Problem Set 2

Applied Stats II

Due: February 18, 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 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.

2. If any of the explanatory variables are statistically 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)

   (b) For the policy in which very few countries participate [20 of 192], how does increasing sanctions from 5% to 15% change the odds that an individual will support the policy? (Interpretation of a coefficient)

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

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.

Task 1: Additive Logistic Regression Model

In order to fulfil the task, I fit an additive model as outlined above. Before this, however, I took a quick look at the data. As can be seen in the code below, I decided to recode the input variables. While I believe that RStudio actually takes the first category as the reference category, I decided to explicitly select the reference category myself (in case my assumption was wrong). To do so, I first unordered the factor variables *countries* and *sanctions* since the command *relevel* does not work on ordered factors. Then I set the reference category to "20 of 192" and "None", respectively.

```
# load data
load(url("https://github.com/ASDS-TCD/StatsII_2026/blob/main/datasets/
    climateSupport.RData?raw=true"))

#inspect data
head(climateSupport)
str(climateSupport)
summary(climateSupport)

#check levels
levels(climateSupport$countries)
```

```
levels(climateSupport$sanctions)
#i believe RStudio uses the first category as the reference cat.
#but in case that's actually not true, this is how I would to do it:

#unorder data
climateSupport$countries <- factor(climateSupport$countries,
                                   ordered = FALSE)
climateSupport$sanctions <- factor(climateSupport$sanctions,
                                   ordered = FALSE)
#set reference categories
climateSupport$countries <- relevel(climateSupport$countries, ref = "20 of 192")
climateSupport$sanctions <- relevel(climateSupport$sanctions, ref = "None")

#check again
levels(climateSupport$countries)
levels(climateSupport$sanctions)
```

As a next step, I specified the additive model using the *glm* function. Specifically, I regressed *choice* on *countries* and *sanctions*, using *binomial* as the distribution for the model.

```
model_add <- glm(choice ~ countries + sanctions,
                 family = binomial(link = logit),
                 data = climateSupport)

summary(model_add)
```

```
#Stargazer for Latex Write-Up
stargazer(model_add,
          title = "Logistic Regression: Support for An Environmental Policy (International)",
          type = "latex")
```

The following table was the result:

Table 1: Logistic Regression: Support for An Environmental Policy (International)

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

As a next step, I focused on testing the global null hypothesis using *anova*. Specifically, the null hypothesis in this scenario is that all slopes are equal to zero, whereas the alternative hypothesis is that at least one coefficient is unequal to zero. In order to assess this, we need a model to compare the additive model to (=the Null Model):

```
##Global Null-Hypothesis
#H0: All slopes = 0
#H1: At least one coefficient is different from zero

#First: need Null Model (=Intercept only)
model_null <- glm(choice ~ 1,
                  family = binomial,
                 data = climateSupport)

#anova: null model vs full model
anova_add <- anova(model_null, model_add, test = "LRT")
```

```
Model 1: choice ~ 1
Model 2: choice ~ countries + sanctions
Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
2      8494      11568  5   215.15 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```r
stargazer(anova_add,
          title = "Anova: Global Null Hypothesis Testing",
          summary = FALSE,
          type = "latex")
```

Table 2: Anova: Global Null Hypothesis Testing

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	8,499	11,783.410			
2	8,494	11,568.260	5	215.150	0

In Conclusion: Firstly, the global null hypothesis test shows that the H_0 can be rejected since $p-value < 0.05$. This means that at least one of the coefficients is significantly related affects the outcome variable (=whether an individual supports the policy in question). Secondly, let's take a closer look at the actual output of the logistic regression model.

1. When 20 (out of 192) countries participate and there are no sanction, the expected odds that an individual supports the policy is $exp(-0.273) = 0.761$ (baseline odds ratio). This means that, on average, individuals do not support the policy in question.

2. Regarding **countries**: Broadly speaking, an increase in participating countries is associated with an increase in the probability that an individual supports the policy in question. Also, please note that the reference category for the following explanations is "20 out of 192" countries. When 80 countries participate, the log-odds increase by 0.336 (holding everything else constant). In other words, individuals are 1.4 times as likely to support the policy ($exp(0,336) = 1.4$). Similarly, the log-odds increase by 0.648 (holding everything else constant). This means that individuals are almost twice as likely to support a policy if most countries participate.

3. Regarding **sanctions**: The relationship between sanctions and choice is a bit more complex. Compared to no sanctions, a 5% penalty is associated with an 0.192 increase in the log-odds (all else constant). Thus, individuals are roughly 1.21 times as likely to support a policy, when compared to no sanctions. However, higher sanctions are associated with a decrease in the log-odds by 0.133 (15% sanctions) and 0.304 (20%

sanctions), respectively. Therefore, when compared to the reference category and holding all else constant, individuals are less likely to support a policy in these cases than when there are no sanctions at all.

4. **Overall**, then, this suggest that international participation has a significant, positive effect on the probability that individuals support a given policy, whereas sanctions have a significant but mixed effect. While the model obviously cannot speak to this, I could imagine that the positive effect on participation could be related to the perceived legitimacy of a policy. In comparison, sanctions might be viewed as too strict if they exceed a certain limit.