

# Problem Set 2

Han Li

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.

```

1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/
  datasets/climateSupport.RData?raw=true"))
3 #recode to factor
4 climateSupport$sanctions <- relevel(factor(climateSupport$sanctions ,
  ordered = F), ref="None")
5 climateSupport$countries <- relevel(factor(climateSupport$countries ,
  ordered = F), ref="20 of 192")
6 additive_log_model <- glm(choice~countries+sanctions , data=climateSupport
  , family = binomial(link = "logit"))
7 summary(additive_log_model)

```

```

1 Call:glm(formula = choice ~ countries + sanctions ,
2 family = binomial(link = "logit"),
3 data = climateSupport) Coefficients:
4      Estimate Std. Error z value Pr(>|z|)
5 (Intercept)      -0.27266    0.05360   -5.087 3.64e-07 ***
6 countries80 of 192    0.33636    0.05380    6.252 4.05e-10 ***
7 countries160 of 192   0.64835    0.05388   12.033 < 2e-16 ***
8 sanctions5%          0.19186    0.06216    3.086 0.00203 **
9 sanctions15%         -0.13325    0.06208   -2.146 0.03183 *
10 sanctions20%         -0.30356    0.06209   -4.889 1.01e-06 ***
11
12 (Dispersion parameter for binomial family taken to be 1)
13 Null deviance: 11783  on 8499  degrees of freedom
14 Residual deviance: 11568  on 8494  degrees of freedom
15 AIC: 11580Number of Fisher Scoring iterations: 4
16

```

The global null hypothesis is :none of the variable increase the model fit/ all variable's coefficient estimate is 0. We can use anova to get p value.

```

1 #the global null is if all explanatory coefficient are zero , hence we can
  use anova
2 null_model<- glm(choice~1,data=climateSupport ,family = binomial(link ='
  logit '))
3 anova(null_model, additive_log_model, test="LRT")

```

```

1 Analysis of Deviance Table
2 Model 1: choice ~ 1
3 Model 2: choice ~ countries + sanctions
4 Resid. Df Resid. Dev Df Deviance Pr(>Chi)

```

```

5      1      8499      11783
6      2      8494      11568  5    215.15 < 2.2e-16 ***
7
8

```

We can see the pvalue is extremely small and below our threshold of 0.05, so we have found evidence to reject the null hypothesis that none of the variables improves the model fit.

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) From the above summary we can see that the coefficient for sanction from 5% to 15% will change the odds by  $-0.13325 - 0.19186 = -0.32511$ , because this is an additive model. So the log odds of someone will support the policy will decrease by 0.32511 on average, holding other factor constant, including in which nearly all countries participate as will did not model interactive term.
- (b) What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions?

```

1 # Problem 2
2 #####
3 #significant variables are: countries80, country160, sanctionNone,
  Sanction15, sanction20
4 additive_b <- predict(additive_log_model, newdata = data.frame(
  countries="80 of 192", sanctions="None"), type="response")

1 additive_b
2 0.5159191

```

So after rounding the probability is about 52

- (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.

```

1 #c adding interactive
2 interactive_log_model <- glm(choice~countries*sanctions, data=
  climateSupport, family = binomial(link = "logit"))
3 summary(interactive_log_model)
4
5 interactive_b<-predict(interactive_log_model, newdata = data.
  frame(countries="80 of 192", sanctions="None"), type="response
  ")
6 change_b<-interactive_b-additive_b
7 # compare to using anova
8 anova(additive_log_model, interactive_log_model, test="LRT")

```

The answers for 2a and 2b will potentially change because once we add the interaction term, the slope for different policy term in relation to sanction may also differ in addition to just additive intercept. We can run the model below to compare.

```

1 #c adding interactive
2 interactive_log_model <- glm(choice ~ countries * sanctions, data=
  climateSupport, family = binomial(link = "logit"))
3 summary(interactive_log_model)

```

```

1 Call:glm(formula = choice ~ countries * sanctions, family =
  binomial(link = "logit"),
2 data = climateSupport)
3 Coefficients:
  z value Pr(>|z|)
4 (Intercept)          -0.27469      0.07534    -3.646
  0.000267
5 countries80 of 192      0.37562      0.10627     3.535
  0.000408
6 countries160 of 192     0.61266      0.10801     5.672 1.41
  e-08
7 sanctions5%           0.12179      0.10518     1.158
  0.246909
8 sanctions15%          -0.09687      0.10822    -0.895
  0.370723
9 sanctions20%          -0.25260      0.10806    -2.338
  0.019412
10 countries80 of 192:sanctions5%  0.09471      0.15232     0.622
  0.534071
11 countries160 of 192:sanctions5%  0.13009      0.15103     0.861
  0.389063
12 countries80 of 192:sanctions15% -0.05229      0.15167    -0.345
  0.730262
13 countries160 of 192:sanctions15% -0.05165      0.15267    -0.338
  0.735136
14 countries80 of 192:sanctions20% -0.19721      0.15104    -1.306
  0.191675
15 countries160 of 192:sanctions20%  0.05688      0.15367     0.370
  0.711279
  (Intercept)
  ***
16 countries80 of 192          ***
17 countries160 of 192         ***
18 sanctions5%
19 sanctions15%
20 sanctions20%          *
21 countries80 of 192:sanctions5%
22 countries160 of 192:sanctions5%
23 countries80 of 192:sanctions15%
24 countries160 of 192:sanctions15%
25 countries80 of 192:sanctions20%
26 countries160 of 192:sanctions20%
27

```

```

28 Null deviance: 11783 on 8499 degrees of freedom
29 Residual deviance: 11562 on 8488 degrees of freedom
30 AIC: 11586Number of Fisher Scoring iterations: 4

```

So we see for a policy where almost all countries participate, the sanction increase will need to include both the interaction term ( but the interaction doesn't seem statistically significant). For 2b, we can also calculate the probability with the new model interactive b, and see the change is about 0.009.

```

1
2 interactive_b<-predict(interactive_log_model, newdata = data.
   frame(countries="80 of 192", sanctions="None"), type="response
   ")
3 change_b<-interactive_b-additive_b

```

Finally, we can perform LRT test to check for pvalue.

```

1 # compare to using anova
2 anova(additive_log_model, interactive_log_model, test="LRT")

```

```

1 Analysis of Deviance Table
2 Model 1: choice ~ countries + sanctions
3 Model 2: choice ~ countries * sanctions
4 Resid. Df Resid. Dev Df Deviance Pr(>Chi)
5 1      8494      11568
6 2      8488      11562  6    6.2928    0.3912

```

As we can see, the pvalue of 0.3912 is not below 0.05, so we cannot reject the null hypothesis that the interactive model does not improve our additive model,