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 the data at first, and because of the data type, I use the load function
load("climateSupport.RData")

# Fit an additive model with the data I loaded
model <- glm(choice ~ countries + sanctions, data = climateSupport, family = binomial())

# Display the summary of the model to get the coefficients, global null hypothesis, and p-value
summary(model)
```

```
##
## Call:
## glm(formula = choice ~ countries + sanctions, family = binomial(),
##     data = climateSupport)
##
## 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
```

   The number of fisher scoring iterations is 4

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)

```r
# I use a simple way to calculate for that instead of using some function because there are some system error in my Rstudio when I finish the homework

# Calculate predicted probability for sanctions at 5%
prob_5 <- predict(model, newdata = data.frame(countries = "160 of 192", sanctions = "5%"), type = "response")
```

```r

# Calculate predicted probability for sanctions at 15%
prob_15 <- predict(model, newdata = data.frame(countries = "160 of 192", sanctions = "15%"), type = "response")

# Calculate odds
odds_5 <- prob_5 / (1 - prob_5)
odds_15 <- prob_15 / (1 - prob_15)

# Calculate odds ratio
odds_ratio <- odds_15 / odds_5

print(odds_ratio)
```

```
##         1
## 0.7224531
```

For the policy in which very few countries participate [20 of 192]

```r
# I use a simple way to calculate for that instead of using some function because there are some system error in my Rstudio when I finish the homework

# Calculate predicted probability for sanctions at 5%
prob_5 <- predict(model, newdata = data.frame(countries = "160 of 192", sanctions = "5%"), type = "response")

# Calculate predicted probability for sanctions at 15%
prob_15 <- predict(model, newdata = data.frame(countries = "160 of 192", sanctions = "15%"), type = "response")

# Calculate odds
odds_5 <- prob_5 / (1 - prob_5)
odds_15 <- prob_15 / (1 - prob_15)

# Calculate odds ratio
odds_ratio <- odds_15 / odds_5

print(odds_ratio)
```

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

```r
# Calculate estimated probability
prob <- predict(model, newdata = data.frame(countries = "80 of 192", sanctions = "None"), type = "response")

print(prob)
```

The value for the prob is:

```
##         1
## 0.5159191
```

(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.

```r
# Fit model with interaction
model_interaction <- glm(choice ~ countries * sanctions, data =
    climateSupport, family = binomial())

# Compare models
anova(model, model_interaction, test = "Chisq")
```

```
## 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
```

Including an interaction term in the model could potentially change my answers to questions 2a and 2b because it allows for the effect of one variable (e.g., sanctions) on the response (e.g., support for the policy) to depend on the level of another variable (e.g., number of participating countries). This means that the impact of increasing sanctions could be different depending on how many countries are participating.
The code I've written calculates odds ratios for changes in sanctions without considering potential interactions between sanctions and the number of countries. If an interaction term were included in the model, I would need to adjust the 'predict' function calls to account for this interaction when estimating probabilities for specific scenarios. This could lead to different probabilities (and therefore different odds and odds ratios) than those calculated without the interaction term, as it would capture more complex relationships between variables.