

# Problem Set 2

Yuanyuan Liu

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
4 # Check the structure of the data
5 summary(climateSupport)
6
7 # Coerce from a character vector to a logical vector (assuming 'choice'
  was originally "Supported"/"Not Supported")
8 climateSupport$choice <- ifelse(climateSupport$choice == "Supported", 1,
  0)
9
10 # Convert to logical
11 climateSupport$choice <- as.logical(climateSupport$choice)
12
13 # Set 'countries' and 'sanctions' as ordered factors
14 climateSupport$countries <- relevel(factor(climateSupport$countries,
  levels = c("20 of 192", "80 of 192", "160 of 192"), ordered = FALSE),
  ref = "20 of 192")
15 climateSupport$sanctions <- relevel(factor(climateSupport$sanctions,
  levels = c("None", "5%", "15%", "20%"), ordered = FALSE), ref = "5%")
16
17
18 # Fit the logistic regression model
19 glmModel <- glm(choice ~ countries + sanctions, data = climateSupport,
  family = binomial(link="logit"))
20
21 # Display the summary of the model
22 summary_glmModel <- summary(glmModel)
23
24 # The global null hypothesis test using an ANOVA type of approach
25 anova_glm <- anova(glmModel, test="Chisq")
26
27 # Print the summary of the model
28 print(summary_glmModel)
29
30 # Print the results of the global test
31 print(anova_glm)
```

**Output:**

Table 1: Model Output

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	-0.08081	0.05316	-1.520	0.12848
countries80 of 192	0.33636	0.05380	6.252	4.05e-10 ***
countries160 of 192	0.64835	0.05388	12.033	< 2e-16 ***
sanctionsNone	-0.19186	0.06216	-3.086	0.00203 **
sanctions15%	-0.32510	0.06224	-5.224	1.76e-07 ***
sanctions20%	-0.49542	0.06228	-7.955	1.79e-15 ***

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

Table 2: Analysis of Deviance Table

	Df	Deviance	Resid. Df	Resid. Dev	Pr(> $\chi^2$ )
NULL			8499	11783	
countries	2	146.724	8497	11637	< 2.2e-16 ***
sanctions	3	68.426	8494	11568	9.272e-15 ***

**Signif. codes:** 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**global null hypothesis:** The global null hypothesis for logistic regression is that none of the explanatory variables have an effect on the response variable. The ANOVA table shows that after adding countries and sanctions, the bias of the model is significantly reduced, and the corresponding p-value is much less than 0.05, so we reject the global null hypothesis, indicating that these explanatory variables do provide significant information on the probability of supporting the policy.

**Result:** All coefficients have very small p-values, which means they are statistically significant.

The coefficient for countries80 of 192 is 0.33636, indicating that the log odds of supporting the policy are increased by about 0.33636 for the category 80 of 192 compared to the baseline category (assumed to be 20 of 192).

The coefficient for countries160 of 192 is 0.64835, indicating that the log odds of 160 of 192 increases by approximately 0.64835 compared to the baseline category.

The coefficient of sanctions None is -0.19186, indicating that the log odds of supporting the policy in the absence of sanctions (None) is reduced by about 0.19186 compared to the 5% sanctions level.

The coefficient on sanctions15% is -0.32510, indicating that the log odds of supporting

the policy are reduced by approximately 0.32510 at the 15% sanctions level compared to the 5% sanctions level.

The coefficient of sanctions20% is -0.49542, indicating that the log odds of supporting the policy are reduced by about 0.49542 at the 20% sanctions level compared to the 5% sanctions level.

**Conclusion:** The number of participating countries significantly affects the degree of support for a policy, and more countries participating are positively associated with higher levels of support.

Relative to the 5% sanction level, there was less support for the policy without sanctions, while higher sanction levels (15% and 20%) were associated with lower levels of support, suggesting that higher sanctions may reduce public support for policy.

Based on the above analysis, we can conclude that the number of countries and the severity of sanctions have a clear impact on predicting whether individuals support a certain policy. Policymakers need to carefully consider these two factors when considering how to increase support for their policies. Especially in terms of sanctions, milder sanctions (such as 5%) seem to be better able to promote policy support than no sanctions and higher proportions of sanctions (such as 15% and 20%). This may be because excessive sanctions may arouse resentment, whereas moderate sanctions may be viewed as fairer or more acceptable.

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)

```

1 # Calculate the odds ratio for countries 160 of 192
2 odds_ratio_countries = exp(0.64835)
3 odds_ratio_countries
4 # Calculate the odds ratio for an increase in sanctions from 5% to
   15%
5 odds_ratio_increase_sanctions = exp(-0.32510)
6 odds_ratio_increase_sanctions
7 # The change in odds
8 change_in_odds = odds_ratio_countries * odds_ratio_increase_sanctions
9 change_in_odds

```

Table 3: Your table caption here

Values	Numbers
odds_ratio_countries	1.91383
odds_ratio_increase_san	0.72246
change_in_odds	1.38161

**Odds Ratio for Countries:** An odds ratio of 1.91238 for countries160 of 192 indicates that when 160 out of 192 countries participate, the odds that an individual will support the policy are about 91.24% higher than when only 20 out of 192 countries participate.

**Odds Ratio for Sanctions Increase:** An odds ratio of 0.72246 for increasing sanctions from 5% to 15% indicates a decrease in the odds of an individual supporting the policy by about 27.75% when considering the effect of sanctions alone, relative to the baseline sanction level of 5%.

**Combined Change in Odds:** The change in odds of 1.38161 indicates the combined effect of having 160 out of 192 countries participating and increasing sanctions from 5% to 15%. Even though increasing sanctions typically reduces support, the strong positive effect of broad country participation (160 out of 192) outweighs the negative effect of increased sanctions in this case.

Therefore, the presence of nearly all countries in the agreement has a significant positive impact on the likelihood of policy support, which is strong enough to overcome the decrease in support due to heightened sanctions from 5% to 15%. The net effect is an overall increase in the odds of an individual supporting the policy by about 38.16%, even with the increase in sanctions.

- (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 # Coefficients
2 intercept <- -0.08081
3 coef_countries80 <- 0.33636
4 coef_sanctionsNone <- -0.19186
5
6 # Calculate the linear predictor (log odds)
7 log_odds <- intercept + (coef_countries80 * 1) + (coef_sanctionsNone
8   * 1)
9
10 # Convert log odds to probability
11 probability <- exp(log_odds) / (1 + exp(log_odds))
12 probability

```

The value of probability: 0.51592 The estimated probability that an individual will support a policy with 80 of 192 countries participating and no sanctions, while the reference level of sanctions is 5%, is approximately 51.59%

- (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 # Fit the original model without interaction

```

```

2 original_model <- glm(choice ~ countries + sanctions, family =
  binomial(link = "logit"), data = climateSupport)
3
4 # Fit the new model with interaction
5 interaction_model <- glm(choice ~ countries * sanctions, family =
  binomial(link = "logit"), data = climateSupport)
6
7 # Compare the two models
8 model_comparison <- anova(original_model, interaction_model, test = "
  Chisq")
9 model_comparison

```

Table 4: Analysis of Deviance Table

	Resid. Df	Resid. Dev	Df	Deviance	$\Pr(\chi^2)$
Model 1: choice ~ countries + sanctions	8494	11568			
Model 2: choice ~ countries * sanctions	8488	11562	6	6.2928	0.3912

The key metric to focus on here is the p-value for the chi-squared test comparing the two models, which is 0.3912. This p-value is quite high, much greater than the common alpha level of 0.05. This indicates that there is no statistical evidence that the interaction terms provide a better fit to the data compared to the model without interactions. In other words, including the interaction term between countries and sanctions does not significantly improve the model's ability to predict the likelihood of an individual supporting the policy.