# Problem Set 2

Applied Stats II

Due: February 19, 2023

## 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 19, 2023. 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.csv` 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.

   Code:

```
# Regularize data

df <- mutate(climateSupport,
  # choice column
  choice = case_when(
   choice == "Supported" ~ 1,
   TRUE ~ 0
  ),

  # countries column
  countries = sub(" .*", "", countries),
  countries = as.numeric(countries) / 192,

  #sanctions column
  sanctions = sub("None", 0, sanctions),
  sanctions = sub("%", "", sanctions),
  sanctions = as.numeric(sanctions) / 100
)

head(df)

# GLM
glm_climate <- glm(choice ~ countries + sanctions,
                   family = binomial(link = 'logit'),
                   data = df
               )
summary(glm_climate)
```

   Results:

```
glm(formula = choice ~ countries + sanctions, family = binomial(link = "logit")
    data = df)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-1.4038  -1.1240  -0.9744   1.1535   1.3950

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.21898    0.04811  -4.552 5.31e-06 ***
countries    0.88428    0.07354  12.025  < 2e-16 ***
sanctions   -1.85725    0.27766  -6.689 2.25e-11 ***
---
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: 11593  on 8497  degrees of freedom
AIC: 11599

Number of Fisher Scoring iterations: 4
```

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)

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

   (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.