# Problem Set 2

## Applied Stats II

## Due: February 18, 2024

## 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 Sunday February 18, 2024. 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**

```
 1    # View the data structure
 2   str(climateSupport)
 3
 4   climateSupport
 5
 6    # Fit the logistic regression model
 7   model <- glm(choice ~ countries + sanctions, family = binomial(link="
        logit"), data = climateSupport)
 8
 9  # Summary output
10   summary(model)
11
12  > summary(model)
13
14   Call:
15   glm(formula = choice ~ countries + sanctions, family = binomial,
16       data = climateSupport)
17
18   Coefficients:
19                 Estimate Std. Error z value Pr(>|z|)
20   (Intercept)  -0.005665   0.021971   -0.258 0.796517
21   countries.L   0.458452   0.038101   12.033  < 2e-16 ***
22   countries.Q  -0.009950   0.038056   -0.261 0.793741
23   sanctions.L  -0.276332   0.043925   -6.291 3.15e-10 ***
24   sanctions.Q  -0.181086   0.043963   -4.119 3.80e-05 ***
25   sanctions.C   0.150207   0.043992    3.414 0.000639 ***
26   ---
27   Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1
                     1
28
29   (Dispersion parameter for binomial family taken to be 1)
30
31       Null deviance: 11783  on 8499  degrees of freedom
32   Residual deviance: 11568  on 8494  degrees of freedom
33   AIC: 11580
34
35   Number of Fisher Scoring iterations: 4
36
37  #fitted logistic regression model
38
```

```
39  logit(p) = −0.005665 + 0.458452 * countries.L − 0.009950 * countries.Q −
       0.276332 * sanctions.L − 0.181086 * sanctions.Q + 0.150207 * sanctions
       .C
40
41  #H0: there is no association between the number of participating
       countries or the level of sanctions and the likelihood of an
       individual supporting the policy.
42  #The p−value for countries.L is < 2e−16. This very small pvalue suggests
       strong proof against the null hypothesis and shows that the number of
       participating countries (countries.L) has a significant association
       with the likelihood of supporting the policy.
```

2. If any of the explanatory variables are 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)

   **Answer 2.a**

   We can check to the coefficient estimate for sanctions.L in the regression model

   sanctions.L -0.276332 0.043925 -6.291 3.15e-10 ***

   The coefficient estimate for sanctions.L is -0.276332. This means that for a one-unit increase in the level of sanctions (from 5

   ```
   1    #odds ratio
   2  odds_ratio = exp(−0.276332)
   3  odds_ratio
   4  1−odds_ratio
   ```

   The odds ratio of 0.7583312 indicates that increasing sanctions from 5

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

   **Answer 2.b**

   ```
   1    probability <− 1 / (1 + exp(−(−0.005665 + 0.458452)))
   ```

   probability [1] 0.6113017

   Therefore, the estimated probability that an individual will support the policy when there are 80 out of 192 countries participating with no sanctions is approximately 0.6 This means that the probability of support is around 60

   (c) Would the answers to 2a and 2b potentially change if we included the interaction term in this model? Why?

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