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.

    First, I reassigned the countries and sanctions variables as unordered factors, because having them as ordered factors was messing up my lm output:

```
climateSupport$countries <- factor(climateSupport$countries, ordered=
  FALSE)
climateSupport$sanctions <- factor(climateSupport$sanctions, ordered=
  FALSE)
```

    Then I fit the additive model:

```
mod1 <- glm(choice ~ ., family =binomial(link="logit"), data =
  climateSupport)
```

    The summary of mod1 is below:

```
Call:glm(formula = choice ~ ., family = binomial(link = "logit"),
     data = climateSupport)
 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)         -0.27266    0.05360  -5.087 3.64e-07 ***
countries80 of 192   0.33636    0.05380   6.252 4.05e-10 ***
countries160 of 192  0.64835    0.05388  12.033  < 2e-16 ***
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 = (H0: all slopes (/estimated relationships in our model) = 0)

Created a Null Model:

```
nullmod1 <- glm(choice ~ 1, family = binomial(link="logit"), data =
    climateSupport)
```

Summary of Null Model output:

```
glm(formula = choice ~ 1, family = binomial(link = "logit"),
data = climateSupport)
Deviance Residuals:
    Min      1Q  Median      3Q     Max
    -1.175  -1.175  -1.175   1.180   1.180
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.006588   0.021693  -0.304    0.761

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 11783  on 8499  degrees of freedom
Residual deviance: 11783  on 8499  degrees of freedom
AIC: 11785
Number of Fisher Scoring iterations: 3
```

Then, going to run an ANOVA test to assess this global null hypothesis:

```
anova1 <- anova(nullmod1, mod1, test = "LRT")
```

Results of anova1 (the summary):

```
[[1]]Analysis of Deviance Table
Model 1: choice ~ 1
Model 2: choice ~ countries + sanctions
Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
2      8494      11568  5   215.15 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

P value is less than 0.01 (is less than 2.23-16, so very close to 0). This means that we can conclude that at least one predictor is reliable in our mod1 model (our additive model), i.e. that our additive model is at least better fit than our null model with no predictors included.

Results and Conclusion Described: The results of the anova (analysis of deviance between the two models) test shows a very small p-value, which is grounds to reject our null hypothesis (that all slopes are 0, or that there are no relationships between the predictor and outcome variables in mod1).

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)

   To do this, need to use the Odds Ratio (OR) – for two diff groups (5 percent sanctions and 15 percent sanctions) holding countries variable constant at 160-192.

   So: first going to plug in the mod1 summary estimates to find predicted probability that Y=support when countries = 160-192 and sanctions = 5 percent. Did this with following code:

```
PProb_sanctions5 <- (1/(1+exp(-(-0.27266+.64835+.19186))))
# Probability when sanctions = 5 ==> 0.638
```

   Next, I found probability that Y=support when countries = 160-192 and sanctions = 15 percent. With the below code:

```
PProb_sanctions15 <- (1/1+exp(-(-0.27266+.64835-.13325)))
# Probability when sanctions = 15 ==> 1.785
```

   Then calculated the odds ratio of moving from 5 to 15 percent sanctions with country support held constant at 160-192:

```
OR_5to15 <- PProb_sanctions15/PProb_sanctions5
```

   Interpretation: the odds of a bill being supported when sanctions move from 5 to 15 percent, with a policy supported by 160-192 countries, increases by 2.796 percent.

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

   Calculated this by plugging in the respective estimates (as I did in 2a) but for the correct terms for this question. Did this with the following code:

```r
PProb_b <- (1/1+exp(-(-.27266+.33636+0)))
# PProb_b = 1.938
```

This means the estimated probability that an indiviudal will support a policy if 80 of 192 countries participate in it with no sanctions is 1.938

(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.

Yes, potentially the answers to 2a and 2b could change if we included the interaction term in our model. Theoretically speaking this is because it might make sense that the effect of one of our covariates (sanctions, for example), increasing from 5 to 15 percent may have a different effect on the outcome (choice), depending on the other predictor (country support) - if more or less countries are supportive of the policy.

I will perform a significance test for different slopes to see if we get a better model fit by using an interactive term rather than an additive term in our model. The p-value of our ANOVA test will indicate whether our interactive model is a better fit or not.

First, I will make a mod2 (which has the interactive term):

```r
mod2 <- glm(choice ~ countries*sanctions, family =binomial(link="logit"), data = climateSupport)
```

Next I run the ANOVA:

```r
anova2 <- anova(mod1, mod2, test ="LRT")
```

Which outputs the following:

```
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
```

Pvalue is over threshold (of 0.01), is .3912, meaning the model is not a better fit when the interaction term is included. In other words, there does not seem to be a different effect of percentage of sanctions depending on the number of countries that participate in the policy on the likelihood of supporting the policy (choice = supported), thus sticking to the additive model would be preferable. (Using an interactive term is not appropriate).