Problem Set 2

Tolga Bag - 23371290

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.

```r
# load data
load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/datasets/climateSupport.RData?raw=true"))
str(climateSupport) #I am checking the structure of the code. I can see that the format
#of the variables won't work unless I change them to fit my model.
# I convert countries to numbers and convert it to a number
climateSupport$countries <- as.numeric(gsub(" of 192", "", climateSupport$countries))
#then, I make 20 low, 80 medium, 160 high in terms of participation rates
climateSupport <- climateSupport %>%
   mutate(countries = case_when(
         countries == 20 ~ "low",
         countries == 80 ~ "medium",
         countries == 160 ~ "high"
   ))
#I need it to be a factor to run the model, so I change it:
climateSupport$countries <- as.factor(climateSupport$countries)
# I convert sanctions by removing '%', and converting 'None' to 0
climateSupport$sanctions <- as.numeric(sub("%", "", climateSupport$sanctions))
climateSupport$sanctions[is.na(climateSupport$sanctions)] <- 0
climateSupport$sanctions <- factor(climateSupport$sanctions, levels = c(0, 5, 15, 20))
# I convert the choice variable from supported/non supported to 1 and 0. I turn
#it into a factor for the model
climateSupport$choice <- ifelse(climateSupport$choice == "Supported", 1, 0)
climateSupport$choice <- as.factor(climateSupport$choice)
str(climateSupport) #now they are all factoral and could be fit to the model.
#I got many errors. I use this and Gemini: https://stat.ethz.ch/pipermail/r-help/2007-October/143569.html
#I decided to change to glmnet and fit a logistical regression since gam didn't work.
logit_model <- glm(choice ~ countries + sanctions,
                   data = climateSupport,
                   family = binomial())
```

```
summary(logit_model)
summary_output <-capture.output(summary(logit_model)) #I provide the
    output.
cat("My R Output:\n", summary_output, file = "output1.txt", sep = "\n")
```

Coefficients:

Estimate Std. Error z value Pr(¿—z—)

(Intercept) 0.37569 0.05416 6.937 4.01e-12 ***

countrieslow -0.64835 0.05388 -12.033 ¡ 2e-16 ***

countriesmedium -0.31199 0.05387 -5.792 6.97e-09 ***

sanctions5 0.19186 0.06216 3.086 0.00203 **

sanctions15 -0.13325 0.06208 -2.146 0.03183 *

sanctions20 -0.30356 0.06209 -4.889 1.01e-06 ***

—

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: 11568 on 8494 degrees of freedom

AIC: 11580

Number of Fisher Scoring iterations: 4

The global null hypothesis is none of the independent variables (the number of participating countries or the sanctions the lack of compliance brings) are linked to the probability of an individual supporting a policy. The p value is ¡ 2e-16, meaning it is a tiny number. In other words, we can reject the null hypothesis since there is a statistically significant relationship between at least one of the independent variables and the probability of "choice". intercept is 0.376, which means that the predicted probability of "choice" when countries is "high" and there are no sanctions is 37.6% in our 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)

   The "higher" in countries refer to the highest number of countries' participation into the policy. As higher is our reference point in the regression, the coefficients 0.19186 for 5% sanctions, -0.13325 for 10% and 15% for 20% sanctions mean

the higher the sanctions the lesser the individual will support the policy per this model. In other words, increasing sanctions from 5% to 15% will decrease the odds that an individual will support the policy within the limits of our model.

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

The model predicts that the lesser the participation of countries, the lesser likelihood of estimated probability that an individual will support a policy. Specifically 80 of 192 countries participating with no sanctions are referred in the countriesmedium without any sanctions. the estimated probability that an individual will support a policy is 31.19% less probable than the higher number of countries' participation (our reference point in the model).

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

- It depends on the data and results. The best way is to analyze it and make the interpretation.
- Perform a test to see if including an interaction is appropriate.

```
#the interpretation.

logit_model2 <- glm(choice ~ countries + sanctions + countries*
    sanctions,
                    data = climateSupport,
                    family = binomial())
```

Coefficients:
Estimate Std. Error z value Pr(¿—z—)
(Intercept) 0.3379696 0.0773968 4.367 1.26e-05 ***
countrieslow -0.6126645 0.1080141 -5.672 1.41e-08 ***
countriesmedium -0.2370437 0.1077348 -2.200 0.0278 *
sanctions5 0.2518725 0.1083852 2.324 0.0201 *
sanctions15 -0.1485161 0.1076839 -1.379 0.1678
sanctions20 -0.1957184 0.1092585 -1.791 0.0732 .
countrieslow:sanctions5 -0.1300853 0.1510307 -0.861 0.3891
countriesmedium:sanctions5 -0.0353744 0.1545468 -0.229 0.8190
countrieslow:sanctions15 0.0516474 0.1526666 0.338 0.7351
countriesmedium:sanctions15 -0.0006461 0.1512924 -0.004 0.9966
countrieslow:sanctions20 -0.0568798 0.1536716 -0.370 0.7113
countriesmedium:sanctions20 -0.2540852 0.1518989 -1.673 0.0944 .

—

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: 11562 on 8488 degrees of freedom
AIC: 11586
Number of Fisher Scoring iterations: 4

- Once interactions between sanctions and number of countries are included, none of the results concerning these interactions are statistically significant since p values range from 0.0944 to 0.9966. We can't reject our null hypothesis. Even though there are bits like countriesmedium or sanctions5 with statistical significant results, the overall interaction terms do not reach a strong level of significance. Hence, it won't change our assessment.