

# Problem Set 2

Applied Stats II  
Dan Zhang 23335541

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:

First, let's load the data by using the url in R:

```
1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/
  datasets/climateSupport.RData?raw=true"))
3 write.csv(climateSupport, "climateSupport.csv")
4
5 # Question 1
6 data <- climateSupport
```

Then, let's do some preprocessing with the data: transfer choice into binary, convert countries and sanctions into factors.

```
1 # Transfer choice into binary variable
2 data$choice <- ifelse(data$choice == "Supported", 1, 0)
3
4 # Convert countries and sanctions columns into factors
5 data$countries <- factor(data$countries, levels = c("20 of 192", "80 of
  192", "160 of 192"), ordered = FALSE)
6 data$sanctions <- factor(data$sanctions, levels = c("None", "5%", "15%",
  "20%"), ordered = FALSE)
7 # Change reference level
8 data$countries <- relevel(data$countries, ref = "20 of 192")
9 data$sanctions <- relevel(data$sanctions, ref = "None")
```

Next, run a glm model in R with "binomial" family, the model function is :

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * countries + \beta_2 * sanctions$$

p: The probability of individual support for the policy

```
1 # Run the glm model
2 glm_model <- glm(choice ~ countries + sanctions, family = binomial(link =
  "logit"), data = data)
3 summary(glm_model)
```

The model results are in table 1:

Table 1:

	<i>Dependent variable:</i>
	choice
countries80 of 192	0.336*** (0.054)
countries160 of 192	0.648*** (0.054)
sanctions5%	0.192*** (0.062)
sanctions15%	−0.133** (0.062)
sanctions20%	−0.304*** (0.062)
Constant	−0.273*** (0.054)
Observations	8,500
Log Likelihood	−5,784.130
Akaike Inf. Crit.	11,580.260
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

As we know, in this case, the **global null hypothesis** is:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ . Either the number of participating countries or sanctions for missing emission reduction targets have no impact on the likelihood of individual support policy.

Check Table 1, we can see that all the estimated coefficients are statistically significant with p-value  $< 0.01$ . Therefore, we have strong evidence to reject the global null hypothesis, that is, at least 1 explanatory variable (countries or sanctions) in the model has a significant impact on the response variable (choice).

Here are the coefficients interpretations:

**Intercept (-0.273)**: When 20 countries support the policy and there is no sanction, the log odds of an individual supporting the policy is -0.273. It means that the individual support probability is approximately 43% ( $p = \frac{e^{-0.273}}{1 + e^{-0.273}}$ ).

**Countries80 of 192 (0.336)**: Comparing 20 countries supporting the policy, when there are 80 countries participating, the log odds of supporting the policy increase by 0.336. This suggests that more state involvement increases the probability of policy support to approximately 58% ( $p = \frac{e^{0.336}}{1 + e^{0.336}}$ ).

**Countries160 of 192 (0.648)**: Comparing 20 countries supporting the policy, when there are 160 countries participating, further significantly increases the log odds of supporting the policy by 0.648. The individual support probability is approximately 66% ( $p = \frac{e^{0.648}}{1 + e^{0.648}}$ ).

**Sanctions5 % (0.192)**: Relative to no sanctions, a sanctions level of 5% increases the log odds of supporting the policy by 0.192. The individual support probability is approximately 55% ( $p = \frac{e^{0.192}}{1 + e^{0.192}}$ ).

**Sanctions15% (-0.133) and sanctions20% (-0.304)**: Relative to no sanctions, higher levels of sanctions (15% and 20%) reduce the log odds of support for the policy. This suggests that if sanctions are too harsh, they may reduce public support for the policy.

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)

In order to compare sanction rate increases from 5% to 15%, we need to check the support probabilities of different sanction rates comparing with no sanction. As the odds ratio of sanctions5% is greater than 1 ( $e^{0.192} > 1$ ), it means that 5% sanctions increase the probability of support compared to no sanctions. And the odds ratio of sanctions15% is less than 1 ( $e^{-0.133} < 1$ ), this means that 15% sanctions reduce the probability of support compared to no sanctions. Thus, increasing sanctions from 5% to 15%, we see a change from potentially increasing the probability of support to decreasing the probability of support.

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

In order to find the probability of an individual supporting a policy when there are 80 of 192 countries participating with no sanctions. We need to find the log odds, which here would be  $0.336 + (-0.273) = 0.063$ . So, the probability of an individual supporting a policy when there are 80 of 192 countries participating with no sanctions would be  $\frac{e^{0.063}}{1+e^{0.063}} = 0.516$ .

- (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.  
Let's run a glm with interaction in R:

```
1 # Run glm model with interaction
2 glm_model_interaction <- glm(choice ~ countries*sanctions, family
  = binomial(link = "logit"), data = data)
3 summary(glm_model_interaction)
```

Here are the results in Table 2:

Table 2:

	<i>Dependent variable:</i>
	choice
countries80 of 192	0.376*** (0.106)
countries160 of 192	0.613*** (0.108)
sanctions5%	0.122 (0.105)
sanctions15%	−0.097 (0.108)
sanctions20%	−0.253** (0.108)
countries80 of 192:sanctions5%	0.095 (0.152)
countries160 of 192:sanctions5%	0.130 (0.151)
countries80 of 192:sanctions15%	−0.052 (0.152)
countries160 of 192:sanctions15%	−0.052 (0.153)
countries80 of 192:sanctions20%	−0.197 (0.151)
countries160 of 192:sanctions20%	0.057 (0.154)
Constant	−0.275*** (0.075)
Observations	8,500
Log Likelihood	−5,780.983
Akaike Inf. Crit.	11,585.970

Check Table 2, we can see that none of the coefficients on the interaction terms are statistically significant (except the intercept: baseline 20 of countries with no ). We can conclude that the interaction effect does not significantly increase the explanation of the model. This means that the impact of different countries' participation does not depend on the level of sanctions and vice versa.