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.

Code in `PS2_ImeldaFinn.R`

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.

```
load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/datasets/climateSupport.RData?raw=true"))

# choice = 1,2
# countries = 1, 2, 3
# sanctions = 1, 2, 3, 4

csFacs <- climateSupport
csFacs$choice <- relevel(csFacs$choice, "Supported")

```

- 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.

```
mod <- glm(choice ~., family = binomial(link="logit"),
```

   (a) Provide the summary output,

```
    Call:
    glm(formula = choice ~ ., family = binomial(link = "logit"), data
    = csFacs)

    Deviance Residuals:
        Min       1Q   Median       3Q      Max
    -1.4259  -1.1480  -0.9444   1.1505   1.4298

    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


```

   (b) the global null hypothesis:

   H_0: the explanatory variables have no effect on the likelihood of an individual supporting a policy

H_a: one or more of the explanatory variables have some effect on the likelihood of an individual supporting a policy

$\alpha = 0.05$

(c) and p-value. The model was run with no explanatory variables (dummy... TODO). The comparison of the two models is shown in **??**

Table 1:

	Dependent variable:	
	choice	
	(1)	(2)
countries.L	-0.458^{***}	
	(0.038)	
countries.Q	0.010	
	(0.038)	
sanctions.L	0.276^{***}	
	(0.044)	
sanctions.Q	0.181^{***}	
	(0.044)	
sanctions.C	-0.150^{***}	
	(0.044)	
Constant	0.006	0.007
	(0.022)	(0.022)
Observations	8,500	8,500
Log Likelihood	−5,784.130	−5,891.705
Akaike Inf. Crit.	11,580.260	11,785.410
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,496.500	3.536	8,494	8,499
Resid. Dev	2	11,675.830	152.134	11,568.260	11,783.410
Df	1	5.000		5	5
Deviance	1	215.150		215.150	215.150

(d) Please describe the results and provide a conclusion.

Table 3

The e^{β_k}s are all non-zero and the 5% Confidence Intervals do not include 0.(Table 3

The estimates for β_k are all significant at $p = 0.01$ except for `countries.Q`, ie there is no significant difference in likelihood going from 80 to 160 countries.

Table 3:

	lower	estimate	upper
intercept	c(‘(Intercept)‘ = 0.9633)	c(‘(Intercept)‘ = 1.0057)	c(‘(Intercept)‘ = 1.0499)
countries 80	c(countries.L = 0.5867)	c(countries.L = 0.6323)	c(countries.L = 0.6812)
countries 160	c(countries.Q = 0.9374)	c(countries.Q = 1.01)	c(countries.Q = 1.0882)
sanctions 5%	c(sanctions.L = 1.2096)	c(sanctions.L = 1.3183)	c(sanctions.L = 1.4369)
sanctions 10%	c(sanctions.Q = 1.0996)	c(sanctions.Q = 1.1985)	c(sanctions.Q = 1.3064)
sanctions 15%	c(sanctions.C = 0.7894)	c(sanctions.C = 0.8605)	c(sanctions.C = 0.938)

It took 4 iterations to find the maximum likelihood estimates.

the log likelihood is -5,784.130

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.

Table 4:

	Dependent variable:		
	choice		
	(1)	(2)	(3)
countries.L	−0.458***		−0.457***
	(0.038)		(0.038)
countries.Q	0.010		0.011
	(0.038)		(0.038)
sanctions.L	0.276***		0.274***
	(0.044)		(0.044)
sanctions.Q	0.181***		0.182***
	(0.044)		(0.044)
sanctions.C	−0.150***		−0.153***
	(0.044)		(0.044)
countries.L:sanctions.L			0.002
			(0.077)
countries.Q:sanctions.L			−0.134*
			(0.076)
countries.L:sanctions.Q			0.008
			(0.076)
countries.Q:sanctions.Q			−0.093
			(0.076)
countries.L:sanctions.C			−0.095
			(0.076)
countries.Q:sanctions.C			−0.010
			(0.077)
Constant	0.006	0.007	0.004
	(0.022)	(0.022)	(0.022)
Observations	8,500	8,500	8,500
Log Likelihood	−5,784.130	−5,891.705	−5,780.983
Akaike Inf. Crit.	11,580.260	11,785.410	11,585.970
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: ANOVA

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,496.500	3.536	8,494	8,499
Resid. Dev	2	11,675.830	152.134	11,568.260	11,783.410
Df	1	5.000		5	5
Deviance	1	215.150		215.150	215.150

Table 6: ANOVA

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	8,491.000	4.243	8,488	8,494
Resid. Dev	2	11,565.110	4.450	11,561.970	11,568.260
Df	1	6.000		6	6
Deviance	1	6.293		6.293	6.293

Table 7: ANOVA

ANOVA