# Problem Set 2

Applied Stats II

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.

   **Answer**:

```r
#1#

## check data
summary(climateSupport)

## prepare data
DF$choice<- ifelse(DF$choice == "Supported", 1, 0)
DF$countries<- factor(DF$countries,ordered=FALSE)
DF$sanctions<- factor(DF$sanctions,ordered=FALSE)

## check data again
summary(climateSupport)

## Run the logit regression
logit <- glm(choice ~ countries+sanctions,
             family = binomial(link = "logit"), data = DF)

## check the result
# summary output
summary(logit)#summary output
```

```
Call:
glm(formula = choice ~ countries + sanctions, family = binomial(link = "logit"),
    data = DF)

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
```

Figure 1: logit

`The global null hypothesis and p-value`: the global null hypothesis is that efficient equals to 0.the p-value for the coefficient of countries80 of 192 is 4.05e-10, the null hypothesis is that the coefficient = 0, that is, is the average difference on log-odds value between countries80 of 192 and countries20 of 192 is 0. The p-value for the coefficient of countries160 of 192 is ¡ 2e-16, the null hypothesis is that the coefficient = 0, that is, is the average difference on log-odds value between countries160 of 192 and countries20 of 192 is 0. the p-value for the coefficient of sanctions5% is 0.00203, the null hypothesis is that the coefficient = 0, that is, is the average difference on log-odds value between sanctions5% and None is 0. The p-value for the coefficient of sanctions15% is 0.03183, the null hypothesis is that the coefficient = 0, that is, is the average difference on log-odds value between sanctions15% and None is 0. The p-value for the coefficient of sanctions20% is 1.01e-06, the null hypothesis is that the coefficient = 0, that is, is the average difference on log-odds value between sanctions20% and None is 0.Conclusion: All coefficient are significantly not equal to 0. Thus, we get the predicion equation:

`Equation`: log(odds)=-0.27266+0.33636*D(countries80 of 192)+0.64835*D(countries160 of 192) +0.19186*D(sanctions5% )-0.13325*D(sanctions15%)-0.30356*D(sanctions20%)

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)

   Answer:

   log(odds)=-0.27266+0.33636*D(countries80 of 192)+0.64835*D(countries160 of 192) +0.19186*D(sanctions5% )-0.13325*D(sanctions15%)-0.30356*D(sanctions20%)

   When countries participate [160 of 192] and sanctions is 5%, the equation is log(odds)=-0.27266+0.64835+0.19186=0.56755 odds=exp(0.56755)=1.76394

   When countries participate [160 of 192] and sanctions is 15%, the equation is log(odds)=-0.27266+0.64835-0.13325=0.24244 odds=exp(0.24244)=1.274355

   The change in odds is 1.274355-1.76394=-0.489585

   The change in log-odds equals to the coefficient of D(sanctions15%) minus the coefficient of D(sanctions5%), so the change in odds is -0.489585

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

Answer:

if there are 80 of 192 countries participating with no sanctions log(odds)=-0.27266+0.33636=0.06 p=exp(log(odds))/(1+exp(log(odds)))=0.5159196 the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions is 0.5159196

(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.

Answer:

Yes, the answers to 2a and 2b would potentially change if we included the interaction term in this model, because when an interaction term is included in the model, it allows for the possibility that the effect of one predictor on the outcome variable depends on the level of another predictor. This means that the relationship between one predictor and the outcome variable is not constant across different levels of the other predictor. As a result, the interpretation of the coefficients associated with each predictor becomes more complex, as the effect of one predictor depends on the level of the other predictor.

The P-value of chisq-test bigger than 0.05, which means we cannot reject the null hypothesis that the equation including interaction items cannot perform better significantly than equatio not including those . Therefore, including an interaction is not appropriate.

```
logit1 <- glm(choice ~ countries+sanctions, family = binomial(link =
    "logit"), data = DF)
logit2 <- glm(choice ~ countries+sanctions+countries:sanctions,
    family = binomial(link = "logit"), data = DF)
anova (logit1, logit2,test="Chisq")
```

```
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
```

Figure 2: Anova