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.

   Answer Question 1:

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

   Model Output in R:

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

Call:
glm(formula = choice ~ countries + sanctions, 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) -0.005665   0.021971  -0.258 0.796517
countries.L  0.458452   0.038101  12.033  < 2e-16 ***
countries.Q -0.009950   0.038056  -0.261 0.793741
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 ***
---
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 null hypothesis states that the explanatory variables of countries and sanction levels do not affect the log odds of someone supporting a policy.

The intercept value of (-.006) in the output indicates the log odds of individuals supporting a policy when the explanatory variables are at reference level, 0.

```
(exp(-.0057)) / (1 + exp(-.0057))
```

The probability that someone will support a policy is .4986.

We can see that moving to the highest levels of countries involved in a policy is not a statistically significant contribution to the model with a p value of .794. The other explanatory variables are significantly associated with the probability of supporting a policy.

To get the model chi square value we subtract the residual deviance score from the null deviance score

```
X2 <- 111783 - 11568
X2

100215
```

There are p = 5 predictor variables degrees of freedom.

The p value here is less than .5 and therefore we reject the null hypothesis.This indicates that at least one of the explanatory variables are significant in the model at predicting policy support.

If any of the explanatory variables are significant in this model, then:

1. 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 Question 2 - Part 1:

```
levels(climateSupport$sanctions)

exp(-.01811)
```

The coefficient is equal to (-.01811) and the p value is ¡.05, this indicates that this change in sanction levels does influence policy support. The negative value indicates that increaseing the sanction would decrease the likelihood of someone supporting the policy.

```
In a policy in which nearly all countries participate, the odds of someone
supporting a policy with 15% sanctions is .9821 times less than the odds of
someone supporting a policy with 5% sanctions.

1-.9821 = 17.9%
Policies with 15% sanctions are associated with a 17.9% reduction in
policy support.
```

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

   Answer Question 2 - Part 2:

```
predicted_data <- with(climateSupport,
expand.grid(countries = unique(countries),
sanctions = unique(sanctions)))

predicted_data <- cbind(predicted_data,
predict(model, newdata = predicted_data,
type = "response",
se = TRUE))

predicted_data <- within(predicted_data,
{PredictedProb <- plogis(fit)
LL <- plogis(fit - (1.96*se.fit))
UL <- plogis(fit + !1.96*se.fit)
                         })
print(predicted_data)

select(predicted_data, countries, sanctions, PredictedProb)
    countries sanctions PredictedProb
1   80 of 192       15%     0.6183663
```

```
2  160 of 192       15%     0.6365253
3   20 of 192       15%     0.5986620
4   80 of 192      None     0.6261930
5  160 of 192      None     0.6440148
6   20 of 192      None     0.6064116
7   80 of 192        5%     0.6372719
8  160 of 192        5%     0.6543455
9   20 of 192        5%     0.6177028
10  80 of 192       20%     0.6083351
11 160 of 192       20%     0.6266853
12  20 of 192       20%     0.5889923
```

The estimated probability that an individual will support a policy if 80/192 countries are participating and there are no sanctions is 0.6262.

3. 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 Question 2 - Part 3:

   If there is an interaction term in this model then the distributions of the variables may crossover eachother. This would result in different log likelihood and probability values in relation to policy support.

   To see if an interaction model is needed here I have performed a likelihood ratio test. The null hypothesis states that the second model, the interactive model, is a better fit for the data.

```
model2 <- glm(choice ~ countries
            * sanctions,family=binomial(link='logit'),data=climateSupport)
summary(model2)
install.packages('lmtest')
library(lmtest)

lrtest(model, model2)

Output:

Likelihood ratio test

Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries * sanctions
```

```
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   6 -5784.1
2  12 -5781.0  6 6.2928     0.3912
```

We can see from the output that the p value is not less than .05 so we fail to reject the null hypothesis. The interactive predictor variables do not offer a significant improvement from the additive model.