Problem Set 2

Applied Stats II
Shelly Veal-Upham
25337422

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.

```
1  df <- data %>% mutate(
2    sanctions = factor(sanctions, ordered = FALSE),
3    countries = factor(countries, ordered = FALSE))
```

```
1  m_1 <- glm(data = df, choice ~ sanctions + countries,
2      family = binomial)
```

Table 1: Factors Impacting Support for Climate Policy (Additive)

	Dependent variable:
	Choice
5 Percent Sanctions	0.192***
	(0.062)
15 Percent Sanctions	−0.133**
	(0.062)
20 Percent Sanctions	−0.304***
	(0.062)
80 Participating Countries	0.336***
	(0.054)
160 Participating Countries	0.648***
	(0.054)
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

Testing against the null hypothesis:

H_A: At least one of the covariates predicts our outcome variable better than the null model (intercept only variable fit).
H_N: None of the predictors contribute significantly to predicting the outcome variable better than the null model.

`glm(data = df, choice ~ 1, family = binomial)` will produce our null model:

```
1 m_null <- glm(data = df, choice ~ 1, family = binomial)
2 m1_v_mnull <- anova(m_1, m_null, test = "LRT")
```

Table 2: ANOVA Comparison of Additive and Null Models

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,496.500	3.536	8,494	8,499
Resid. Dev	2	11,675.830	152.134	11,568.260	11,783.410
Df	1	-5.000		-5	-5
Deviance	1	-215.150		-215.150	-215.150
Pr($>$Chi)	1	0.000		0	0

The resulting p-value of this ANOVA is $2.2e^{-16}$, suggesting significance in the use of at least one of our covariates in modeling the outcome variable.

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)

   For the policy in which nearly all countries participate [160 of 192], increasing sanctions from 5% to 15% results in a 0.325 ($\beta_2 - \beta_1$) decrease in the log odds of that individual supporting the policy. In other terms, we would anticipate that individuals in the 15% sanctions group would be $e^{-0.325} = 0.723$ times as likely as (or 27.7% less likely than) the 5% sanctions group to support the policy.

   (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)

   For the policy in which very few countries participate [20 of 192], increasing sanctions from 5% to 15% results in the same 0.325 ($\beta_2 - \beta_1$) decrease in the log odds of that individual supporting the policy. In other terms, we would anticipate that individuals in the 15% sanctions group would be $e^{-0.325} = 0.723$ times as likely as (or 27.7% less likely than) the 5% sanctions group to support the policy, given constancy in the `countries` variable.

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

   The estimated probability that an individual will support a policy if there are 80 participating countries and no sanctions is 0.425 (42.5%):

$$\beta_4 = 0.336 = ln(\tfrac{p}{1-p})$$
$$e^{\beta_4} = e^{0.336} = 0.738 = \tfrac{p}{1-p}$$
$$p = \tfrac{0.738}{1.738} = 0.425$$

3. Would the answers to 2a and 2b potentially change if we included an interaction term in this model? Why?

Yes the answers to 2a and 2b could potentially change if the interaction is significant – presently, as we have fit an additive model, the impact of a change in one of our covariates is independent of the value of the other covariate (so long as it remains constant, the impact is constant). If we were to include an interaction term (and it proved significant in predicting our variable of outcome), there could be a difference in the answers to questions 2a and 2b, as we would model for change in the impact of increases/decreases in sanctions across variation in the `countries` variable.

- Perform a test to see if including an interaction is appropriate.

Table 3: ANOVA Comparison of Additive and Interactive Models

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,491.000	4.243	8,488	8,494
Resid. Dev	2	11,565.110	4.450	11,561.970	11,568.260
Df	1	6.000		6	6
Deviance	1	6.293		6.293	6.293
Pr(>Chi)	1	0.391		0.391	0.391

Here the ANOVA results in a p-value of 0.391. As 0.391 is much greater that 0.05, we cannot reject the null hypothesis that there exists no significant difference in the interactive and additive models. This suggests it is best to retain the simpler additive model.