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.

Answer 1

```
# load data
climate_data <- load(url("https://github.com/ASDS-TCD/StatsII_2026/blob/main/datase

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

summary(logit_add)      # this replaces summary(all) from part 1
```

We fit the following additive logistic regression model:

$$\text{logit}\{\Pr(Y_i = 1)\} = \beta_0 + \beta_1\,\text{countries.L}_i + \beta_2\,\text{countries.Q}_i + \beta_3\,\text{sanctions.L}_i + \beta_4\,\text{sanctions.Q}_i + \beta_5\,\text{sanctio}$$

The R output for the additive model is:

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.005665   0.021971  -0.258 0.796517
countries.L  0.458452   0.038101  12.033  < 2e-16 ***
countries.Q -0.009950   0.038056  -0.261 0.793741
sanctions.L -0.276332   0.043925  -6.291 3.15e-10 ***
sanctions.Q -0.181086   0.043963  -4.119 3.80e-05 ***
sanctions.C  0.150207   0.043992   3.414 0.000639 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11783  on 8499  degrees of freedom
```

```
Residual deviance: 11568  on 8494  degrees of freedom
AIC: 11580

Number of Fisher Scoring iterations: 4
```

The global null hypothesis that "countries and sanctions have no effect" is

$$H_0: \ \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0,$$

i.e. the probability of supporting the policy is unrelated to the number of countries and the sanction level.

We compare this additive model to a null (intercept–only) model using a likelihood ratio test. The null deviance is 11783 on 8499 degrees of freedom and the residual deviance of the additive model is 11568 on 8494 degrees of freedom. This yields a likelihood–ratio statistic of approximately $\chi^2 \approx 215$ on 5 degrees of freedom with p–value < 0.001, so we reject H_0 and conclude that the set of `countries` and `sanctions` variables jointly improves prediction of support.

From the coefficient table we see that:

- countries.L is highly significant and positive (Estimate ≈ 0.458, $p < 2 \times 10^{-16}$), so moving along the main linear contrast in the number of participating countries increases the log–odds, and thus the probability, of supporting the policy.
- countries.Q is not significant ($p \approx 0.79$), suggesting little curvature beyond that linear trend in the number of countries.
- All three sanctions contrasts are significant: sanctions.L is negative and highly significant (Estimate ≈ -0.276, $p \approx 3.1 \times 10^{-10}$), sanctions.Q is negative and significant ($p \approx 3.8 \times 10^{-5}$), and sanctions.C is positive and significant ($p \approx 0.0006$). This indicates that the pattern of support varies systematically with the level of sanctions, and not in a purely linear way.

In words, there is strong evidence that both the number of participating countries and the sanction level are associated with support for the climate agreement, jointly and individually. As the number of countries increases, people are more likely to support the policy, and changes in sanctions levels also meaningfully change support, with a non–linear pattern captured by the orthogonal polynomial contrasts for sanctions.

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)

```
# build two hypothetical cases and look at linear predictor difference
lp_160_5  <- predict(logit_add, newdata = data.frame(
  countries = "160 of 192",
  sanctions = "5%"
), type = "link")

lp_160_15 <- predict(logit_add, newdata = data.frame(
  countries = "160 of 192",
  sanctions = "15%"
), type = "link")

log_OR_160_5_to_15 <- lp_160_15 - lp_160_5
OR_160_5_to_15     <- exp(log_OR_160_5_to_15)
OR_160_5_to_15
```

```
Using the fitted model in R, we obtain an estimated odds ratio
OR_160_5_to_15 = 0.7224531.
```

Answer 2a

- Group 1: 160 of 192 countries participate, sanctions = 5%.
- Group 2: 160 of 192 countries participate, sanctions = 15%.

Their logits can be written as

$$\log\left(\frac{\pi_{160,5}}{1-\pi_{160,5}}\right) = \beta_0+\beta_1\,\text{countries.L}_{160}+\beta_2\,\text{countries.Q}_{160}+\beta_3\,\text{sanctions.L}_5+\beta_4\,\text{sanctions.Q}_5+\beta$$

$$\log\left(\frac{\pi_{160,15}}{1-\pi_{160,15}}\right) = \beta_0+\beta_1\,\text{countries.L}_{160}+\beta_2\,\text{countries.Q}_{160}+\beta_3\,\text{sanctions.L}_{15}+\beta_4\,\text{sanctions.Q}_{15}$$

The log of the odds ratio comparing 15% versus 5% sanctions (for 160 countries) is

$$\log(\text{OR}_{160,\,5\%\to15\%}) = \log\left(\frac{\pi_{160,15}/(1-\pi_{160,15})}{\pi_{160,5}/(1-\pi_{160,5})}\right) = \left[\log\left(\frac{\pi_{160,15}}{1-\pi_{160,15}}\right)\right]-\left[\log\left(\frac{\pi_{160,5}}{1-\pi_{160,5}}\right)\right].$$

All of the countries terms cancel, because they are the same in both groups. What remains is a linear combination of the sanctions contrasts, i.e. β_3, β_4 and β_5 determine the difference in log–odds between 5% and 15% sanctions.

Using the fitted model in R, we obtain an estimated odds ratio

$$\widehat{\text{OR}}_{160,\,5\%\to15\%} \approx 0.72.$$

This means that increasing sanctions from 5% to 15% multiplies the odds of supporting the policy by a factor of 0.72, i.e. it reduces the odds of support by about 28%, when 160 of 192 countries participate.

(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)

```
lp_20_5  <- predict(logit_add, newdata = data.frame(
  countries = "20 of 192",
  sanctions = "5%"
), type = "link")

lp_20_15 <- predict(logit_add, newdata = data.frame(
  countries = "20 of 192",
  sanctions = "15%"
), type = "link")

log_OR_20_5_to_15 <- lp_20_15 - lp_20_5
OR_20_5_to_15     <- exp(log_OR_20_5_to_15)
OR_20_5_to_15
```

$\widehat{\mathrm{OR}}_{20,\,5\%\to15\%} = 0.7224531$

Answer 2(b)

Now consider the analogous comparison when only 20 of 192 countries participate:

- Group 1: 20 of 192 countries, sanctions = 5%.
- Group 2: 20 of 192 countries, sanctions = 15%.

The fitted model can be written generically as

$$\operatorname{logit}(\pi) = \beta_0 + f_{\text{countries}}(\text{countries}) + f_{\text{sanctions}}(\text{sanctions}),$$

with separate terms for `countries` and `sanctions`, but no countries $\times$ sanctions interaction. When we compare 5% vs. 15% sanctions at 160 countries, the difference in log–odds is

$$\Delta_{160} = f_{\text{sanctions}}(15\%) - f_{\text{sanctions}}(5\%).$$

When we compare 5% vs. 15% sanctions at 20 countries, the difference in log–odds is

$$\Delta_{20} = f_{\text{sanctions}}(15\%) - f_{\text{sanctions}}(5\%)$$

as well, because the countries part is the same in both groups and cancels out in the odds ratio.

Thus

$$\mathrm{OR}_{160,\,5\%\to15\%} = \exp(\Delta_{160}) = \exp(\Delta_{20}) = \mathrm{OR}_{20,\,5\%\to15\%} \approx 0.72.$$

In an additive logit model, the multiplicative change in odds for moving from 5% to 15% sanctions does not depend on the number of participating countries. It is a single global odds ratio. Therefore, both parts 2(a) and 2(b) yield the same estimated odds ratio of about 0.72, interpreted in different country scenarios.

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

```
# Make sure factors have the same levels as in the model
newdat_80_none <- data.frame(
  countries = factor("80 of 192", levels = levels(climateSupport$countries)),
  sanctions = factor("None",      levels = levels(climateSupport$sanctions))
)

p_80_none <- predict(logit_add, newdata = newdat_80_none, type = "response")
p_80_none
```

$$\hat{P}\big(\text{choice} = 1 \mid \text{countries} = 80/192,\ \text{sanctions} = \text{None}\big) = 0.5159191.$$

Answer 2(c)

For a policy with 80 of 192 countries participating and no sanctions, we compute the fitted probability as

$$\hat{P}(\text{choice} = 1 \mid \text{countries} = 80/192,\ \text{sanctions} = \text{None}) = \hat{p}_{80,\text{None}} \approx 0.516.$$

That is, the estimated probability that an individual supports the policy when 80 of 192 countries participate and there are no sanctions is approximately 0.516, or about a 51.6% chance of support.

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.

Problem 3

```
# Additive model
logit_add <- glm(
  choice ~ countries + sanctions,
  data   = climateSupport,
  family = binomial(link = "logit")
```

```
)

# Model with interaction
logit_int <- glm(
  choice ~ countries * sanctions,
  data   = climateSupport,
  family = binomial(link = "logit")
)

# Likelihood ratio test
anova(logit_add, logit_int, test = "Chisq")
```

To assess whether including an interaction between `countries` and `sanctions` is appropriate, we compare the additive model

$$\text{Model 1:} \quad \text{choice} \sim \text{countries} + \text{sanctions}$$

to the interaction model

$$\text{Model 2:} \quad \text{choice} \sim \text{countries} * \text{sanctions}$$

using a likelihood ratio test.

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

Table 1: Analysis of deviance table for additive vs. interaction model.

From the analysis of deviance:

- Additive model: residual deviance = 11568 on 8494 df.
- Interaction model: residual deviance = 11562 on 8488 df.
- Likelihood ratio test: ΔDev = 6.29 on 6 df, with p–value = 0.3912.

The null hypothesis is

$$H_0: \text{ no interaction between countries and sanctions,}$$

i.e. the effect of sanctions on support is the same across all levels of `countries`. The test statistic is small and the p–value is $p = 0.39 > 0.05$, so we fail to reject H_0. There is no strong evidence that the sanctions effect depends on the number of participating countries.

Therefore, adding the countries × sanctions interaction does not significantly improve model fit. The additive model is adequate, and it is reasonable that the odds ratio for moving from 5% to 15% sanctions is the same in parts 2(a) and 2(b).