

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.

My solutions to Question 1

```
1 # Converting the choice variables to dummy variables
2 climateSupport$choice <- ifelse(climateSupport$choice == 'Not
  supported', 0, 1)
3
4 # Converting the countries and sanctions to numeric variables
5 climateSupport$countries <- as.numeric(climateSupport$countries)
6 climateSupport$sanctions <- as.numeric(climateSupport$sanctions)
7
8 # Fitting the additive model
9 climate_additive_model <- glm(choice ~ countries + sanctions, data =
  climateSupport,
10 family=binomial(logit))
11
12 # Printing the summary output
13 summary(climate_additive_model)
```

```

Call:
glm(formula = choice ~ countries + sanctions, family = binomial(logit),
    data = climateSupport)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3976  -1.1490  -0.9679   1.1536   1.4025

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.34540    0.07539  -4.581 4.62e-06 ***
countries      0.32436    0.02689  12.062 < 2e-16 ***
sanctions    -0.12353    0.01964  -6.291 3.15e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11783  on 8499  degrees of freedom
Residual deviance: 11597  on 8497  degrees of freedom
AIC: 11603

Number of Fisher Scoring iterations: 4

```

The p-values for both the countries (<0.00000000000000002) and sanctions (0.0000000000315) variables are both significant at the 99.99 % level (i.e., less than 0.000005), indicating we can reject the global null hypothesis (i.e., the countries and sanctions variables have no predictive power for the choice variable). This p-value suggests that were we to conduct this test 10,000 times, we'd get the same result 9,999 times.

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) Holding countries constant, sanctions increasing from 5% to 15% increases the likelihood an individual will support a policy by approximately 11.6%. We figure this out by exponentiating the log-odds coefficient for the sanctions variable (-0.12353). I.e., the following R code:

```
1 exp(-0.12353)
```

Exponentiating $-0.12353 = 0.884$. We then calculate the log-odds ratio as follows:
 $100 \times (\text{regression coefficient} - 1)$

$100 \times (0.884 - 1) = -11.6\%$.

This means for a one-unit increase - in this instance, sanctions increasing from 5% to 15% - the odds of the choice variable decrease by a factor of 0.884 or 11.6%.

- (b) What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions? This gives us the regression coefficient of $y = -0.34540 + 0.32436(\text{countries} \times 2) + -0.12353(\text{sanctions} \times 0)$ We calculate this using the following R code:

```
1 Q2_coefficient <- -0.34540 + (0.32436 * 2) + (-0.12353 * 0)
2 Q2_odds_ratio <- exp(Q2_coefficient)
3 100*(Q2_odds_ratio - 1)
```

This results in a 35.4% probability an individual will support a policy if 80 of 192 countries participate if sanctions aren't stipulated.

- (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.

My solutions to Question 2(c)

```
1 # Creating the interaction model
2 climate_interaction_model <- glm(choice ~ countries + sanctions +
3   countries*sanctions, data = climateSupport,
4   family=binomial(logit))
5
6 # Print the interaction model summary
summary(climate_interaction_model)
```

```

Call:
glm(formula = choice ~ countries + sanctions, family = binomial(logit),
     data = climateSupport)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.3976  -1.1490  -0.9679   1.1536   1.4025

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.34540    0.07539  -4.581 4.62e-06 ***
countries      0.32436    0.02689  12.062 < 2e-16 ***
sanctions    -0.12353    0.01964  -6.291 3.15e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11783  on 8499  degrees of freedom
Residual deviance: 11597  on 8497  degrees of freedom
AIC: 11603

Number of Fisher Scoring iterations: 4

```

According to the interactive model, the sanctions variable p-value (0.0228) is much higher than in the additive model (0.000000000315). However, it's still significant at the 95% level.

The countries variable p-value (0.000000582) is higher than in the additive model (<0.0000000000000002). However, it's still significant at the 99.99% level.

Most relevant for our purposes is the countries:sanctions interaction p-value is 0.9195. Thus, we can't reject the null hypothesis (the countries and sanctions variables don't interact significantly). This p-value suggests the interaction term isn't contributing significantly to the model. In other words, the countries and sanctions variables don't appear to interact with each other, and any interaction is likely by chance.