Problem Set 2

Sara Cid

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 to Question 1:

In this part, we fit a logistic regression where we examine the influence of the number of participating countries and the level of sanctions on individual support for policies. The global null hypothesis we are testing, H_0, is that both variables do not significantly affect policy support. This can be expressed as: $H_0 : \beta_{\text{countries}} = \beta_{\text{sanctions}} = 0$. As such, the alternative hypothesis H_A is that least one of these predictors significantly impacts support likelihood. When fitting the regression model, we are also performing hypothesis tests for all the individual coefficients, using Wald tests.

First, we run the regression, making sure we set our preferred reference categories (in this case, 'None' for sanctions and '20/192 countries', which is the baseline for country support):

```r
# Re-leveling independent variables so that "none" is the reference category for
# sanctions, and 20/192 is the reference category for countries:

# First we must make into unordered factors:

climateSupport$sanctions <- factor(climateSupport$sanctions, ordered = FALSE)
climateSupport$countries <- factor(climateSupport$countries, ordered = FALSE)

# Then we relevel:

climateSupport$sanctions <- relevel(climateSupport$sanctions, ref = "None")
climateSupport$countries <- relevel(climateSupport$countries, ref = "20 of 192")

# Fitting the model (taking code from slides):

climate_logit <- glm(choice ~ countries + sanctions, data = climateSupport,
                     family = binomial(link = logit))
```

Next we perform a likelihood ratio test (LRT) to evaluate the global null hypothesis:

```r
null_logit <- glm(choice ~ 1, data = climateSupport, family = binomial(
    link = "logit"))
summary(null_logit)

anova_test <- anova(null_logit, climate_logit, test = "LRT")
anova_test

# Reporting:
```

Showing the results:

Table 1: Summary of Null Logistic Regression Model

	Dependent variable:
	choice
Constant	−0.007
	(0.022)
Observations	8,500
Log Likelihood	−5,891.705
Akaike Inf. Crit.	11,785.410
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: LRT Results

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

From the results, we see that the p-value is smaller than 0.000. This means we can reject the global null hypothesis that all of the slopes are equal to zero, and we have find evidence for the alternative hypothesis that at least one of them is not.

Then we can look at individual coefficients. Table 1 shows the coefficient estimates, p-values, and the number of iterations it took to find the maximum likelihood estimate:

Table 3: Model Summary

	Dependent variable:
	choice
countries80 of 192	0.336***
	(0.054)
countries160 of 192	0.648***
	(0.054)
sanctions5%	0.192***
	(0.062)
sanctions15%	−0.133**
	(0.062)
sanctions20%	−0.304***
	(0.062)
Constant	−0.273***
	(0.054)
Number of Fisher Scoring iterations	4
Observations	8,500
Log Likelihood	−5,784.130
Akaike Inf. Crit.	11,580.260
Note:	*p<0.1; **p<0.05; ***p<0.01

Interpreting the coefficients from our logistic regression model:

- countries80 of 192: This coefficient of 0.336 means that moving from the reference category of 20 out of 192 countries to the one of 80 out of 192 countries participating increases the log odds of supporting the policy by 0.336, on average, while controlling for sanctions.
- countries160 of 192: This coefficient of 0.648 indicates that moving from 20 out of 192 countries to 160 out of 192 countries increases the log odds of support by 0.648, on average, while controlling for sanctions.
- sanctions5%: The coefficient of 0.192 for 5% sanctions means that introducing 5% sanctions, compared to none, is associated with an increase of 0.192 in the log odds, on average, while controlling for the number of participating countries.
- sanctions15%: This coefficient of -0.133 for 15% sanctions means that increasing sanctions from none to this level is associated with a decrease of -0.133 in the log odds, on average, while controlling for the number of participating countries.
- sanctions20%: This coefficient of -0.304 for 20% sanctions means that increasing sanctions from none to this level is associated with a decrease of -0.304 in the log odds, on average, while controlling for the number of participating countries. This decrease is more substantive than the one going from no sanctions to 15%.
- Constant (Intercept): Finally, the intercept of -0.273 means that the log odds of supporting the policy when all of our independent variables are at their reference levels (no sanctions and 20 out of 192 countries participating) is -0.273, on average.
- Moreover, all coefficients are statistically significant, and the corresponding p-values are always less than 0.01. Thus, for each of the coefficients, we can reject the null hypothesis that the coefficient is equal to zero.

Additionally, we can "transform" our results to make them more informative on the relationship between changing from the different categories of our independent variables and both the odds of supporting a policy and the probability of supporting a policy:

```r
b_countries_80_of_192 <- 0.336
b_countries_160_of_192 <- 0.648
b_sanctions_5_percent <- 0.192
b_sanctions_15_percent <- -0.133
b_sanctions_20_percent <- -0.304

# Transforming from "effect on log odds" to "effect on odds":

odds_80_countries <- exp(b_countries_80_of_192)
odds_160_countries <- exp(b_countries_160_of_192)
odds_sanctions_5_percent <- exp(b_sanctions_5_percent)
odds_sanctions_15_percent <- exp(b_sanctions_15_percent)
odds_sanctions_20_percent <- exp(b_sanctions_20_percent)

# Transforming from "effect on odds" to "effect on probability":

prob_80_countries <- odds_80_countries / (1 + odds_80_countries)
prob_160_countries <- odds_160_countries / (1 + odds_160_countries)
prob_sanctions_5_percent <- odds_sanctions_5_percent / (1 + odds_
    sanctions_5_percent)
prob_sanctions_15_percent <- odds_sanctions_15_percent / (1 + odds_
    sanctions_15_percent)
prob_sanctions_20_percent <- odds_sanctions_20_percent / (1 + odds_
    sanctions_20_percent)

# Creating table to report:

transf_results <- data.frame(
  Variable = c("80 of 192 countries", "160 of 192 countries", "Sanctions
    5%", "Sanctions 15%", "Sanctions 20%"),
  Odds = c(odds_80_countries, odds_160_countries, odds_sanctions_5_
    percent, odds_sanctions_15_percent, odds_sanctions_20_percent),
  Probability = c(prob_80_countries, prob_160_countries, prob_sanctions_5
    _percent, prob_sanctions_15_percent, prob_sanctions_20_percent)
)

stargazer(transf_results, type = "latex",
          title = "Effects of Countries Participation and Sanctions on
    Policy Support",
```

Showing the results:

Table 4: Effects of Countries Participation and Sanctions on Policy Support

	Variable	Odds	Probability
1	80 of 192 countries	1.399	0.583
2	160 of 192 countries	1.912	0.657
3	Sanctions 5%	1.212	0.548
4	Sanctions 15%	0.875	0.467
5	Sanctions 20%	0.738	0.425

Interpretation:

'Effects on Odds'

- 80 of 192 countries: The odds of supporting the policy increase by 1.399, the factor shown for '80 of 192 countries' compared to the baseline, meaning the odds are increased by around 40%, on average, while controlling for sanctions.
- 160 of 192 countries: For '160 of 192 countries"', the odds of support increase by a factor of 1.912, or by 91.2%, on average, while controlling for sanctions.
- Sanctions 5%: Introducing sanctions at a 5% level increases the odds of policy support by a factor of 1.212, or 21.1%, on average, while controlling for participating countries.
- Sanctions 15%: Introducing sanctions at a 15% level decreases the odds of policy support by a factor of 0.875, or 12.5%, on average, while controlling for participating countries.
- Sanctions 20%: Introducing sanctions at a 20% level increases the odds of policy support by a factor of 0.738, or 26.2%, on average, while controlling for participating countries.

'Effects on Probabilities'

- 80 of 192 countries: Changing from 20 to 80 countries out of 192 is associated with an increase of 0.583 in the probability of support, on average, while controlling for sanctions.
- 160 of 192 countries: Changing from 20 to 80 countries out of 192 is associated with an increase of 0.657 in the probability of support, on average, while controlling for sanctions.
- Sanctions 5%: Changing from no sanctions to 5% is associated with an increase of 0.548 in the probability of support, on average, while controlling for participating countries.

- Sanctions 15% Changing from no sanctions to 15% is associated with a decrease of 0.467 in the probability of support, on average, while controlling for participating countries.
- Sanctions 20%: Changing from no sanctions to 20% is associated with a decrease of 0.425 in the probability of support, on average, while controlling for participating countries.

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)

   Answer to Question 2A

   Changing the reference category to 5% to be able to make a direct interpretation and showing the results of fitting the model again:

```
climateSupport$sanctions2 <- relevel(climateSupport$sanctions, ref =
    "5%")

# Fitting again:

climate_logit2 <- glm(choice ~ countries + sanctions2, data =
    climateSupport,
                      family = binomial(link = logit))

# Looking at the results:
```

   As we can see, the coefficient for moving from 5% (reference category) to 15% is -0.325, so when making this change in sanctions (from 5% to 15%), we expect the log odds of supporting a policy to decrease by 0.325. We can also "translate" this so we get information about what this change in categories would do to the odds and the probability:

```
coef5to15 <- -0.32510
odds <- exp(coef5to15)
prob <- odds / (1 + odds)

odds
prob

# PART B ##
```

   The results of "transforming" are that, for increasing sanctions from 5% to 15%, the odds ratio is **0.722**. This means there is a decrease in the OR by approximately 28% when sanctions are increased from 5% to 15%, all else in the model constant. In terms of the probability of support, this probability under 15% sanctions, compared to 5%, is **0.4194333** or approximately 42%

Table 5: Model Summary

	Dependent variable:
	choice
countries80 of 192	0.336***
	p = 0.000
countries160 of 192	0.648***
	p = 0.000
sanctions2None	−0.192***
	p = 0.003
sanctions215%	−0.325***
	p = 0.00000
sanctions220%	−0.495***
	p = 0.000
Constant	−0.081
	p = 0.129
Number of Fisher Scoring iterations	4
Observations	8,500
Log Likelihood	−5,784.130
Akaike Inf. Crit.	11,580.260
Note:	*p<0.1; **p<0.05; ***p<0.01

(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 to Question 2B

We are asked to see what the probability of support is when Sanctions is ‘None’ and Countries is ’180 out of 192‘. We can calculate this as follows:

```
filtered_data <- climateSupport %>%
  filter(countries == "80 of 192", sanctions == "None")

# Creating function based on formula in slide 40, W4:

probability <- function(b0, b1, xi) {
  1 / (1 + exp(-(b0 + b1 * xi)))
}

coefficients <- coef(climate_logit)
b0 <- coefficients['(Intercept)']
b1 <- coefficients['countries80 of 192'] # This is dependent to this case

prob_support <- probability(b0, b1, xi = 1)
prob_support

# Double-checking with predict():

predicted_prob2 <- predict(climate_logit, newdata = filtered_data, type = "response")
summary(predicted_prob2)

# PART C ##
```

We see that the estimated probability is 0.515, or 51.5%, which indicates that when 80 out of 192 countries participate in a policy without implementing any sanctions, there is approximately a fifty-fifty chance that an individual will support it.

(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 to Question 2C

The results could change, because including an interaction between sanctions and countries would allow for the effect of one variable (for instance, countries) on the outcome to vary according to the other variable (for instance, sanction). To evaluate whether including an interaction is appropriate, we can fit a model with the interaction and evaluate its performance compared to the additive one, by performing another likelihood ratio test:

```
climate_logit_int <- glm(choice ~ countries + sanctions +
    countries*sanctions,
                         data = climateSupport,
                         family = binomial(link = logit))

# LRT:

anova_test2 <- anova(climate_logit, climate_logit_int, test = "
    LRT")
anova_test2

anova_table <- as.data.frame(anova_test2)

# Reporting:
```

Looking at the results from the test, and at all the estimated coefficients for the interactive model:

Table 6: LRT Results

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

From the LRT, we get a p-value of 0.3912, so we are not able to reject the null hypothesis, and the additive model performs better than the interactive one. Additionally, looking at all the estimated coefficients, we can see that the answers to questions 2a and 2b do change when including an interaction, as the estimated coefficients are different in the interactive model compared to the additive one.

Table 7: Summary of Null Logistic Regression Model

	Dependent variable:
	choice
countries80 of 192	0.376***
	(0.106)
countries160 of 192	0.613***
	(0.108)
sanctions5%	0.122
	(0.105)
sanctions15%	−0.097
	(0.108)
sanctions20%	−0.253**
	(0.108)
countries80 of 192:sanctions5%	0.095
	(0.152)
countries160 of 192:sanctions5%	0.130
	(0.151)
countries80 of 192:sanctions15%	−0.052
	(0.152)
countries160 of 192:sanctions15%	−0.052
	(0.153)
countries80 of 192:sanctions20%	−0.197
	(0.151)
countries160 of 192:sanctions20%	0.057
	(0.154)
Constant	−0.275***
	(0.075)
Observations	8,500
Log Likelihood	−5,780.983
Akaike Inf. Crit.	11,585.970
Note:	*p<0.1; **p<0.05; ***p<0.01

This is clearer when looking at the results of both models side by side:

Table 8: Model Summary

	Dependent variable:	
	choice	
	(1)	(2)
countries80 of 192	0.336^{***} p = 0.000	0.376^{***} p = 0.0005
countries160 of 192	0.648^{***} p = 0.000	0.613^{***} p = 0.000
sanctions5%	0.192^{***} p = 0.003	0.122 p = 0.247
sanctions15%	-0.133^{**} p = 0.032	−0.097 p = 0.371
sanctions20%	-0.304^{***} p = 0.00001	-0.253^{**} p = 0.020
countries80 of 192:sanctions5%		0.095 p = 0.535
countries160 of 192:sanctions5%		0.130 p = 0.390
countries80 of 192:sanctions15%		−0.052 p = 0.731
countries160 of 192:sanctions15%		−0.052 p = 0.736
countries80 of 192:sanctions20%		−0.197 p = 0.192
countries160 of 192:sanctions20%		0.057 p = 0.712
Constant	-0.273^{***} p = 0.00000	-0.275^{***} p = 0.0003
Number of Fisher Scoring iterations	4	
Observations	8,500	8,500
Log Likelihood	−5,784.130	−5,780.983
Akaike Inf. Crit.	11,580.260	11,585.970
Note:	*p<0.1; **p<0.05; ***p<0.01	