Problem Set 2

Applied Stats II

Due: February 19, 2023

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 19, 2023. 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.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.

```r
# part 1 - additive model
modl <- glm(choice ~ countries + sanctions, data = dat,
            family = "binomial")
predict(modl, dat)
stargazer::stargazer(modl, type = "text")
summary(modl)
```

```
> summary(modl)
Call:glm(formula = choice ~ countries + sanctions,
family = "binomial",     data = dat)
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)  0.37569    0.05416   6.937 4.01e-12 ***
countries20 -0.64835    0.05388 -12.033  < 2e-16 ***
countries80 -0.31199    0.05387  -5.792 6.97e-09 ***
sanctions15 -0.13325    0.06208  -2.146  0.03183 *
sanctions20 -0.30356    0.06209  -4.889 1.01e-06 ***
sanctions5   0.19186    0.06216   3.086  0.00203 **
---
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
```

```
> test2 # can reject knowledge
Analysis of Deviance Table
Model 1: choice ~ 1
Model 2: choice ~ countries + sanctions
Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
```

```
2      8494      11568  5   215.15 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) The results of the additive model and the ANOVA suggest we can reject the null hypothesis and consider the Zscores of the prediction betas. All betas are statistically significant at the alpha = 0.05 level, holding all others constant. That suggests a change in the levels of the countries or sanctions variables will be associated with a high log odds of a change in the participant's choice, given we account for the other variable. Thus, a shift from 20 countries to 80 countries is expected to elicit a change in the log odds of choice as will a change in the number of sanctions from 5 to 20.

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)

```
For the policy in which nearly all countries participate:
Yhat = 0.376 + -0.648(0 or 1) - 0.312(0 or 1) - 0.133(0 or 1)
 - 0.304(0 or 1) + 0.192(0 or 1)
Yhat15percent = 0.376 + 0.192(1) = 0.568
Yhat5percent = 0.376 - 0.133(1) = 0.243
prob = odds / (1 + odds)
= -0.325 / (1 - 0.325) = 0.4814815 decrease in probability
```

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

```
Y_hat = 0.376 - 0.312(1) = 0.064
0.064 / (1 + 0.064) = 0.06015038 probability
```

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

Potentially, yes. Adding an interaction term adds another beta coefficient and potentially affects the shape of the model.

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

```r
# part 2c
modl1 <- glm(choice ~ countries * sanctions, dat, family = "
    binomial")
compar <- anova(modl, modl1, test = "LRT")
# summary(compar)
compar # fail to reject H0
```

The ANOVA suggests that there is no statistically significant difference between the two models, so we fail to reject H0 that including an interaction would have no effect on the model outcomes.