Problem Set 2

Daniel Murray (13303981)

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.

   Answer:

   A logit model was fitted, with "choice" regressed on all other variables, and summary output produced using the below code.

   The sanctions variable was transformed from an ordered factor to an unordered factor so that factor levels could be reordered with "5%" as the reference, to aid with interpreting coefficients for Q2.

```r
# Change sanctions variable to unordered factor and re-order to set
    5% as reference
climateSupport$sanctions <- factor(climateSupport$sanctions,
                                   ordered=FALSE)
climateSupport$sanctions <- relevel(climateSupport$sanctions,
                                    ref="5%")

# Run logit model with choice regressed on all other variables
logit_mod <- glm(choice ~ .,
                 data=climateSupport,
                 family=binomial(link="logit"))

# Check summary output of model
summary(logit_mod)
stargazer(logit_mod, type="latex")
```

   Summary output:

Table 1: Summary output of logit model

	Dependent variable:
	choice
countries.L	0.458^{***}
	(0.038)
countries.Q	-0.010
	(0.038)
sanctionsNone	-0.192^{***}
	(0.062)
sanctions15%	-0.325^{***}
	(0.062)
sanctions20%	-0.495^{***}
	(0.062)
Constant	0.247^{***}
	(0.044)
Observations	8,500
Log Likelihood	$-5{,}784.130$
Akaike Inf. Crit.	11,580.260
Note:	*p<0.1; **p<0.05; ***p<0.01

Global null hypothesis: coefficients for all predictor variables are equal to zero. No values for either number of countries or level of sanctions for non-compliance are statistically differentiable from zero.

We can test this global null hypothesis by carrying out a likelihood ratio test. First a null model was created and then an anova test was carried out comparing our model to the null model, using the below code:

```
# Create null model
null_mod <- glm(choice ~ 1,
                data=climateSupport,
                family=binomial(link="logit"))

# Run anova test comparing our model to the null model
```

```r
anova(null_mod, logit_mod, test = "LRT")
```

The p-value from the anova test is 2.2e-16. As this value is below our threshold of α=0.05, we can reject the null hypothesis. Our model provides a better explanation of the variation in our outcome variable than the null model.

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)

   Answer:

   Increasing the level of sanctions from 5% to 15%, on average, leads to a 0.325 reduction in the log odds that an individual will support the policy. This equates to a reduction in the odds by a multiplicative factor of 0.722. This would hold true for all levels of country participation in the policy, as our model is additive.

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

   Answer:

   The probability was calculated using the code below and was found to be 0.63. Values of the coefficients were taken directly from the summary output.

```
# Convert coefficients of interest from log odds to odds
odds <- exp(0.247+0.458-0.192)
odds

# Calculate probability from odds
prob <- odds/(1+odds)
round(prob,2)
```

   (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.

   Answer:

   The answers might change as there would be different lines with potentially differing intercepts and slopes for each combination of sanction level and number of participating countries.

   A likelihood ratio test was performed to see if including the interaction is appropriate. An interactive model was first fit, and then an anova test was run:

```r
# Run logit model with interaction terms
logit_mod_inter <- glm(choice ~ countries*sanctions,
                       data=climateSupport,
                       family=binomial(link="logit"))

# Run anova test comparing first model to the interactive model
anova(logit_mod, logit_mod_inter, test = "LRT")
```

Figure 1: Output of anova test

```
Analysis of Deviance Table

Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries * sanctions
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8494      11568
2      8488      11562  6   6.2928   0.3912
```

As the p-value from the anova test (0.3912) is above our threshold of α=0.05, we fail to reject the null hypothesis that the interactive model does not provide a better fit. Therefore, we do not have enough evidence to suggest that we should add an interactive term.