

# Problem Set 2 - Applied Stats II

Sarah Magdihs

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

To fulfil the task, I fit an additive model in which I regress *choice* on *countries* and *sanctions*. Initially, I checked out the data and decided to explicitly set a reference category. Since the factors were ordered, I believe R would have taken the first category as reference anyway, but this way I can be absolutely sure. 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.

```
1 # load data
2 load(url("https://github.com/ASDS-TC/StatsII-2026/blob/main/datasets/
  climateSupport.RData?raw=true"))
3
4 #inspect data
5 head(climateSupport)
6 str(climateSupport)
7 summary(climateSupport)
8
9 #check levels
10 levels(climateSupport$countries)
11 levels(climateSupport$sanctions)
```

```

12 #i believe RStudio uses the first category as the reference cat.
13 #but in case that's actually not true, this is how I would to do it:
14
15 #unordered data
16 climateSupport$countries <- factor(climateSupport$countries ,
17                                   ordered = FALSE)
18 climateSupport$sanctions <- factor(climateSupport$sanctions ,
19                                   ordered = FALSE)
20 #set reference categories
21 climateSupport$countries <- relevel(climateSupport$countries , ref = "20 of 192
22                                   ")
23 climateSupport$sanctions <- relevel(climateSupport$sanctions , ref = "None")
24
25 #check again
26 levels(climateSupport$countries)
27 levels(climateSupport$sanctions)

```

Then, I specified the additive model using the *glm* function, using *binomial* as the distribution for the model since the output is binary (and we are thus fitting a logistic regression model).

```

1 model_add <- glm(choice ~ countries + sanctions ,
2                  family = binomial(link = logit) ,
3                  data = climateSupport)
4
5 summary(model_add)

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

```

As a result, Table 1 was provided by R. Before actually discussing its implications, I want to first also discuss the Global Null Hypothesis Test. Specifically, the null hypothesis 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; includes only the intercept). To do so, I did the following:

```

1 ##Global Null-Hypothesis
2 #H0: All slopes = 0
3 #H1: At least one coefficient is different from zero
4
5 #First: need Null Model (=Intercept only)
6 model_null <- glm(choice ~ 1 ,
7                   family = binomial ,
8                   data = climateSupport)
9
10 #anova: null model vs full model
11 anova_add <- anova(model_null , model_add , test = "LRT")

1 stargazer(anova_add ,

```

```

2 title = "Anova: Global Null Hypothesis Testing",
3 summary = FALSE,
4 type = "latex")

```

Through this process, I got Table 1 and 2:

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

	<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

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

Additonally, here is the verbatim output, in order to get a better sense of the p-value:

```

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

```

## Interpretation of Coefficients:

### Global Null Hypothesis Test:

I compared the additive model to a null model using a likelihood ratio test. The output shows that we ought to reject the null hypothesis since the p-value is smaller than 0.05. This means that at least one predictor significantly affects individuals' policy support.

### Interpretation of Coefficients of the Additive Model:

**Intercept:** When 20 countries participate and there are no sanctions, the expected odds that an individual supports the policy are  $e^{-0.273} = 0.761$  (baseline odds ratio), indicating that, on average, individuals do not support the policy.

**Countries:** Generally, an increase in participating countries is associated with an increase in the probability that an individual supports the policy (reference category = 20 out of 192). When 80 (rather than 20) countries participate, the log-odds increase by 0.336 (all else constant). This means that individuals are 1.4 times as likely to support the policy ( $\exp(0.336) = 1.4$ ). Similarly, increasing participation from 20 to 160 countries is associated with an increase in log-odds by 0.648 (approximately 1.91 times as likely to support;  $\exp(0.648) = 1.91$ )

**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). This means that individuals are roughly 1.21 times as likely to support a policy under these circumstances. In contrast, higher sanctions (when compared to the reference category) reduce support for the policy. Specifically, 15% sanctions are associated with a decrease in the log-odds by 0.133 ( $1 - \exp(-0.133) \approx 0.125$ , meaning a decrease by 12.5%) and 20% sanctions with a decrease by 0.304 ( $1 - \exp(-0.304) \approx 0.26$ )

**Overall,** then, this suggests 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.

## Task 2: Sanctions and Probabilities

### 2a and 2b:

Task 2a and 2b can be answered simultaneously. The odds ratio does not change, since we are working with an additive model. Thus, we assume that the effect does not depend on the number of countries that participate.

I calculated how increasing the sanctions from 5% to 15% changes the odds that an individual will support the policy by 'isolating' the relevant coefficients and then exponentiating their difference.

```
1 ##2a: Change in odds
2
3 #check model coefficients
4 coef(model_add)
5
6 #I need coefficient for sanctions 5% and 15%
7 beta_5 <- coef(model_add)["sanctions5%"]
8 beta_15 <- coef(model_add)["sanctions15%"]
9
10 #Calculate Odds Ratio
11 odds_ratio <- exp(beta_15 - beta_5)
12 odds_ratio
13
14 #because I don't want to do the math myself
15 decrease <- 1 - odds_ratio
16 decrease
17
18 ##2b is the same answer since its an additive model
```

As a result, these were the outputs:

```
> odds_ratio
sanctions15%
0.7224531
> decrease
sanctions15%
0.2775469
```

**Interpretation:** Increasing sanctions from 5% to 15% decreases the odds of supporting the policy by about 27.7% (independent of how many countries participate).

### 2c:

Moving on , I estimated the probability that an individual will support the policy when 80 of 192 countries participate and there are no sanctions by using the *predict* function:

```
1 ##2c: estimated probailty for 80 countries + no sanctions
2 est_prob_80 <- predict(model_add,
3                          newdata = data.frame(countries = "80 of 192", sanctions
4                                                = "None"),
```

```

4                                     type = "response")
5 est_prob_80

```

As a result, RStudio provided this output:

```

> est_prob_80
1
0.5159191

```

**Interpretation:** Given 80 participating countries with no sanctions, the estimated probability of support is 51.6%.

## Task 3: Additive vs. Interaction Model

Yes, the answer to Question 2a and 2b could change, if we were to include an interaction term. Unlike an additive model (which assumes that the effect of sanctions is not dependent on countries), a model with interaction terms would expect that the effect of sanctions depends on the number of participating countries.

To be sure whether the results change, we would have to test whether the interaction actually improves our model. To do this, we can compare both models using *anova*.

```

1 #need: interaction model
2 model_int <- glm(choice ~ countries * sanctions ,
3                 data = climateSupport ,
4                 family= binomial(link = "logit"))
5
6 summary(model_int)
7
8 #Stargazer output
9 stargazer(model_int ,
10          title = "Interaction Model" ,
11          type = "latex")
12
13 #anova: additive model vs interaction model
14 anova_int <- anova(model_add, model_int, test = "Chisq")
15 anova_int
16
17
18 #stargazer output for Write-up
19 stargazer(anova_int ,
20          summary = FALSE,
21          title = "Anova: Additive vs. Interaction" ,
22          type = "latex")

```

Using the code above, R provides us with the following output:

Table 3: Anova: Additive vs. Interaction

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	8,494	11,568.260			
2	8,488	11,561.970	6	6.293	0.391

**Interpretation:** The p-value is bigger than 0.05, so we fail to reject the null hypothesis, meaning that adding the interaction term does not improve model fit. Therefore, the additive model is sufficient, and the odds ratio results from Task 2 remain the same in this case.