Problem Set 2

Tianxin Zhang 21352232 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.

Table 1: Addictive Model

	Dependent variable:
	choice
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
	(0.022)
Observations	8,500
Log Likelihood	$-5,784.130$
Akaike Inf. Crit.	11,580.260
Note:	*p<0.1; **p<0.05; ***p<0.01

Interpretation of Model:

```
1
2 mod1 <- glm(choice ~ countries + sanctions, data = climateSupport,
3             family = "binomial")
```

```
summary(mod1)
stargazer(mod1, title="Addictive Model")


# transfer log odds ratios to odds ratios
itc <- exp(-0.006)
itc

cL <- exp(0.458)
cL

cQ <- exp(-0.010)  # p > 0.05
cQ

sL <- exp(-0.276)
sL

sQ <- exp(-0.181)
sQ

sC <- exp(0.150)
sC


## Get exp() here: https://www.educative.io/answers/how-to-calculate-the-
    natural-exponential-in-r-using-exp
```

When there are only 20 countries participating and none sanctions are applied, the odds of recipients to support the climate policy is 0.994 times of the odds of recipients against the policy.

Holding other explanatory variables constant, when there are 80 countries participating, the odds of recipient supporting the climate policy are 1.581 times of the odds of recipients against the policy. The p-value of this coefficient is lower than 0.01. Therefore, we can deny the null hypothesis that there is no correlation between the log odds of supporting climate policy and the condition that there are 80 countries participating.

Holding other explanatory variables constant, when there are 160 countries participating, the odds of recipient supporting the climate policy are 0.990 times of the odds of recipients against the policy. The p-value of this coefficient is higher than 0.1. Therefore, we cannot deny the null hypothesis that there is no correlation between the log odds of supporting climate policy and the condition that there are 160 countries participating.

Holding other explanatory variables constant, when there is a sanction of 5% of the monthly household costs given 2% GDP growth for missing emission reduction tar-

gets, the odds of recipient supporting the climate policy are 0.759 times of the odds of recipients against the policy. The p-value of this coefficient is lower than 0.01. Therefore, we can deny the null hypothesis that there is no correlation between the log odds of supporting climate policy and the condition that there is a sanction of 5% of the monthly household costs given 2% GDP growth for missing emission reduction targets.

Holding other explanatory variables constant, when there is a sanction of 15% of the monthly household costs given 2% GDP growth for missing emission reduction targets, the odds of recipient supporting the climate policy are 0.834 times of the odds of recipients against the policy. The p-value of this coefficient is lower than 0.01. Therefore, we can deny the null hypothesis that there is no correlation between the log odds of supporting climate policy and the condition that there is a sanction of 15% of the monthly household costs given 2% GDP growth for missing emission reduction targets.

Holding other explanatory variables constant, when there is a sanction of 20% of the monthly household costs given 2% GDP growth for missing emission reduction targets, the odds of recipient supporting the climate policy are 1.162 times of the odds of recipients against the policy. The p-value of this coefficient is lower than 0.01. Therefore, we can deny the null hypothesis that there is no correlation between the log odds of supporting climate policy and the condition that there is a sanction of 20% of the monthly household costs given 2% GDP growth for missing emission reduction targets.

Testing the Coefficients

Null Hypothesis: All the coefficients are 0.
Alternative Hypothesis: at least one coefficient is not 0.

```r
#H0: All the coefficients are 0.
# Ha: at least one coefficient is not 0.

mod_null <- glm(choice ~ 1, family = binomial(link = "logit") , data =
    climateSupport)

CST <- anova(mod1, mod_null, test = "LRT")
stargazer(CST, title = "Chi-Square Test")

# we can deny the null hypothesis that all coefficients equal to 0, and
    we can conclude that at least one predictor is reliable.
```

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)

Table 2: Chi-Square Test

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
Pr(>Chi)	1	0.000		0	0

```
# a) # odds change

odds_change <- exp(-0.181) - exp(-0.276)
odds_change # 0.07562243
```

The times of odds of recipients supporting the climate policy against the odds of recipients against the policy changes by 0.076.

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

```
# b) # probability
logit_ns_80 <- -0.181 - 0.006

prob_perc <- exp(logit_ns_80)/(1+exp(logit_ns_80))*100
prob_perc  # 45.34%
```

The probability of recipients supporting the policy with given condition is 45.34

(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.
  First, let's create a model with interaction. See below:

```
# c) Interaction Model or not?

mod2 <- glm(choice ~ countries  + sanctions + countries:sanctions
    , data = climateSupport,
             family =  binomial(link = "logit"))
summary(mod2)
stargazer(mod2, title="Interactive Model")
```

Then, let's conduct a significance for different coefficients.
$H_0 : B_{sacntions|20countries} = B_{sacntions|80countries} = B_{sacntions|160countries}$
$H_a : Effectofsacntionsonreceipients'supportingcliamtepolicyisdifferentbythenumberof$

```

test2 <- anova(mod1, mod2, test="LRT")
stargazer(test2, title="Test: Interactive Model or not")
# p-value is 0.3912
# no significant difference between the two models
```

Table 3: Interactive Model

	Dependent variable:
	choice
countries.L	0.457^{***}
	(0.038)
countries.Q	-0.011
	(0.038)
sanctions.L	-0.274^{***}
	(0.044)
sanctions.Q	-0.182^{***}
	(0.044)
sanctions.C	0.153^{***}
	(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	-0.004
	(0.022)
Observations	8,500
Log Likelihood	$-5{,}780.983$
Akaike Inf. Crit.	11,585.970
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Test: Interactive Model or not

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
Pr(>Chi)	1	0.391		0.391	0.391

Conclusion: The p-value is larger than 0.1. Therefore, there is no significant difference between the two models. We can conclude that it is not appropriate to include interaction in the model.