Problem Set 2

Yana Konshyna

February 17, 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.

   Uploading data from GitHub, then convert two variables `countries` and `sanctions` into the factors and setting reference categories "20 of 192" and "5%" appropriately. The global null hypothesis H_0: all $\beta_j = 0$. H_a: at least one $\beta_j \neq 0$.
   Fitting an additive logistic model `climateSupport_logit`.

```r
data <- load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/
    main/datasets/climateSupport.RData?raw=true"))

# Converting variables 'countries' and 'sanction' into factors and
    setting reference categories
climateSupport$countries <- factor(climateSupport$countries, levels = c("
    20 of 192", "80 of 192", "160 of 192"), ordered = FALSE)
climateSupport$countries <- relevel(climateSupport$countries, ref = "20
    of 192")
climateSupport$sanctions <- factor(climateSupport$sanctions, levels = c("
    5%", "None", "15%", "20%"), ordered = FALSE)
climateSupport$sanctions <- relevel(climateSupport$sanctions, ref = "5%")

# Fitting an additive logistic model
climateSupport_logit <- glm(choice ~ countries + sanctions, family =
    binomial(link = "logit"), data = climateSupport)
summary(climateSupport_logit)
```

Output:

Table 1

```
Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          -0.08081    0.05316  -1.520  0.12848
countries80 of 192    0.33636    0.05380   6.252 4.05e-10 ***
countries160 of 192   0.64835    0.05388  12.033  < 2e-16 ***
sanctionsNone        -0.19186    0.06216  -3.086  0.00203 **
anctions15%          -0.32510    0.06224  -5.224 1.76e-07 ***
sanctions20%         -0.49542    0.06228  -7.955 1.79e-15 ***
---
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
```

Checking results running `anova` test on the additive model compared to the null model using `chi-squared test`:

```
#  Create a null model
nullMod <- glm(choice ~ 1, family = binomial(link = "logit"), data =
    climateSupport)
# Checking results running anova test on the additive model compared to
    the null model
anova(nullMod, climateSupport_logit, test = "Chisq") #chi-squared test
```

Output:

```
Analysis of Deviance Table
Model 1: choice ~ 1
Model 2: choice ~ countries + sanctions
      Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
2      8494      11568  5   215.15 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking results running `anova` test on the additive model compared to the null model using `Likelihood Ratio test (LRT)`:

```
anova(nullMod, climateSupport_logit, test = "LRT") #LRT test (equivalent)
```

Output:

```
Analysis of Deviance Table
Model 1: choice ~ 1
Model 2: choice ~ countries + sanctions
    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
2      8494      11568  5   215.15 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: After running an additive logistic model and conducting two tests we obtain P-value that is below critical value of 0.05, that means we can reject null hypothesis that there is no variable that improve the fit of our model. Therefore, we can conclude that at least one predictor is reliable in the model.

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)

   Interpretation of a coefficient from `Table 1`: On average, for all policies in all numbers of participating countries [20 of 192; 80 of 192; 160 of 192], increasing sanctions from 5% to 15%, will decrease in the log odd by 0.325 that an individual will support 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?

   Calculating predicted probability using function `predict`:

```
predictedProb <- predict(climateSupport_logit, newdata = data.frame(
    countries = "80 of 192", sanctions = "None"), type = "response",
    se = TRUE)
print(predictedProb$fit)
```

   Output:

```
0.5159191

```

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

   If we included the interaction term in the model, the answers to 2a and 2b would potentially change because interaction term allows for two fit lines with intercepts ans slopes that differ for each group, and which we should estimate.
   Fitting an interactive logistic model:

```
climateSupport_logit_interact <- glm(choice ~ countries*sanctions,
    family = binomial(link = "logit"), data = climateSupport)
stargazer::stargazer(climateSupport_logit, climateSupport_logit_
    interact, title = "Comparison of Additive and Interactive Models"
    ,
                     model.names = TRUE)
```

   Looking at the interactive effects in Model 2 (`Table 2`) we can see that potentially one of them is statistically differantiable from zero (countries80 of 192:sanctions20%).

Table 2: Comparison of Additive and Interactive Models

	Dependent variable:	
	choice	
	logistic	
	Model 1	Model 2
Constant	-0.081	-0.153^{**}
	(0.053)	(0.073)
countries80 of 192	0.336^{***}	0.470^{***}
	(0.054)	(0.109)
countries160 of 192	0.648^{***}	0.743^{***}
	(0.054)	(0.106)
sanctionsNone	-0.192^{***}	-0.122
	(0.062)	(0.105)
sanctions15%	-0.325^{***}	-0.219^{**}
	(0.062)	(0.107)
sanctions20%	-0.495^{***}	-0.374^{***}
	(0.062)	(0.107)
countries80 of 192:sanctionsNone		-0.095
		(0.152)
countries160 of 192:sanctionsNone		-0.130
		(0.151)
countries80 of 192:sanctions15%		-0.147
		(0.154)
countries160 of 192:sanctions15%		-0.182
		(0.151)
countries80 of 192:sanctions20%		-0.292^{*}
		(0.153)
countries160 of 192:sanctions20%		-0.073
		(0.152)
Observations (N)	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	

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

To check if including an interaction is appropriate, I run `anova` test to compare interactive model with additive model using `Likelihood Ratio test (LRT)`:

```
anova(climateSupport_logit, climateSupport_logit_interact, test = "
    LRT")
```

Output:

```
Analysis of Deviance Table
Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries * sanctions
     Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8494      11568
2      8488      11562  6   6.2928   0.3912
```

Conclusion: Our p-value (0.3912) is not below critical value 0.05, so we cannot reject the null hypothesis that the interactive model does not improve the model fit comparing to addtive model. It means we don't have suficient evidence that including interaction effect of the number of participating countries and sanctions is a significant predictor for odds that an individual agreed with the policy.