Problem Set 2

Applied Stats II

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.

```r
load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/datasets/climateSupport.RData?raw=true"))
####
pset2_data<-climateSupport
############################
# In the 'choice' column/variable contains logical values Not supported and Supported
# I Convert 'Supported' to 1 and 'Not supported' to 0
pset2_data$choice <- ifelse(pset2_data$choice == "Supported", 1, 0)
######
head(pset2_data)
View(pset2_data)
dim(pset2_data) # Rows/Observations= 8500; Columns/Variables=3
names(pset2_data) # "choice" ;   "countries" ; "sanctions"
##
with(pset2_data, table(pset2_data$choice)) # This line of the code provide the frequency of Not supported and Supported
#  0          1
# 4264     4236
```

   The given dataset has three variables such as: choice that need to be changed to binary where I replaced Not supported by zero and Supported by 1. After running R code using with() function to check the frequency table, I have found that 4264 or 50.16% don't support and 4236 or 49.84% support international environment agreement or policies people support (Bechtel and Scheve 2013).

```r
## Fit logistic regression model
fit_myFull_logit_supporting_policy <- glm (choice ~ countries + sanctions, family=binomial(link="logit"),
                                           data=pset2_data)
# Summary of the logistic regression model
summary(fit_myFull_logit_supporting_policy)
####
###Answer of Summary Output
###
#Call:
#  glm(formula = choice ~ countries + sanctions, family = binomial(link = "logit"),
#      data = pset2_data)

# 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
#
   ##############################################################################################

#####
# In a logistic regression, the response being modeled is the log(
   odds) that Y=1. The regression coefficients gives the
#change in log(odds) in the response for a unit change in the
   predictor variables, holding all other predictor variables
#constant.Since log(odds) are difficult to interpret, I have decided
   to exponentiate them (i.e., estimated parameters)
#to put the results on odd scale:
```

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.