

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 Question 1

```
1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/
  datasets/climateSupport.RData?raw=true"))
3 #fit an additive model using glm() with the family = binomial()
4 model_additive <- glm(choice ~ countries + sanctions, data =
  climateSupport, family = binomial)
5 #examine the summary output
6 summary(model_additive)
7
```

Explanation Answer 1

The global null hypothesis is that all coefficients are zero, and the p -values indicate the significance of each term. To interpret the results, we'll focus on the coefficients for countries and sanctions. If the coefficients are positive, it means that an increase in the explanatory variable leads to an increase in the probability of supporting the policy. The global null hypothesis for the additive model is:

$$H_0 : \beta_{\text{countries}} = \beta_{\text{sanctions}} = 0$$

The p -values will indicate whether the coefficients are significantly different from zero. To describe the results, we'll look at the coefficients and their significance: If the coefficient for countries is positive and significant, it means that as the number of participating countries increases, the probability of supporting the policy also increases. If the coefficient for sanctions is positive and significant, it means that as the sanctions increase, the probability of supporting the policy also increases. Recap: The additive model shows that both the number of participating countries and the level of sanctions are significant predictors of an individual's likelihood of supporting the policy. As the number of participating countries increases, the probability of supporting the policy also increases. Similarly, as the level of sanctions increases, the probability of supporting the policy also increases.

Call:

```
glm(formula = choice ~ countries + sanctions, family = binomial, data = climateSupport)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.844e-03	2.933e-04	9.694	6.77e-16 ***
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

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)

Explanation Answer 2 (a)

If the coefficient for sanctions is positive and significant, it means that increasing sanctions from 5% to 15% will lead to an increase in the probability of supporting the policy.

- (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 Question 2(b)

```
1 # Ensure 'countries' and 'sanctions' are treated as numeric variables
2 climateSupport$countries <- as.numeric(climateSupport$countries)
3 climateSupport$sanctions <- as.numeric(climateSupport$sanctions)
4
5 # Fit the additive model
6 model_additive <- glm(choice ~ countries + sanctions, data =
7   climateSupport, family = binomial)
```

```

8 # Predict the probability for 80 of 192 countries participating with
   no sanctions
9 pred_80_no_sanctions <- predict(model_additive, newdata = data.frame(
   countries = 80, sanctions = 0), type = "response")
10

```

- (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.

Answer Question 2(c)

```

1 # Ensure 'countries' and 'sanctions' are treated as numeric variables
2 climateSupport$countries <- as.numeric(climateSupport$countries)
3 climateSupport$sanctions <- as.numeric(climateSupport$sanctions)
4
5 # Fit the additive model
6 model_additive <- glm(choice ~ countries + sanctions, data =
   climateSupport, family = binomial)
7
8 # Fit the model with interaction term
9 model_interaction <- glm(choice ~ countries * sanctions, data =
   climateSupport, family = binomial)
10
11 # Perform ANOVA test
12 anova_result <- anova(model_additive, model_interaction)
13
14 # Check if interaction term is significant
15 if (anova_result$`Pr(>Chisq)`[2] < 0.05) {
16   cat("The interaction term is significant.\n")
17 } else {
18   cat("The interaction term is not significant.\n")
19 }
20

```

Explanation Answer 2 (c)

If the interaction term is significant, the answers to 2a and 2b may change because the effect of sanctions would depend on the number of participating countries. The interpretation of the coefficients would change, as the effect of sanctions would not be constant across different levels of the number of participating countries. It is significant, it means that the effect of sanctions on the probability of supporting the policy is not the same for all levels of the number of participating countries. The interaction term captures the combined effect of countries and sanctions, and including it in the model most likely will change the significance and interpretation of the individual coefficients.