Problem Set 2

Applied Stats II

Due: February 28, 2022

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 class on Monday February 28, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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.csv` 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.

   The code I used to answer this question is as follows:

```
  load(url("https://github.com/ASDS-TCD/StatsII_Spring2022/blob/main/
   datasets/climateSupport.RData?raw=true"))

summary(climateSupport)

head(climateSupport)

tail(climateSupport)

#converting the factors to numeric values

climateSupport$choice <- as.numeric(as.factor(climateSupport$choice))-1 #
    1 = participant supported policy , 0 = did not support

climateSupport$countries <- as.numeric(as.factor(climateSupport$countries
   ))-1 #  0 = 20 of 192 countries , 1 = 80 of 192, 2 = 160 of 192
climateSupport$sanctions <- as.numeric(as.factor(climateSupport$sanctions
   ))-1 # 0= None, 1 = 5%, 2 = 15% and 3 = 20%

#logit regression because the data is binary - working from lecture 4

ad_model <- glm(choice ~ countries + sanctions, data = climateSupport,
   family = binomial(logit))

summary(ad_model)
```

   Here is my table of results:

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.14458    0.04518  -3.200  0.00137 **
countries    0.32436    0.02689  12.062  < 2e-16 ***
sanctions   -0.12353    0.01964  -6.291 3.15e-10 ***
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05   .   0.1
         1
```

In terms of the global null hypothesis- I know the default null hypothesis = that a quantity to be measured is zero i.e. that the explanatory variables are not significant. I am taking it that global null hypothesis is referring to when a group of hypotheses are being tested together, i.e. multiple variables. The global null hypothesis states that in this case, none of the individual null hypotheses is false, so for this problem set, the global null is that none of the variables are significant. Source: https://journals.sagepub.com/doi/full/10.1177/0962280218768326

I am taking the column labelled "Pr(¿—z—)" as referring to p value significance. Taking alpha as 0.05, I can see that the p value on my table of results for countries is ¡ 2e-16 and for sanctions is 3.15e-10. Both of these values are less than 0.05, meaning that we reject the global null hypothesis and I conclude that we can take both variables to be statistically significant.

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)

   The code I used to answer this question is as follows:

```
#policy supported by 160 of 192 countries

#I have chosen to use log odds ratios to answer this question based
    on lecture 4

#Regression equation y = x + bx1 + bx2 + error
#subbing values in to the above to set the conditions we have been
    given in the question and get an answer
#using formula for log odds from slides 30-36 of lecture 4

odds_1 <- exp((-0.144558) + (0.32436)*2 + (-0.12353)*1)/ (1+ exp
    ((-0.144558) + (0.32436)*2 + (-0.12353)*1)) # this is the model
    when sanctions are set at 5%
#we add exp() function to make it log odds and not simply odds on its
     own

odds_2 <- exp((-0.144558) + (0.32436)*2 + (-0.12353)*2)/ (1+ exp
    ((-0.144558) + (0.32436)*2 + (-0.12353)*2)) #going from 5% to 15%
     sanctions with same model

odds_1
odds_2



```

```
#using formula for log odds from slides 66-69 of lecture 4
odds_diff <- odds_1 - odds_2

odds_diff # change from 5% to 15% here causes a 0.03010176 increase in log odds likelihood of
#supporting a climate policy


```

(b) For the policy in which very few countries participate [20 of 192], how does increasing sanctions from 5% to 15% change the odds that an individual will support the policy? (Interpretation of a coefficient)
The code I used to answer this question is as follows:

```
#policy supported by only 20 of 192 countries

#Again using regression equation y = x + bx1 + bx2 + error
#again subbing in to the above equation
#exp() function allows us to create logarithmic odds
#using formula for log odds from slides 30-36 of lecture 4

odds_a <- exp((-0.144558) + (0.32436)*0 + (-0.12353)*1)/ (1+ exp((-0.144558) + (0.32436)*0 + (-0.12353)*1)) #sanctions set at 5%

odds_a

exp(odds_a)

odds_b <- exp((-0.144558) + (0.32436)*0 + (-0.12353)*2)/ (1+exp((-0.144558) + (0.32436)*0 + (-0.12353)*2)) #sanctions set at 15%

odds_b

exp(odds_b)

odds_change <- odds_a - odds_b

odds_change

#In this case there is a 0.03004869 increase in the estimated log odds of participants supporting the policy
#when increasing sanctions from 5% to 15%


```

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

```
#80 participating countries, no sanctions

```

```r
#subbing in to regression equation:

est_prob <- exp((-0.144558) + (0.32436)*1 + (-0.12353)*0)/(1+(exp(-0.144558) + (0.32436)*1 + (-0.12353)*0))

est_prob #answer found = estimated probability of 0.5466251


```

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

```r
int_model <- glm(choice ~ countries*sanctions, data = climateSupport, family = binomial(logit))

summary(int_model)

#summary output of the interactive model is slightly different to that of additive model,
#with slightly different coefficients :

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         -0.148144   0.057311  -2.585  0.00974 **
countries            0.328007   0.045036   7.283 3.26e-13 ***
sanctions           -0.121111   0.030987  -3.908 9.29e-05 ***
countries:sanctions -0.002455   0.024288  -0.101  0.91950
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05   .   0.1      1


#checking if the answers would change:
#part a-

odds_3 <- exp((-0.148144) + ((0.328007)*2)*((-0.121111 )*1)) / (1+ exp((-0.148144) + ((0.328007)*2)*((-0.121111 )*1))) # this is the model when sanctions are set at 5%
#we add exp() function to make it log odds and not simply odds on its own

odds_4 <- exp((-0.148144) + ((0.328007)*2)*((-0.121111)*2)) / (1+ exp((-0.148144) + ((0.328007)*2)*((-0.121111)*2))) #going from 5% to 15% sanctions with same model

odds_3
odds_4



#using formula for log odds from slides 66-69 of lecture 4
odds_diff_2 <- odds_3 - odds_4
```

```
odds_diff_2 #answer here = 0.01950954 = different to original







#part b-

odds_c <- exp((-0.148144) + ((0.328007)*0)*(( -0.121111)*1))/ (1+ exp((-0.148144) + ((0.328007)*0)*(( -0.121111)*1))) #sanctions set at 5%

odds_c



odds_d <- exp((-0.148144) + ((0.328007)*0)*((-0.121111)*2))/ (1+ exp((-0.148144) + ((0.328007)*0)*((-0.121111)*2))) #sanctions set at 15%

odds_d


odds_change_2 <- odds_c - odds_d

odds_change_2 #answer here =  0 = different to original


#I got two different answers here to the original, but I am not sure why.




```

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

  I was not sure if I was meant to run an anova test here, but this is the code I tried and its output:

```
#Testing significance
anova_test <- anova(ad_model, int_model, test = "Chisq")

anova_test



```

```
Analysis of Deviance Table

Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries * sanctions
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8497      11597
2      8496      11597  1 0.010214   0.9195



#not sure how to interpret the outcome of this, but I can see that the deviance
#score is 0.010214.


```