

Problem Set 2

Hanyu Li (25346841)

February 14, 2026

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 Wednesday February 18, 2026. 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 supported, 0 otherwise).
- Explanatory variables: `countries` and `sanctions`.

Question 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

Given in the raw data, predictors are coded as factors, so to make the model interpretation more clear we first transformed variables *countries* and *sanctions* (with unit of percentage) as numeric ones.

```
1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_2026/blob/main/datasets/
  climateSupport.RData?raw=true"))
3 str(climateSupport)
4
5 # Data cleaning
6 climateSupport$countries_n <- as.numeric(as.character(climateSupport$ 
  countries))
7 library(dplyr)
8 climateSupport <- climateSupport |>
9   mutate(
10     countries_n = case_when(
11       grepl("20", countries) ~ 20,
12       grepl("80", countries) ~ 80,
13       grepl("160", countries) ~ 160
14     ),
15     sanctions_n = case_when(
16       grepl("None", sanctions) ~ 0,
17       grepl("5%", sanctions) ~ 5,
18       grepl("15%", sanctions) ~ 15,
19       grepl("20%", sanctions) ~ 20
20     ))
21 str(climateSupport)
```

In the framework of Generalized Linear Models (GLM), we employ the logit link function to model the probability of policy support. Specifically, we estimate the following equation to interpret the relationship between the explanatory variables and the log-odds of support:

$$\ln \left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)} \right) = \beta_0 + \beta_{\text{countries}} \cdot \text{Countries}_i + \beta_{\text{sanctions}} \cdot \text{Sanctions}_i \quad (1)$$

where $P(Y_i = 1)$ represents the probability that individual i supports the policy, Countries _{i} shows the number of participating countries, and Sanctions _{i} denotes the sanction level. Then we fit an additive model and got outcomes as follows:

```

1 # run the logit regression model
2 add_model <- glm(choice ~ countries_n + sanctions_n,
3                   data = climateSupport,
4                   family = binomial(link = "logit"))
5 summary(add_model)

```

Table 1: Logistic Regression Results

<i>Dependent variable:</i>	
Support for Policy	
Countries (N)	0.005*** (0.0004)
Sanctions (Percentage)	-0.017*** (0.003)
Constant	-0.272*** (0.045)
Observations	8,500
Log Likelihood	-5,800.915
Akaike Inf. Crit.	11,607.830

Note: *p<0.1; **p<0.05; ***p<0.01

To assess the model fit, we can have the global null hypothesis:

- $H_0: \beta_{\text{countries}} = \beta_{\text{sanctions}} = 0$
- $H_a: \text{at least one } \beta \text{ not equal to 0}$

After implementing full (additive model including all predictors) and reduced (reduced model only has intercept) likelihood ratio test, we got the following results.

```

1 # run the reduced model
2 red_model <- glm(choice ~ 1, data = climateSupport,
3                    family = binomial(link = "logit"))
4 anova(red_model, add_model, test ="LRT")

```

Model 1: choice ~ 1

```

Model 2: choice ~ countries_n + sanctions_n
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       8499      11783
2       8497      11602  2   181.58 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

From the above analysis, we can conclude that:

- As the deviance table shows, p -value < 0.001 , we have enough evidence to reject H_0 and conclude that at least one predictor in the full model explains variation in log-odds of policy support.
- In the additive model outcome, both the number of participating countries ($p < 0.001$) and the level of sanctions ($p < 0.001$) are statistically associated with the log-odds of supporting the policy.
- Specifically, holding the level of sanctions constant, for every one-unit increase in the number of participating countries, the odds of an individual supporting the policy increase by a multiplicative factor of $e^{0.0046} \approx 1.005$ (or approximately a 0.5% increase in the odds).
- Controlling for the number of participating countries, for every one percentage increase in sanctions, the odds of supporting the policy change by a multiplicative factor of $e^{-0.0174} \approx 0.983$. This indicates that higher sanctions are associated with a decrease in the odds of support by approximately 1.7%.

Question 2

If any of the explanatory variables are statistically significant in this model, then: (a) calculate the odds change for 160 countries with sanctions increasing from 5% to 15%; (b) calculate the same for 20 countries; (c) calculate estimated probability for 80 countries with no sanctions.

Answer

Based on the regression results from Question 1, the fitted prediction equation for the log-odds of supporting the policy is:

$$\ln \left(\frac{\hat{P}(Y = 1)}{1 - \hat{P}(Y = 1)} \right) = -0.2719 + 0.0046 \cdot \text{Countries} - 0.0174 \cdot \text{Sanctions} \quad (2)$$

where Countries is the number of participating countries and Sanctions is the sanction level (in percentage).

(a) For the policy in which nearly all countries participate [160 of 192], we have:

$$\begin{aligned} \Delta \text{log-odds} &= (-0.2719 + 0.0046 \cdot 160 - 0.0174 \cdot 15) - (-0.2719 + 0.0046 \cdot 160 - 0.0174 \cdot 5) \\ &= -0.0174 \cdot (15 - 5) \\ &= -0.0174 \cdot 10 = -0.174 \end{aligned}$$

which means the log-odds decrease by 0.174, and the odds of support change by a multiplicative factor of $e^{-0.174} \approx 0.840$ (a 16.0% decrease).

(b) For the policy in which very few countries participate [20 of 192], we have:

$$\begin{aligned} \Delta \text{log-odds} &= (-0.2719 + 0.0046 \cdot 20 - 0.0174 \cdot 15) - (-0.2719 + 0.0046 \cdot 20 - 0.0174 \cdot 5) \\ &= -0.0174 \cdot (15 - 5) \\ &= -0.174 \end{aligned}$$

which means the log-odds decrease by 0.174, and the odds change by a factor of $e^{-0.174} \approx 0.840$ (identical to part a).

(c) To estimate the probability for 80 countries with no sanctions, we substitute values into the equation:

$$\begin{aligned} \text{log-odds} &= -0.2719 + 0.0046 \cdot 80 - 0.0174 \cdot 0 = 0.0961 \\ \hat{P}(Y = 1) &= \frac{e^{0.0961}}{1 + e^{0.0961}} \approx 0.524 \end{aligned}$$

The estimated probability that an individual will support the policy is approximately **52.4%**.

Question 3

Would the answers to 2a and 2b potentially change if we included an interaction term in this model? Why? Perform a test to see if including an interaction is appropriate.

Answer

To assess whether including an interaction term improves model fit, we estimated a full interaction model and performed a Likelihood Ratio Test (LRT) comparing it with the reduced additive model.

H_0 : Effect of number of participant countries is the same among sanction levels
 H_a : Effect of number of participant countries is different among sanction levels

```
1 #####
2 #####
3 # Problem 3
4 #####
5 # run the interaction model
6 int_model <- glm(choice ~ countries_n * sanctions_n,
7                     data = climateSupport,
8                     family = binomial(link = "logit"))
9 summary(int_model)
10
11 # run the LRT
12 anova(add_model,int_model, test = "LRT")
```

Table 2: Comparison of Additive and Interaction Models

	Additive (1)	Interaction (2)
Countries (N)	0.005***	0.004***
Sanctions (%)	-0.017***	-0.019***
Countries x Sanctions		0.00002
Constant	-0.272***	-0.257***
Observations	8,500	8,500
AIC	11,607.830	11,609.620

Through the above table in comparison, we can see slight change in intercept and coefficients between additive and interaction models.

The results of Likelihood Ratio Test shows that p -value = 0.6475, which is greater than the common threshold significance level 0.05. We don't have sufficient evidence to reject H_0 , or there is no statistical evidence that the effect of number of countries depends on sanction levels. We can thus conclude that an interaction term is not appropriate, and the additive model fits the data sufficiently well. It's not necessary to change answers in Question 2.

Analysis of Deviance Table
Model 1: choice ~ countries_n + sanctions_n
Model 2: choice ~ countries_n * sanctions_n
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1 8497 11602
2 8496 11602 1 0.20905 0.6475