Problem Set 2

Applied Stats II

Due: February 18, 2024

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub in `.pdf` form.
- This problem set is due before 23:59 on Sunday February 18, 2024. No late assignments will be accepted.

We're interested in what types of international environmental agreements or policies people support (Bechtel and Scheve 2013). So, we asked 8,500 individuals whether they support a given policy, and for each participant, we vary the (1) number of countries that participate in the international agreement and (2) sanctions for not following the agreement.

Load in the data labeled `climateSupport.RData` on GitHub, which contains an observational study of 8,500 observations.

- Response variable:
  - `choice`: 1 if the individual agreed with the policy; 0 if the individual did not support the policy
- Explanatory variables:
  - `countries`: Number of participating countries [20 of 192; 80 of 192; 160 of 192]
  - `sanctions`: Sanctions for missing emission reduction targets [None, 5%, 15%, and 20% of the monthly household costs given 2% GDP growth]

Please answer the following questions:

1. Remember, we are interested in predicting the likelihood of an individual supporting a policy based on the number of countries participating and the possible sanctions for non-compliance.

   Fit an additive model. Provide the summary output, the global null hypothesis, and p-value. Please describe the results and provide a conclusion.

   First I relevel and change the reference catagory to "None" for variable "sanctions" and "20 out of 192" for variable "country".

```r
data = climateSupport
View(data)

# Re order the levels and change reference category
data$countries <- as.character(data$countries)
data$countries <-  c("20 of 192", "160 of 192", "80 of 192"))
data$countries <- factor(data$countries)
data$countries <- relevel(data$countries, ref = "20 of 192")

# Do it again for sanctions
data$sanctions <- as.character(data$sanctions)
data$sanctions <-  c("None", "5%", "15%", "20%")
data$sanctions <- factor(data$sanctions)
data$sanctions <- relevel(data$sanctions, ref = "None")
```

   Now to create an additive model, the coefficient table is listed below

```r
add_mod <- glm(choice ~ countries + sanctions, family = binomial(link = "
    logit"), data = data)
summary(add_mod)
```

   NULL HYPOTHESIS
   Ho: There is no statistically significant association between either of the two independent variables (number of countries who participate and sanctions).
   We can reject the Null Hypothesis because some variables have a statistically significant corrolation (see below).

   COEFFICIENTS and P-VALUES

   (Intercept) -0.26597: When sanctions are "None" and 20 countries participate, the estimated log odds that the person will support the policy are -0.26597. The expected odds of a person supporting the legislature is $\exp(-0.005665) = 0.7664621$ (baseline odds ratio). This value has high statistical significance with a p-value of 6.08×10^{-7}.

   Countries160 of 192 0.64568: When sanctions are "None" and there is a change from 20 to 160 countries participating, the log odds of a person supporting the legislation increase by 0.64568. This value has high statistical significance with a p-value of $< 2 \times 10^{-16}$.

Countries80 of 192 0.32106: When sanctions are "None" and there is a change from 20 to 80 countries participating, the log odds of a person supporting the legislation increase by 0.32106. This value has high statistical significance with a p-value of 2.03×10^{-09}.

Sanctions20% -0.12203: When 20 countries participate, and sanctions increase from 0% to 20%, the log odds of a person supporting the legislation decrease by -0.12203. This value is statistically significant with a p-value of 0.0488.

Sanctions15% -0.02222: When 20 countries participate, and sanctions increase from 0% to 20%, the log odds of a person supporting the legislation decrease by -0.02222. This value is not statistically significant with a p-value of 0.7197.

Sanctions5% -0.10383: When 20 countries participate, and sanctions increase from 0% to 5%, the log odds of a person supporting the legislation decrease by -0.10383. This value is generally not considered statistically significally significant with a p-value of 0.0936 (some may see this as significant depeding on alpha).

Table 1:

	Dependent variable:
	choice
countries160 of 192	0.646***
	(0.054)
countries80 of 192	0.321***
	(0.054)
sanctions5%	$-$0.104*
	(0.062)
sanctions15%	$-$0.022
	(0.062)
sanctions20%	$-$0.122**
	(0.062)
Constant	$-$0.266***
	(0.053)
Observations	8,500
Log Likelihood	$-$5,815.530
Akaike Inf. Crit.	11,643.060
Note:	*p<0.1; **p<0.05; ***p<0.01

2. If any of the explanatory variables are significant in this model, then:

   (a) For the policy in which nearly all countries participate [160 of 192], how does increasing sanctions from 5% to 15% change the odds that an individual will support the policy? (Interpretation of a coefficient)

   First I change the reference catagorys to make "5 %" and "160 of 192".

```r
# Change the reference category
data$countries <- relevel(data$countries, ref = "160 of 192")
# Do it again for sanctions
data$sanctions <- relevel(data$sanctions, ref = "5%")
# Run the model again
add_mod <- glm(choice ~ countries + sanctions, family = binomial(link
    = "logit"), data = data)
summary(add_mod)
```

   When the number of participating countries is 160 and the sanctions change from 5% to 15%, the estimated change in log odds that a person will support the policy is 0.08161. This equates to an odds ratio of $\exp(0.08161) = 1.085033$. This means that a change from 5% to 15% in sanctions is estimated to increase the odds that a person will support the policy by approximately 8.5%. $1.085033 - 1 = 0.085$ and $0.085 \times 100 = 8.5\%$. However, this value is not statistically significant with a p-value of 0.1874. (See Table 2 Below)

Table 2:

	Dependent variable:
	choice
countries160 of 192	0.646***
	(0.054)
countries80 of 192	0.321***
	(0.054)
sanctionsNone	0.104*
	(0.062)
sanctions15%	0.082
	(0.062)
sanctions20%	−0.018
	(0.062)
Constant	−0.370***
	(0.054)
Observations	8,500
Log Likelihood	−5,815.530
Akaike Inf. Crit.	11,643.060
Note:	*p<0.1; **p<0.05; ***p<0.01

(b) What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions?

First I change the reference catagorys to make "None" and "80 of 192" respectivley.

```r
# Change the reference category
data$countries <- relevel(data$countries, ref = "80 of 192")
# Do it again for sanctions
data$sanctions <- relevel(data$sanctions, ref = "None")
# Run the model again
add_mod <- glm(choice ~ countries + sanctions, family = binomial(link
    = "logit"), data = data)
summary(add_mod)
```

When 80 countries participate and there are no sanctions, the estimated expected odds of a person supporting the legislature is $\exp(0.05509) = 1.056636$ (baseline odds ratio). To find the probability, I use the formula $\frac{1.056636}{1+1.056636} * 100$. The estimated probability is approximately 51%. This value is not statistically significant with a p-value of 0.3069. (See Table 3 Below)

Table 3:

	Dependent variable:
	choice
countries160 of 192	0.325***
	(0.054)
countries20 of 192	−0.321***
	(0.054)
sanctions5%	−0.104*
	(0.062)
sanctions15%	−0.022
	(0.062)
sanctions20%	−0.122**
	(0.062)
Constant	0.055
	(0.054)
Observations	8,500
Log Likelihood	−5,815.530
Akaike Inf. Crit.	11,643.060
Note:	*p<0.1; **p<0.05; ***p<0.01

(c) Would the answers to 2a and 2b potentially change if we included the interaction term in this model? Why?

- Perform a test to see if including an interaction is appropriate.

  As can be seen none of the interaction terms below are statistically significant so while the answer would change the interaction should not be added into the model.

```
compare <- anova(add_mod, int_mod , test = "LRT")
print(compare)
```

Table 4:

	Dependent variable:
	choice
countries160 of 192	0.688***
	(0.107)
countries80 of 192	0.357***
	(0.107)
sanctions5%	−0.025
	(0.106)
sanctions15%	−0.020
	(0.107)
sanctions20%	−0.100
	(0.107)
countries160 of 192:sanctions5%	−0.159
	(0.151)
countries80 of 192:sanctions5%	−0.078
	(0.152)
countries160 of 192:sanctions15%	−0.002
	(0.152)
countries80 of 192:sanctions15%	−0.006
	(0.152)
countries160 of 192:sanctions20%	−0.006
	(0.152)
countries80 of 192:sanctions20%	−0.060
	(0.151)
Constant	−0.292***
	(0.075)
Observations	8,500
Log Likelihood	−5,814.592
Akaike Inf. Crit.	11,653.180
Note:	*p<0.1; **p<0.05; ***p<0.01

I run an ANOVA test with the additive and interactive model using the original reference catagories of "none" and "20 of 192"

```
compare <- anova(add_mod, int_mod , test = "LRT")
print(compare)
```

The p-value is not statistically significant so we should not use the interactive model.

Table 5:

	Dependent variable:
	choice
countries160 of 192	0.646***
	(0.054)
countries80 of 192	0.321***
	(0.054)
sanctionsNone	0.104*
	(0.062)
sanctions15%	0.082
	(0.062)
sanctions20%	−0.018
	(0.062)
Constant	−0.370***
	(0.054)
Observations	8,500
Log Likelihood	−5,815.530
Akaike Inf. Crit.	11,643.060
Note:	*p<0.1; **p<0.05; ***p<0.01