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.

   Since the response is binary rather than continuous (1 if agree with policy and 0 if not) to fit an additive model it would be preferable to use logistic regression rather than OLS. Logistic regression bound predicted probabilities between possible outcomes [0,1] and does not require the same assumptions as OLS such as constant variance and normal errors. Instead we can use a logistic model that determines the probability of getting a specific outcome conditional on X, writing as a logistic function of X_i

   $$\pi_i = P(Y_i = 1|X_i) = E(Y_i|X_i) = \frac{e^{\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}}}$$

   By taking a log transformation we return the log odds (logit function) that is linearly associated with covariates. This is the canonical link for the Bernoulli distribution is used for MLE estimation as it transforms bounded (0-1) probabilities to values that have infinite range, where the logit of p is linear in X.

   $$log(\frac{\pi_i}{1 - \pi_i}) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}$$

   This enables us to set up a GLM that uses maximum likelihood estimation instead of least squares to determine the best parameter estimates that predicts observed outcomes. The log odds of an individual agreeing with the policy (choice = 1) is modelled below where the reference category is $Countries 20/192 and Sanctions None$

   $$log(\frac{p}{1 - p}) = \beta_0 + \beta_1 Countries 80/192 + \beta_2 Countries 160/192$$

   $$+ \beta_3 Sanctions 5\% + \beta_4 Sanctions 15\% + \beta_5 Sanctions 20\%$$

   Using contr.treatment factors with many levels are automatically converted into a sequence of dummies (as above).

```
1  load(url("https://github.com/ASDS-TCD/StatsII_2026/blob/main/datasets
      /climateSupport.RData?raw=true"))
2
3  str(climateSupport)
4  table(is.na(climateSupport)) #check for missingness
5
```

```r
6 contrasts(climateSupport$countries) <- contr.treatment #quick method
      to convert factor into set of dummys
7 contrasts(climateSupport$sanctions) <- contr.treatment #autoadapt to
      one for each type
8
9 m_add <-  glm(
10   choice ~ countries + sanctions, #additive
11   data = climateSupport,
12   family = binomial(link = "logit") #specify link types
13 )
14 summary(m_add)
15 stargazer::stargazer(m_add,
16                      type = "latex", #output code for use in latex
17                      title = "Impact of Number of Countries and
      Sanctions on Choice",
18                      covariate.labels = c("Countries 80 of 192", "
      Countries 160 of 192", "Sanctions 5%", "Sanctions 10%", "
      Sanctions 20%"),
19                      dep.var.labels = "Choice on policy support")
```

Table 1: Impact of Number of Countries and Sanctions on Choice

	Dependent variable:
	Choice on policy support
Countries 80 of 192	0.336***
	(0.054)
Countries 160 of 192	0.648***
	(0.054)
Sanctions 5%	0.192***
	(0.062)
Sanctions 10%	−0.133**
	(0.062)
Sanctions 20%	−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

The returning model finds every individual coefficient to be statistically significant to the 0.05 level. This means there is evidence that countries and sanctions are associated with the log-odds of policy choice, with changes in each predictor having a multiplicative change in the odds of policy agreement holding other variables constant.

To test the significance of model we want to test the model against a global null that none the predictors have an effect on the outcome.

H_0 : All $\beta_i = 0$
H_1 : At least one of the $\beta \neq 0$

To test this null we use a likelihood ratio test (LRT) that compares the full fitted model with a null model (with no explanatory variables) to compute a test statistic and p-value.

```
m_null <- glm(
  choice ~ 1, #remove all predictors and replace it with 1 i.e. no
    slopes
  data = climateSupport,
  family = binomial(link = "logit") #logit link still
)

anova(m_null, m_add, test = "LRT") #global LRT test against null with
    no covariates
```

```
Analysis of Deviance Table

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
```

Since $p = 2.2e - 16$ is significantly lower than the standard threshold $\alpha = 0.05$ we have sufficient evidence to reject the global null and accept that at least one of the predictors (countries or sanctions) affects the probability that participants will agree with the policy. Moreover we can see in the estimate table that every covariate is statistically impactful on the probability of the agreeing with the policy.

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)

To calculate the change in odds need to calculate the log odds of both sanctions conditions for a policy with 160 countries partipating using the estimated coefficients, all else held constant.

160 of 192 with 5% sanctions : $-0.27266 + 0.64835 + 0.19186 = 0.56755$
160 of 192 with 15% sanctions: $-0.27266 + 0.64835 - 0.13325 = 0.24244$
Difference in log odds: $\Delta_{LO} = 0.24244 - 0.56755 = -0.32511$
(intercept and countries estimate cancel out)

Odds ratio: $OR = e^{\Delta_{LO}} = e^{-0.32511} = 0.72244788$

For a policy with 160 of 192 countries participating, going from 5% to 15% sanctions reduces the odds of policy agreement, on average, by a multiplicative factor of 0.72244788 holding all else constant (i.e. decreasing the odds by 27.75%).

(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)

The change in odds will be the same as above as the model is additive on the log-odds scale. The different in log-odds i.e. moving from 5% to 15% sanctions is independent of the category of number of countries in the policy has therefore cancels out and is no difference between 160 or 20 of the 192. The change would still therefore be that going from 5% to 15% sanctions when very few countries participate (20 of 192) reduces the odds of policy agreement, on average, by a multiplicative factor of 0.72244788 holding all else constant.

The odds ratio for a change in one factors is constant across all levels of the other factors in this additive model and would only change if the model included an interaction term.

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

To calculate the estimated probability need to first calculate the log odds of agreement for when there 80/192 countries and no sanctions, all else held constant. Since no sanctions is the baseline category we just need the intercept and

6

effect moving from 20 to 80 countries.

80 of 192 with No sanctions: $\beta_0 + \beta_1 = -.27266 + .33636 = 0.0637$

We can then convert the log odds to the probability using:

$$P(Y = 1 | Countries80of192, SanctionsNone) = \frac{e^{logit(P)}}{1 + e^{logit(P)}} = \frac{e^{.0637}}{1 + e^{.0637}} = 0.515991$$

The estimate probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions is 0.5159 i.e 51.59%.

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.

Yes the answers would potentially change with an interaction included. The log-odds change when moving from sanction categories now would depend on the number of countries participating as the interaction term would add a different amount for each combination. This would greatly increase the complexity of the model interpretations and should only be done if including the interaction improves model fit to a degree that it is worth the additional complexity. This can be tested using a partial likelihood ratio test (following a central Chi-Squared distribution under the null being true).

H_0 : The interaction coefficients are zero
H_1: At least one of the interaction terms are non-zero.

If the null is true it means adding the interaction terms added are not significantly different from just the additive model. First we must run an new model with interactions include then compare using an ANOVA.

```
1  m_int <-  glm( #interaction model version
2    choice ~ countries + sanctions + countries*sanctions, #adding in
       interaction term
3    data = climateSupport,
4    family = binomial(link = "logit")
5  )
6
7  anova(m_add, m_int, test = "LRT") #partial LRT test
```

The results of the test determine a p-value $p = 0.3912$ that is larger than the standard threshold $\alpha = 0.05$, thus providing sufficient evidence that the interactive model fit is not significantly better than the additive model. Therefore, it would be more appropriate to continue with the additive model rather than adding an interaction for parsimony and ease of interpretation.

```
Analysis of Deviance Table

Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries + sanctions + countries * sanctions
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8494      11568
2      8488      11562  6   6.2928   0.3912
```

Moreover when we see the results of the interaction model there are many estimates and none of the interaction terms appear to be individually significant.

```
1  stargazer::stargazer(m_int,
2                       type = "latex", #output code for use in latex
3                       title = "Impact of Number of Countries and Sanctions
     on Choice with Interaction",
4                       dep.var.labels = "Choice on policy support")
```

Table 2: Impact of Number of Countries and Sanctions on Choice with Interaction

	Dependent variable:
	Choice on policy support
countries80 of 192	0.376***
	(0.106)
countries160 of 192	0.613***
	(0.108)
sanctions5%	0.122
	(0.105)
sanctions15%	−0.097
	(0.108)
sanctions20%	−0.253**
	(0.108)
countries80 of 192:sanctions5%	0.095
	(0.152)
countries160 of 192:sanctions5%	0.130
	(0.151)
countries80 of 192:sanctions15%	−0.052
	(0.152)
countries160 of 192:sanctions15%	−0.052
	(0.153)
countries80 of 192:sanctions20%	−0.197
	(0.151)
countries160 of 192:sanctions20%	0.057
	(0.154)
Constant	−0.275***
	(0.075)
Observations	8,500
Log Likelihood	−5,780.983
Akaike Inf. Crit.	11,585.970

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