Problem Set 2

Applied Stats II

Due: February 19, 2023

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 19, 2023. 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.csv` 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.

Table 1: Additive Model

	Dependent variable:
	choice
countries.L	0.458^{***}
	(0.038)
countries.Q	-0.010
	(0.038)
sanctions.L	-0.276^{***}
	(0.044)
sanctions.Q	-0.181^{***}
	(0.044)
sanctions.C	0.150^{***}
	(0.044)
Constant	-0.006
	(0.022)
Observations	8,500
Log Likelihood	$-5{,}784.130$
Akaike Inf. Crit.	11,580.260
Note:	*p<0.1; **p<0.05; ***p<0.01

```
logis_1 <- glm(climateSupport$choice ~ climateSupport$countries +
  climateSupport$sanctions,
family = "binomial")
```

Null hypothesis for the additive model: There is no relationship between support for climate agreements and any component the sanctions and numbers of participating countries variables

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,496.500	3.536	8,494	8,499
Resid. Dev	2	11,675.830	152.134	11,568.260	11,783.410
Df	1	5.000		5	5
Deviance	1	215.150		215.150	215.150
Pr(>Chi)	1	0.000		0	0

```r
# Null model
null_logis_1 <- glm(climateSupport$choice ~ 1, family = "binomial")

# Likelihood Ratio Test
anova(null_logis_1, logis_1, test = 'LRT')
```

P value ¡.001 (p = 2.2e-16) suggests there is a significant difference between the two models and we can reject the null hypothesis that there is no relationship between support for climate agreements and any category of sanctions or no. of participating countries.

```r
 # Coverting log odds to probs for interpretation
 (exp(logis_1$coefficients)) / (1 + (exp(logis_1$coefficients)))
```

Short Analysis: Each coefficient has a significant reletionship with climate support, with the exception of the category for 160-190 countries particpating. Converting the log odds to probabilities, coefficients for 80 countries partipating and 20% sanctions were greater than the intercept, representing increased the likelihood of support for the agreement when compared to baseline of support with 20 countries and no sanctions.

Of the other significant categories, the probabilities for each is bellow the intercept, meaning there effect was reduce the liklihood of support when compared to the baseline of 20 countries and no sanctions.

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)

```
# Trying to do probability for Y1 = odds of it occurring over the
    odds of it not, not sure
# though. Formula adds intercept and all coefficients except one for
    80 countries.
# My thinking was the probability of it not occurring would be equal
    to the cumulative probability
# of the other events occurring
# Probability of support when sanctions 5% and countries 160
1 / 1+exp(-(-.005665 + -0.458452 + -0.181086 + -0.150207)) # 3.2154
# Probability of support when sanctions 15% and countries 160
1 / 1+exp(-(-.005665 + -0.458452 + -0.276332 + -0.150207))

3.436 - 3.215 # = 0.221
```

   Holding participating countries at 160, changing sanctions from 5% to 15% increases the odds of support slightly from 3.2153 to 3.436.

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

```
    # Same thinking again
    1 / 1+exp(-(-.005665 + -0.009950 + -0.276332 + -0.181086 +
    -0.150207)) # = 2.864
```

   When sanctions are none and participating countries are 80, individuals are 2.8 times more likely to support climate agreement compared other combinations

   (c) Would the answers to 2a and 2b potentially change if we included the interaction term in this model? Why? The effect of both sanctions and no. of supporting n support countries may be greater than the sum effect each.

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

	Dependent variable:	
	choice	
	(1)	(2)
countries.L	0.458***	0.457***
	(0.038)	(0.038)
countries.Q	−0.010	−0.011
	(0.038)	(0.038)
sanctions.L	−0.276***	−0.274***
	(0.044)	(0.044)
sanctions.Q	−0.181***	−0.182***
	(0.044)	(0.044)
sanctions.C	0.150***	0.153***
	(0.044)	(0.044)
sanctions.L		−0.002
		(0.077)
sanctions.L		0.134*
		(0.076)
sanctions.Q		−0.008
		(0.076)
sanctions.Q		0.093
		(0.076)
sanctions.C		0.095
		(0.076)
sanctions.C		0.010
		(0.077)
Constant	−0.006	−0.004
	(0.022)	(0.022)
Observations	8,500	8,500
Log Likelihood	−5,784.130	−5,780.983
Akaike Inf. Crit.	11,580.260	11,585.970
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
# Interactive Model
logis_2 <- glm(climateSupport$choice ~ climateSupport$countries +
    climateSupport$sanctions
+ climateSupport$countries*climateSupport$sanctions,
family = binomial(link = "logit"))
summary(logis_2)

# Likelihood Ratio Test for difference between two models
anova(logis_2, logis_1, test = 'LRT')
"P value > .0001, meaning there is not a significant difference between
    between the models"
```

P value>.001, meaning there is not a significant difference between the additive and interactive models