Problem Set 2

Cianna Devitt (17321885)

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.

   Coercing 'choice' column to 'Supported' = 1, 'Not Supported'= 0

```
    climateSupport$choice <- as.numeric(ifelse(climateSupport$choice == '
   Supported', 1,0))

```

   Creating an additive model with glm():

```
model <- glm(choice ~ sanctions + countries,
             family = binomial(link = 'logit'),
             data = climateSupport)
summary(model)

>summary(model)


Deviance Residuals:
Min       1Q   Median       3Q      Max
-1.4259  -1.1480  -0.9444   1.1505   1.4298

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.005665   0.021971  -0.258 0.796517
sanctions.L -0.276332   0.043925  -6.291 3.15e-10 ***
sanctions.Q -0.181086   0.043963  -4.119 3.80e-05 ***
sanctions.C  0.150207   0.043992   3.414 0.000639 ***
countries.L  0.458452   0.038101  12.033  < 2e-16 ***
countries.Q -0.009950   0.038056  -0.261 0.793741

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11783  on 8499  degrees of freedom
Residual deviance: 11568  on 8494  degrees of freedom
AIC: 11580

Number of Fisher Scoring iterations: 4
```

   Global Null Hypothesis: H_o all slopes = 0
   H_a $\beta_j \neq 0$

```
  Null deviance: 11783  on 8499  degrees of freedom
  Residual deviance: 11568  on 8494  degrees of freedom
```

```
3
4  # X^2 = Null deviance - Residual deviance
5  # p-value:
6  1-pchisq(11783-11568, 8499-8494)
7  [1] 0
```

With a p-value of 0, we can reject null hypothesis that all of the slopes in the model are equal to zero. At least one variable in model is a significant predictor.

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)
   A unit increase in sanctions variable (5percent to 15percent) is associated with a 0.15 increase in log odds of outcome variable support for policy, holding all other variables constant.

   (b) What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions?
   $P(Y_i = 1|X_i) = \frac{exp(\beta_o+\beta_1 X_1)}{1+exp(\beta_o+\beta_1 X_1)} = \frac{1}{1+exp[-(\beta_o+\beta_1 X_1)]} = \frac{1}{1+exp[-(-0.005)+(-0.27)(\frac{80}{192})]}$
   $= 1.898077$

   (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.

   Liklihood ratio test: Comparing reduced model and interaction model with ANOVA()

```
1  model_int <- glm(choice ~ sanctions * countries,
2                   family = binomial(link = 'logit'),
3                   data = climateSupport)
4  anova(model_null, model_int, test = 'LRT')
5  Analysis of Deviance Table
6
7  Model 1: choice ~ 1
8  Model 2: choice ~ sanctions * countries
9  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
10 1      8499      11783
11 2      8488      11562 11   221.44 < 2.2e-16 ***
12 ---
13
```

   As p-value is ¡ 0.01, we can conclude at least one predictor in interaction model is reliable