Write your answer here

4. (b) Implement your test. Interpret each coefficient.

Hint: To create dummy variables based on particular characteristics, it is easiest to first create the dummy variable and set it equal to 0 for all observations: `data$dummy <- 0` . Then, replace the values for that dummy with 1 for the observations that match the requirements you are looking for, as in `data[data$x1==0 & data$x2==1,]$dummy<-1` .

Hint: If you need to, you can include an interaction term in your regression using `:` . For example `lm(y~x1+x2+x1:x2,data=data)` includes an interaction

between `x1` and `x2`. You will need to load the `car` package.

In [7]: `# Write your code here`

Write your answer here

4. (c) Adapt your regression to test whether Democratic women vote in the same way as Republican women with respect to the AAUW score. Write out the estimating equation and report your results. Conclude as to whether Democratic and Republican women vote similarly.

In [6]: `# Write your code here`

Write your answer here

Downloading and uploading your Notebook to Gradescope

Before downloading a copy, make sure to run all the cells above, including the 'Markdown'/text ones.

Download a PDF copy of your notebook by using **File > Save and Export notebook as PDF/Webpdf.**

When submitting on Gradescope, please make sure to link the right pages to the right questions, we will impose a penalty to students who not link pages to questions in their assignment.