Problem Set 1 - My Answer

Junjie LIU[1,*]

18 February, 2024

[1] Department of Political Science, Trinity College Dublin, 2 Clare, Street, Dublin 2, Ireland

[*] Correspondence: Junjie LIU <liuj13@tcd.ie>

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.
- Total available points for this homework is 80.

In this problem set, you will run several regressions and create an add variable plot (see the lecture slides) in `R` using the `incumbents_subset.csv` dataset. Include all of your code.

Question 1

%(20 points)} 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.

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.

\section*{\textcolor{My Answer}}

```
# load data
library("fastDummies")
```

```
## Thank you for using fastDummies!

## To acknowledge our work, please cite the package:

## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from
```

```
library("stringr")
load("../../../datasets/climateSupport.RData")
data <- climateSupport
head(data)
```

```
##           choice  countries sanctions
## 1  Not supported  80 of 192       15%
## 9  Not supported 160 of 192       15%
## 17 Not supported 160 of 192      None
## 25 Not supported  80 of 192       15%
## 33 Not supported 160 of 192        5%
## 41 Not supported  20 of 192       15%
```

```
## preprocessing
data_dup <- data
data_dup$choice <- ifelse(data_dup$choice == "Not supported", 0, 1)
data_dup <-  fastDummies::dummy_columns(data_dup,
                      select_columns = c("countries", "sanctions"),
                      remove_first_dummy=TRUE, remove_selected_columns=TRUE)
names(data_dup) <- str_replace_all(names(data_dup), c(" " = "." , "," = "", "%" = ""))
head(data_dup)
```

```
##   choice countries_80.of.192 countries_160.of.192 sanctions_5 sanctions_15
## 1      0                   1                    0           0            1
## 2      0                   0                    1           0            1
## 3      0                   0                    1           0            0
## 4      0                   1                    0           0            1
## 5      0                   0                    1           1            0
## 6      0                   0                    0           0            1
##   sanctions_20
## 1            0
## 2            0
## 3            0
## 4            0
## 5            0
## 6            0
```

```r
# Question 1
## modelling - additive model
model <- glm(as.formula(paste(names(data_dup)[1], "~",
                        paste(names(data_dup)[-1], collapse = "+"))),
             family=binomial(link="logit"), data=data_dup)

summary(model)
```

```
##
## Call:
## glm(formula = as.formula(paste(names(data_dup)[1], "~", paste(names(data_dup)[-1],
##     collapse = "+"))), family = binomial(link = "logit"), data = data_dup)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -0.27266    0.05360  -5.087 3.64e-07 ***
## countries_80.of.192   0.33636    0.05380   6.252 4.05e-10 ***
## countries_160.of.192  0.64835    0.05388  12.033  < 2e-16 ***
## sanctions_5           0.19186    0.06216   3.086  0.00203 **
## sanctions_15         -0.13325    0.06208  -2.146  0.03183 *
## sanctions_20         -0.30356    0.06209  -4.889 1.01e-06 ***
## ---
## 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
```

```r
# Question 2
print("The model's summary suggests that if almost every country is
      involved in a policy (participate [160 of 192]), increasing the
      penalty rate from 5% to 15% results in the reduction of the likelihood,
```

```
      expressed in log odds, that a person will endorse the policy by
      approximately 0.32511 (= -0.13325 - 0.19186).")
```

```
## [1] "The model's summary suggests that if almost every country is \n      involved in a policy (parti
```

```
given_cond <- data.frame(
  countries_80.of.192=1, countries_160.of.192=0,
  sanctions_5 =0, sanctions_15=0, sanctions_20=0
)
prediction <- predict(model, given_cond, type="response")
print(paste("The prediction result is: ", round(prediction[1], 4)))
```

```
## [1] "The prediction result is:  0.5159"
```

```
model_inter <- glm(choice~.*., family=binomial(link="logit"), data=data_dup)
summary(model_inter)
```

```
##
## Call:
## glm(formula = choice ~ . * ., family = binomial(link = "logit"),
##     data = data_dup)
##
## Coefficients: (4 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                              -0.27469    0.07534  -3.646 0.000267
## countries_80.of.192                       0.37562    0.10627   3.535 0.000408
## countries_160.of.192                      0.61266    0.10801   5.672 1.41e-08
## sanctions_5                               0.12179    0.10518   1.158 0.246909
## sanctions_15                             -0.09687    0.10822  -0.895 0.370723
## sanctions_20                             -0.25260    0.10806  -2.338 0.019412
## countries_80.of.192:countries_160.of.192       NA         NA      NA       NA
## countries_80.of.192:sanctions_5           0.09471    0.15232   0.622 0.534071
## countries_80.of.192:sanctions_15         -0.05229    0.15167  -0.345 0.730262
## countries_80.of.192:sanctions_20         -0.19721    0.15104  -1.306 0.191675
## countries_160.of.192:sanctions_5          0.13009    0.15103   0.861 0.389063
## countries_160.of.192:sanctions_15        -0.05165    0.15267  -0.338 0.735136
## countries_160.of.192:sanctions_20         0.05688    0.15367   0.370 0.711279
## sanctions_5:sanctions_15                       NA         NA      NA       NA
## sanctions_5:sanctions_20                       NA         NA      NA       NA
## sanctions_15:sanctions_20                      NA         NA      NA       NA
##
## (Intercept)                              ***
## countries_80.of.192                      ***
## countries_160.of.192                     ***
## sanctions_5
## sanctions_15
## sanctions_20                             *
## countries_80.of.192:countries_160.of.192
## countries_80.of.192:sanctions_5
```

```
## countries_80.of.192:sanctions_15
## countries_80.of.192:sanctions_20
## countries_160.of.192:sanctions_5
## countries_160.of.192:sanctions_15
## countries_160.of.192:sanctions_20
## sanctions_5:sanctions_15
## sanctions_5:sanctions_20
## sanctions_15:sanctions_20
## ---
## 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: 11562  on 8488  degrees of freedom
## AIC: 11586
##
## Number of Fisher Scoring iterations: 4
```

```
lr_test <- anova(model, model_inter, test="LRT")
print(lr_test)
```

```
## Analysis of Deviance Table
##
## Model 1: choice ~ countries_80.of.192 + countries_160.of.192 + sanctions_5 +
##     sanctions_15 + sanctions_20
## Model 2: choice ~ (countries_80.of.192 + countries_160.of.192 + sanctions_5 +
##     sanctions_15 + sanctions_20) * (countries_80.of.192 + countries_160.of.192 +
##     sanctions_5 + sanctions_15 + sanctions_20)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      8494      11568
## 2      8488      11562  6   6.2928   0.3912
```

```
print("the p-value of the likelihood ratio test is higher than the given
      threshold: 0.05, thus, we could not reject the null hypothesis, and
      there is no significant difference between the additive model and
      model with interaction")
```

```
## [1] "the p-value of the likelihood ratio test is higher than the given \n      threshold: 0.05, thus
```