Assignment 3 - Due Friday, October 21 at 11:59 PM

Background

The data for this exercise were used in Ebonya Washington's paper: "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues." published in the American Economic Review in 2008. The paper 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.

For this exercise, we will focus on votes that took place in the 108th Congress, which held session in 2003/04. 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 the 108th Congress, the AAUW selected 9 pieces of legislation in the areas of education, equality 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.dta` contains the following characteristics for a random sample of 386 members of the 108th Congress:

- $ngirls$ number of daughters
- $totchi$ total number of children
- age age
- $female$ indicator for being female
- $repub$ indicator for being a Republican
- $moredef$ proportion of people in the legislator's district who are in favor of "more spending on defense"
- $aauw$ AAUW score

(For the purposes of this exercise, you can assume all members of the 108th Congress were either Democrats or Republicans and were either male or female.)

(a) Estimate and report results for the following regression models:

Load in the data set `legislators.dta` . Remember, you will first need to call the `haven` package to do so.

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

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()` :

$$
\begin{aligned}
aauw &= \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + u & (1\\
aauw &= \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + \beta_4 ngirls2 + \beta_5 totchi + u\\
aauw &= \beta_0 + \beta_1 female + \beta_2 repub + \beta_3 ngirls + \beta_4 ngirls2 + \beta_5 repubngirls + \beta_6 repubngirl
\end{aligned}
$$

(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]:

```
library(haven)
library(tidyverse)

data <- read_dta('legislators.dta')
head(data)

#creating variable ngirls2
data <-mutate(data, ngirls2=ngirls^2)
#creating variable repubngirls
data <-mutate(data,repubngirl=repub*ngirls)
#creating variable repubngirls2
data <-mutate(data, repubngirls2=repub*ngirls2)
```

A tibble: 6 × 7

ngirls	totchi	repub	female	age	moredef	aauw
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	3	0	0	60	17.09234	75
1	1	1	0	37	31.40097	0
2	6	1	0	55	23.44828	0
2	2	0	0	45	16.47510	100
2	4	0	0	55	23.11688	100
2	5	1	0	55	31.40097	0

In [4]:

```
#reg1
reg1 <- lm(aauw~female+repub+ngirls, data=data)
summary(reg1)

#reg2
reg2 <- lm(aauw~female+repub+ngirls + ngirls2 + totchi, data=data)
summary(reg2)

#reg3
reg3 <- lm(aauw~female+repub+ngirls + ngirls2 + repubngirl + repubngirls2 + totch
summary(reg3)
```

```
Call:
lm(formula = aauw ~ female + repub + ngirls, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-86.215  -6.668  -5.976  13.439  56.024 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  86.5608     1.6251  53.266  < 2e-16 ***
female       11.4167     2.8473   4.010 7.31e-05 ***
repub       -79.5468     1.7993 -44.210  < 2e-16 ***
ngirls       -0.3460     0.7894  -0.438    0.661    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.4 on 382 degrees of freedom
Multiple R-squared:  0.8449,	Adjusted R-squared:  0.8437 
F-statistic: 693.9 on 3 and 382 DF,  p-value: < 2.2e-16



Call:
lm(formula = aauw ~ female + repub + ngirls + ngirls2 + totchi, 
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-88.508  -7.606  -1.361  11.839  54.859 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  88.1609     1.9844  44.428  < 2e-16 ***
female       11.3457     2.8444   3.989 7.97e-05 ***
repub       -78.8260     1.8076 -43.609  < 2e-16 ***
ngirls        2.6152     1.7157   1.524   0.1283    
ngirls2      -0.1932     0.3217  -0.601   0.5485    
totchi       -2.0752     0.8056  -2.576   0.0104 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.28 on 380 degrees of freedom
Multiple R-squared:  0.8478,	Adjusted R-squared:  0.8458 
F-statistic: 423.5 on 5 and 380 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = aauw ~ female + repub + ngirls + ngirls2 + repubngirl + 
    repubngirls2 + totchi + moredef, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-85.436  -7.964  -1.367  11.292  54.591 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   95.5997     3.6777  25.995  < 2e-16 ***
female        11.6079     2.8334   4.097 5.13e-05 ***
repub        -79.4364     3.0424 -26.110  < 2e-16 ***
ngirls         0.4452     3.1682   0.141   0.8883    
ngirls2        0.5286     0.8568   0.617   0.5376    
repubngirl     2.1281     3.6217   0.588   0.5571    
repubngirls2  -0.7477     0.9302  -0.804   0.4220    
totchi        -2.0364     0.8066  -2.525   0.0120 *  
moredef       -0.3166     0.1247  -2.540   0.0115 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.2 on 377 degrees of freedom
Multiple R-squared:  0.8505,	Adjusted R-squared:  0.8474 
F-statistic: 268.2 on 8 and 377 DF,  p-value: < 2.2e-16
```

(b) Suggest which model is the best fit to the data. How did you determine this? (no more than 1 sentence is required)

Model 3 has the best fit to the data because it has the highest adjusted R-Sqaured out of the three model, near 1.

(c) 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 [5]:

```
# confused on what this question wants --> is it simply the mean of ngirls???
mean_daughters <- mean(data$ngirls, na.rm = TRUE)
mean_daughters

model_1 <- summary(reg1)$coefficients[4,1]
model_1

model_2 <- (summary(reg2)$coefficients[4,1]) + (2*summary(reg2)$coefficients[5,1
model_2

# Democractic
model_demo_3 <- (summary(reg3)$ coefficients[4,1]) +
           (2*summary(reg3)$coefficients[5,1] * mean_daughters)
model_demo_3

# Republicans
model_repub_3 <- (summary(reg3)$coefficients[4,1]) + (summary(reg3)$coefficients
             (2*summary(reg3)$coefficients[5,1]* mean_daughters)
model_repub_3
```

```
1.21502590673575

-0.346022390451581

2.1457022456137

1.72973734506802

3.85785426879315
```

Mode1:

Model 1 tells us that for each additonal daughter you add it reduces the aauw percentage by -0.34 while holding other variables constant.

Model2:

Model 2 tells us that for each additional daughter added incrases the aauw percentage by 2.15 while holding other variables constant.

Model 3:

Model 3 tell us that for democrats for each added additional unit daughters, the aauw score increases by 1.729 holding everything else cosntant while republicans for each additional unit of daughters added gives us an increase of 3.857 while holding everything constant.

(d) 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.

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

Unrestricted Model: aauw~female + repub + ngirls + ngirls2 + totchi

Restriticed model: aauw~female + repub + totchi

Null hypthoesis is H0: beta3,beta4 = 0

alternative hpyothesis is H1: not H0

In [6]:

```r
restricted_model <- lm(aauw~female + repub +totchi,data=data)

SSR_U <- sum(reg2$residuals^2)
SSR_R <-sum(restricted_model$residuals^2)
n <- nobs(reg2)
k <- 5 ## ask the gsi what this value is. -->
q <- 2 ## ask the gsi what this value is --> coefiencents difference between rist

top <- (SSR_R-SSR_U)/q
bottom <- SSR_U/(n-k-1)
f<- top / bottom
f
n-k-1

## Alternative Formula
R2_Ur<-summary(reg2)$r.squared
R2_R<-summary(restricted_model)$r.squared
F_2<-((R2_Ur-R2_R)/q)/((1-R2_Ur)/(n-k-1))
F_2
```

```
1.47735255535295

380

1.47735255535288
```

Our critical value will be 3.019, our F-statistics value was 1.477 so we reject the null hypoythesis. This tell us that there is an effect on the number of daughters when it comes to the aauw score, so we would favor the alternative.

(e) Using the third model, predict the AAUW score for male democrats who have 3 daughters and 0 son, and who have 36% of constituents who want more spending on defense, on average. Suggest a 95% CI for that predicted value.

(Hint: See part 3-A of Section Notes 8.)

In [7]:

```
data<-mutate(data,ngirls0=ngirls-3)
data<-mutate(data,ngirls20=ngirls2-9)
data<-mutate(data,totchi0=totchi-3)
data<-mutate(data,moredef0=moredef-36)

reganswer <- summary(lm(aauw~ female + repub + ngirls0 + ngirls20 +
            repubngirl + repubngirls2 +  totchi0 + moredef0 , data=data))
reganswer
```

```
Call:
lm(formula = aauw ~ female + repub + ngirls0 + ngirls20 + repubngirl + 
    repubngirls2 + totchi0 + moredef0, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-85.436  -7.964  -1.367  11.292  54.591 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   84.1854     3.3010  25.503  < 2e-16 ***
female        11.6079     2.8334   4.097 5.13e-05 ***
repub        -79.4364     3.0424 -26.110  < 2e-16 ***
ngirls0        0.4452     3.1682   0.141   0.8883    
ngirls20       0.5286     0.8568   0.617   0.5376    
repubngirl     2.1281     3.6217   0.588   0.5571    
repubngirls2  -0.7477     0.9302  -0.804   0.4220    
totchi0       -2.0364     0.8066  -2.525   0.0120 *  
moredef0      -0.3166     0.1247  -2.540   0.0115 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.2 on 377 degrees of freedom
Multiple R-squared:  0.8505,	Adjusted R-squared:  0.8474 
F-statistic: 268.2 on 8 and 377 DF,  p-value: < 2.2e-16
```

In [8]:

```
ci_left <- (84.1854 -  1.96*(3.3010))
ci_right <- (84.1854 +  1.96*(3.3010))
c(ci_left,ci_right)
```

```
77.71544  90.65536
```

Using the third model, predicting the aauw score for 3 daughters, and 0 sons, who have 36% consituents who want more spending on defense on average would predict a value between the interval 77.71 and 90.65 with a confidence of 95%.

(f) Suppose a particular male Democrat has 3 daughters and no son in a state where 36% of constituents want more spending on defense. Generate a 95% CI for his *particular* AAUW score, continuing to use the third model.

(Hint: See part 3-B of Section Notes 8. Note that you can use most the values you already calculated in part (e) to answer this question.)

In [12]:
```
summary(lm(aauw~female+repub+ngirls +
           ngirls2 + repubngirl + repubngirls2
           + totchi + moredef , data=data))$sigma^2

standard_error <- sqrt((3.3010^2) + 295.671049048731)
standard_error

ci_left <- (84.1854 -  1.96*(17.5090733635087))
ci_right <- (84.1854 +  1.96*(17.5090733635087))
c(ci_left,ci_right)
```

```
295.671049048731

17.5090733635087

49.8676162075229   118.503183792477
```

Using the third model, predicting the aauw score for 3 daughters, and 0 sons, who have 36% consituents who want more spending on defense on average would predict a value between the interval 49.8676 and 118.50 with a confidence of 95%.

(g) 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.

Regression model:

aauw = beta0 + beta1(republicanwomen) + beta2(democraticwomen) + beta3(democraticemen) + u

Our null hpypothesis is that republican and non-republic vote differently based on their gender patterns of their AAUW score. That republic men vote different then republican women and who vote differnetly then democractic women and who vote differently than democractic men.

h0: beta1 = 0

h1: beta1 ≠ 0 not equal to zero

h0: beta2 = 0

h1: beta2 ≠ 0 not equal to zero

h0: beta3 = 0

h1: beta3 ≠ 0 not equal to zero

(h) Implement your test. Interpret each coeffcient.

(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 [14]:

```
data$republicanwomen <-0
data[data$repub == 1 & data$female == 1,]$republicanwomen <-1

data$democraticwomen <-0
data[data$repub == 0 & data$female == 1,]$democraticwomen <-1

data$democraticmen <-0
data[data$repub == 0 & data$female == 0,]$democraticmen <-1
head(data)
```

A tibble: 6 × 18

ngirls	totchi	repub	female	age	moredef	aauw	ngirls2	repubngirl	repubngirls2	ngirls0	ng
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	3	0	0	60	17.09234	75	1	0	0	-2	
1	1	1	0	37	31.40097	0	1	1	1	-2	
2	6	1	0	55	23.44828	0	4	2	4	-1	
2	2	0	0	45	16.47510	100	4	0	0	-1	
2	4	0	0	55	23.11688	100	4	0	0	-1	
2	5	1	0	55	31.40097	0	4	2	4	-1	

In [15]:
```
reg4<-lm(aauw~republicanwomen+democraticwomen+democraticmen, data=data)
summary(reg4)
```

```

Call:
lm(formula = aauw ~ republicanwomen + democraticwomen + democraticmen, 
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-86.586  -6.246  -6.246  13.414  55.754 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)        6.246      1.257    4.97 1.01e-06 ***
republicanwomen   16.111      4.809    3.35 0.000889 ***
democraticwomen   89.168      3.462   25.76  < 2e-16 ***
democraticmen     80.339      1.888   42.55  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.37 on 382 degrees of freedom
Multiple R-squared:  0.8455,    Adjusted R-squared:  0.8443 
F-statistic: 696.7 on 3 and 382 DF,  p-value: < 2.2e-16
```

In model4, beta0 would be 6.246 aauw points while holding every other variable constant. Beta1, republicanwomen, each one added unit of republicanwomen would increase the aauw score by 16.11 while holding other variables constant. Beta2, for everyone unit of democratic women added the aauw score will increase by 89.16 points. Beta3, for every one unit of democraticmen will icnrease the aauw score by 80.339. This shows us there is a difference between reoublic women and demoractic women when it comes to their affects on the aauw score.

(i) 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 [16]:
```
### getting the difference between republic and demoratic women
data$republicanman<-0
data[data$repub==1 & data$female==0,]$republicanman<-1
reg5<-lm(aauw~ republicanman + democraticwomen + democraticmen, data=data)
```

```
In [17]: summary(reg5)
```

```
Call:
lm(formula = aauw ~ republicanman + democraticwomen + democraticmen, 
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-86.586  -6.246  -6.246  13.414  55.754 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       22.357      4.642   4.816 2.11e-06 ***
republicanman    -16.111      4.809  -3.350 0.000889 ***
democraticwomen   73.057      5.653  12.924  < 2e-16 ***
democraticmen     64.228      4.851  13.239  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.37 on 382 degrees of freedom
Multiple R-squared:  0.8455,    Adjusted R-squared:  0.8443 
F-statistic: 696.7 on 3 and 382 DF,  p-value: < 2.2e-16
```

We reject the null hypthoesis because all the coeficients are starred with statistically sginficance. In model5, beta0 would be 22.357 aauw points while holding every other variable constant. Beta1, republicanmen, one each added unit of republicanmen would decrease the aauw score by -16.11 while holding other variables constant. Beta2, for everyone unit of democratic women added the aauw score will increase by 73.057 points. Beta3, for every one unit of democraticmen will icnrease the aauw score by 64.228. This model from (i) and (j) showcase there is an ommitted variable and difference between the betas and regression model values hence we reject the null hypothesis (the value is not 0). We also reject that there is difference between democractic women and republican women when it coems to voting.

Downloading your Notebook

Download a PDF copy of your notebook by using **File > Download as HTML.** Alternatively, go to print preview and **Save as PDF** (but make sure your code does not get cut off at the side of the page!)