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.

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.

Answer 1

Begin by reorganising factors ...

```
1
2 data$choice <- as.factor(ifelse(data$choice == "Supported", 1, 0))
3
4 data$countries <- factor(data$countries, ordered = FALSE, levels = c("20 of
    192", "80 of 192", "160 of 192"), labels = c("_20_192", "_80_192", "_160_
    192"))
5 data$sanctions <- factor(data$sanctions, ordered = FALSE, levels = c("None", "
    5%", "15%", "20%"), labels = c("none", "_5_percent", "_15_percent", "_20_
    percent"))
6
7 data$countries <- as.factor(data$countries)
8 data$sanctions <- as.factor(data$sanctions)
9
10
```

Fit an additive model ...

```
1 logit_base <- glm(choice ~ sanctions + countries, data = data, family =
    binomial(link = "logit"))
2
3 summary(logit_base)
4
5 Coefficients        Estimate    Std. Error z value Pr(>|z|)
6 (Intercept)          -0.27266     0.05360  -5.087 3.64e-07 ***
7 sanctions_5_percent   0.19186     0.06216   3.086  0.00203 **
8 sanctions_15_percent -0.13325     0.06208  -2.146  0.03183 *
9 sanctions_20_percent -0.30356     0.06209  -4.889 1.01e-06 ***
10 countries_80_192     0.33636     0.05380   6.252 4.05e-10 ***
11 countries_160_192    0.64835     0.05388  12.033  < 2e-16 ***
12 ---
13
14 (Dispersion parameter for binomial family taken to be 1)
15
16 Null deviance: 11783  on 8499  degrees of freedom
17 Residual deviance: 11568  on 8494  degrees of freedom
18 AIC: 11580
19
20 Number of Fisher Scoring iterations: 4
21
```

-0.273 is the average unit log odds of an individual supporting a policy, given no sanctions and holding countries participating constant at reference level (20 out of 192).

A comparison of null deviance and residual deviance is used to test global null hypothesis.

H0: All slopes = 0
HA: At least one Beta is not equal to 0.

A likelihood test is used ...

```
1
2 nullMod <- glm(formula = data$choice ~ 1, family = "binomial", data = data)
3
4 summary(nullMod)
5 Coefficients Estimate   Std. Error z value Pr(>|z|)
6 (Intercept) -0.006588   0.021693   -0.304    0.761
7
8 (Dispersion parameter for binomial family taken to be 1)
9
10 Null deviance: 11783  on 8499  degrees of freedom
11 Residual deviance: 11783  on 8499  degrees of freedom
12 AIC: 11785
13
14 Number of Fisher Scoring iterations: 3
15
16
```

Now run an anova test on the logit model compared to the null model ...

```
1 anova(nullMod, logit_base, test = "Chisq")
2
3 Model 1: choice ~ 1
4 Model 2: choice ~ sanctions + countries
5 Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
6 1       8499      11783
7 2       8494      11568  5   215.15 < 2.2e-16 ***
8
9
```

Interpretation: p-value is less than 0.01, we can conclude that at least one predictor in logit base model is reliable, and reject the global null hypothesis.

Answer 2

2. (a) For the policy in which nearly all countries participate (162 out of 192), increasing the sanction rate from 5 to 15 percent, changes the odds that an individual will support the policy. This can be calculated by subtracting the log-odds of supporting the policy when sanctions are set at 15 percent and at 5 percent.

$$\begin{aligned}\text{Change in log-odds} &= \text{sanctions\_15\_percent} - \text{sanctions\_5\_percent} \\ &= (-0.13325) - (0.19186) \\ &= -0.13325 - 0.19186 \\ &= -0.32511\end{aligned}$$

2. (b) The estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanction is associated with an increased in odds by a factor of 0.516 (51 percent). Calculation below ...

$$\begin{aligned}\text{Log-odds} &= \text{Intercept} + \text{countries\_80\_192} \times 1 \\ &= -0.27266 + 0.33636 \times 1 \\ &= 0.0637\end{aligned}$$

$$P(\text{Support}) = \frac{e^{\text{Log-odds}}}{1 + e^{\text{Log-odds}}}$$

$$P(\text{Support}) = \frac{e^{0.0637}}{1 + e^{0.0637}} \approx \frac{1.0655}{1 + 1.0655}$$

$$P(\text{Support}) \approx \frac{1.0655}{2.0655} \approx 0.516$$

```
#log-odds in R

> exp(0.0637)/(1+exp(0.0637))
[1] 0.5159196
```

2. (c) What if added an interaction? Perform a test to see if appropriate...

```
glm(formula = choice ~ sanctions * countries, family = binomial(link = "logit"
    ),
data = data)

Coefficients                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                                -0.27469    0.07534  -3.646 0.000267***
sanctions_5_percent                         0.12179    0.10518   1.158 0.246909
sanctions_15_percent                       -0.09687    0.10822  -0.895 0.370723
sanctions_20_percent                       -0.25260    0.10806  -2.338 0.019412 *
countries_80_192                            0.37562    0.10627   3.535 0.000408***
countries_160_192                           0.61266    0.10801   5.672 1.41e-08***
sanctions_5_percent:countries_80_192        0.09471    0.15232   0.622 0.534071
sanctions_15_percent:countries_80_192      -0.05229    0.15167  -0.345 0.730262
sanctions_20_percent:countries_80_192      -0.19721    0.15104  -1.306 0.191675
sanctions_5_percent:countries_160_192       0.13009    0.15103   0.861 0.389063
sanctions_15_percent:countries_160_192     -0.05165    0.15267  -0.338 0.735136
sanctions_20_percent:countries_160_192      0.05688    0.15367   0.370 0.711279
```

Compare the full and interactive models ...

```
anova(logit_base, Model_Interactive, test = "LRT")

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

```

P-Value is not significant (0.392), so there doesn't appear to be evidence to include the interaction term. We could expect that the answers for 2(a) and 2(b) would change with the inclusion of an interactive term, as it changes the interpretation of all the coefficients (as seen above).