Problem Set 2

Applied Stats II

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]

Question 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.

Solution

Loading the Dataset

```
#Loading the dataset

climateSupport <- read.csv("C:/Users/hp/Downloads/climateSupport.csv")


climateSupport$choice      <-      as.factor(climateSupport$choice)
climateSupport$countries   <-   as.factor(climateSupport$countries)
climateSupport$sanctions   <-   as.factor(climateSupport$sanctions)
str(climateSupport)
```

```
## 'data.frame':        8500 obs. of 4 variables:
## $ X                : int 1 9 17 25 33 41 49 57 65 73 ...
## $ choice           : Factor w/ 2 levels "Not supported",..: 1 1 1 1 1 1 2 1 2 1 ...
## $ countries: Factor w/ 3 levels "160 of 192","20 of 192",..: 3 1 1 3 1 2 1 3 1 1 ...
## $ sanctions: Factor w/ 4 levels "15%","20%","5%",..: 1 1 4 1 3 1 3 2 3 2 ... levels(climateSupport$choice)
```

```
## [1] "Not supported" "Supported"
```

Logistic regression model

```
logmodel <- glm(formula = choice~ countries + sanctions, data = climateSupport, family = binomial) summary(logmodel)
```

```
##
## Call:
## glm(formula = choice ~ countries + sanctions, family = binomial,
##             data = climateSupport)
##
## Deviance Residuals:
##         Min       1Q       Median  3Q        Max ## -1.4259 -
1.1480 -0.9444    1.1505   1.4298
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.24244 0.05360 4.523 6.09e-06 *** ## countries20 of 192 -
0.64835 0.05388 -12.033 < 2e-16 *** ## countries80 of 192 -0.31199   0.05387 -
5.792 6.97e-09 ***
## sanctions20%                -0.17032             0.06216 -2.740 0.00615 **
## sanctions5%                  0.32510         0.06224         5.224 1.76e-07 ***
## sanctionsNone            0.13325 0.06208 2.146 0.03183 * ## ---
## 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 model can be written as: $logit(p) =$

$0.24 - 0.648 \times countries20of192 - 0.31 \times countries80of192 - 0.170 \times sanctions20\% + 0.325 \times sanctions5\% + 0.133 \times sanc$

The Global null hypothesis is:
$ H_0: $ The model is not significant *VS* $ H_1: $ The model is significant

To test the significance for the overall model, we use the p- value.

```
pvalue <- 1-pchisq((logmodel$null.deviance-logmodel$deviance), (logmodel$df.null-logmodel$df.residual)); ## [1] 0
```

The p-value is 0 which is less than $\alpha = 0.05$. We reject H_0, hence the model is significant.

Question 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)

Solution

Increasing sanctions from 5% to 15% will reduce the odds that an individual will support the policy by 0.32510 and then by -0.31199. The decrease is significant since the coefficient is significant at $\alpha = 0.05$

(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)

Solution

Increasing sanctions from 5% to 15% will reduce the odds that an individual will support the policy by 0.32510 and then by 0.64835. The decrease is significant since the coefficient is significant at $\alpha = 0.05$

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

Solution

```
pr <- 0.24244-0.31199+0.13325;pr
```

```
## [1] 0.0637
```

The estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions is 0.0637.

(d) 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. ## Solution

Model with Interactions

```
logmodel2 <- glm(formula = choice~ countries * sanctions, data = climateSupport, family = binomial) summary(logmodel2)
```

```
##
## Call:
## glm(formula = choice ~ countries * sanctions, family = binomial,
##     data = climateSupport)
##
## Deviance Residuals:
##     Min       1Q   Median   3Q      Max ## -1.4359 -
1.1570 -0.9632   1.1349   1.4079
##
## Coefficients:
##                                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                                     0.1894535 0.0748703   2.530 0.011392 *
## countries20 of 192                               -0.5610171 0.1078890 -5.200 1.99e-07 ***
## countries80 of 192                               -0.2376898 0.1062197 -2.238 0.025239 *
## sanctions20%                                    -0.0472023 0.1074835 -0.439 0.660547
## sanctions5%                                     0.4003886 0.1065957   3.756 0.000173 ***
## sanctionsNone 0.1485161 0.1076839 1.379 0.167838 ## countries20 of 192:sanctions20% -
0.1085272 0.1535837 -0.707 0.479795
## countries80 of 192:sanctions20% -0.2534390 0.1508281 -1.680 0.092896 .
## countries20 of 192:sanctions5% -0.1817328 0.1509412 -1.204 0.228591 ## countries80 of
192:sanctions5%  -0.0347283 0.1534944 -0.226 0.821006 ## countries20 of 192:sanctionsNone -
0.0516474 0.1526666 -0.338 0.735136 ## countries80 of 192:sanctionsNone 0.0006461
0.1512924   0.004 0.996592
## ---
## 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

The answers to 2a and 2b would not potentially change if we included the interaction term in this model because the interaction effect between the two variables is not significant.