Problem Set 2 - Laura McPhillips

Applied Stats II

Due: February 28, 2022

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 class on Monday February 28, 2022. No late assignments will be accepted.

- Total available points for this homework is 80.

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.

    Null hypothesis: Sanctions and participating countries have no effect on support for the policy.

    Output Estimates:
    Intercept = -0.14458 - Sanctions = -0.12353 - Countries = 0.32436
    All p-values are < 0.01, therefore, are statistically significant. So we reject the null hypothesis. Number of participating countries and sanctions do have an effect on policy choice.

    - When explanatory variables, sanctions and countries are 0, the response variable, support for policy is -0.144.
    - For ever unit increase in sanctions, support for policy decreases by 0.123.
    - For every unit increase in the number of participating countries, support for policy increases by 0.32.

    ```
    1  climateSupport$choice <- as.numeric(as.factor(climateSupport$choice))
         -1
    2  climateSupport$sanctions <- as.numeric(as.factor(climateSupport$
         sanctions))-1
    3  climateSupport$countries<- as.numeric(as.factor(climateSupport$
         countries))-1
    4
    5  reg <- glm(choice~ sanctions + countries, data = climateSupport,
         family = "binomial")
    6  summary(reg)
    7
    ```

2. If any of the explanatory variables are significant in this model, then:

$$ln(\frac{P(Y_i=1)}{1-P(Y_i=1)}) = \beta_o + \beta_1 X_{1i} + ... + \beta_k X_{ki}$$

    (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)

```
1  y_1 <- exp(-0.144 + 0.32435*2 - 0.12353*1)
2  y_2 <- exp(-0.144 + 0.32435*2 - 0.12353*2)
3  y_2 - y_1 = -0.17
```

Increasing sanctions from 5% to 15% decreases the odds that an individual will support the policy by 0.17.

(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)

```
1  y_3 <- exp(-0.144 + 0.32435*0 - 0.12353*1)
2  y_4 <- exp(-0.144 + 0.32435*0 - 0.12353*2)
3  y_4 - y_3 = -0.0889
```

Increasing sanctions from 5% to 15% decreases the odds that an individual will support the policy by 0.0889.

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

```
1  y_5 <- exp(-0.144 + 0.32435*1 - 0.12353*0)
2  prob <- y_5/(1+y_5)
3  prob = 0.5449
```

The probability is 0.54 that an individual will support a policy if there are 80 of 192 countries patricipating with no sanctions.

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

The answers to 2a and 2b could potentially change if there is a statistically significant interaction. If number of participating countries and sanctions affect each other, the estimate for each could increase/decrease when the interaction is accounted for.

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

```
1  interact <- glm(choice~. + countries:sanctions,
2               data = climateSupport,
3               family = "binomial")
4  summary(interact)
5
6  y_interact <- -0.148 + 0.328*(countries) - 0.12*(sanctions) -
       0.002455*(countries)*(sanctions)
```

There is a negative interaction of -0.0002 but the p-value of 0.9 > 0.05. This is not statistically significant, so we use an additive model, not an interactive model.