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.

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

- 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, Fit an additive model.
head(climateSupport)

summary(lm(choice ~ sanctions + countries, climateSupport)) #need a non non linear regression for analysis
summary(climateSupport)


#climateSupport <- as.logical(as.numeric(as.factor(climateSupport$choice)))

?optim #checking what model to use

reg <- glm(choice ~ .,
           data = climateSupport,
           family = "binomial")

summary(reg)

logit_reg <- glm(choice ~ sanctions + countries, data= climateSupport, family =    #slide 30 week 4
                                                   binomial (link = "logit"))

summary(logit_reg)
```

*Apologies the stargazer code would not function

Call: glm(formula = choice   sanctions + countries, family = binomial(link = "logit"), 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) -0.005665 0.021971 -0.258 0.796517 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 *** countries.L 0.458452 0.038101 12.033 ¡ 2e-16 *** countries.Q -0.009950 0.038056 -0.261 0.793741 — 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

As the $p = 0.796517$, if a is set at 0.5

We fail to reject the null hypothesis that “countries participating in sanctions support climate change” as p is greater than a.

It is also possible to conclude from this p value that the variables of class are independent.

Null Hypothesis

$H1 : u = /u0$

$H2 : u = u0$

A one until increase in our B will result in a -0.005665 increase in our choice outcome, with all other factors being held constant.

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
#2a
logit_reg_1a <- glm(choice ~ (sanctions == "5%") + (countries == "160 of 192"), data= climateSupport, family = binomial (link = "logit"))
summary(logit_reg_1a)

logit_reg_1 <- glm(choice ~ (sanctions == "15%") + (countries == "160 of 192"), data= climateSupport, family = binomial (link = "logit"))
summary(logit_reg_1)
```

   If p-value is less than a, then we can conclude that at least one predictor is a significant predictor in logistic regression model. When sanctions are at 5 percent the choice coefficent is at 0.24 but when it is raised to 15 percent the coefficent drops to 0.09. While the support coefficent goes from -0.24 at 5 percent to -0.14 at 15 percent.

   First Model

   Call: glm(formula = choice   (sanctions == "5family = binomial(link = "logit"), data = climateSupport)

   Deviance Residuals: Min 1Q Median 3Q Max -1.422 -1.074 -1.074 1.145 1.284

   Coefficients: Estimate Std. Error z value Pr(¿—z—) (Intercept) -0.24820 0.02950 -8.414 ¡ 2e-16 *** sanctions == "5countries == "160 of 192"TRUE 0.48154 0.04655 10.345 ¡ 2e-16 *** — 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: 11631 on 8497 degrees of freedom AIC: 11637

Number of Fisher Scoring iterations: 4

Second Model

Call: glm(formula = choice (sanctions == ”15family = binomial(link = ”logit”), data = climateSupport)

Deviance Residuals: Min 1Q Median 3Q Max -1.325 -1.116 -1.078 1.240 1.280

Coefficients: Estimate Std. Error z value Pr(¿—z—) (Intercept) -0.14618 0.02940 -4.972 6.63e-07 *** sanctions == ”15countries == ”160 of 192”TRUE 0.48765 0.04645 10.499 ¡ 2e-16 *** — 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: 11669 on 8497 degrees of freedom AIC: 11675

Number of Fisher Scoring iterations: 4

(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)

```
#2b
logit_reg_2a <- glm(choice ~ (sanctions == "5%") + (countries == "20 of 192"), data= climateSupport, family = binomial (link = "logit"))
summary(logit_reg_1a)

logit_reg_2 <- glm(choice ~ (sanctions == "15%") + (countries == "20 of 192"), data= climateSupport, family = binomial (link = "logit"))
summary(logit_reg_1)
```

When sanctions are at 5 percent the choice coefficent is at 0.24 but when it is raised to 15 percent the coefficent drops to 0.09. While the support coefficent goes from -0.24 at 5 percent to -0.14 at 15 percent. The number of countries being reduced appears to have no effect.

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

```
#2c change data
reg_3 <- glm(choice ~ (sanctions == "None") + (countries == "80 of 192"), data= climateSupport, family = binomial (link = "logit"))

predicted_data <- with(climateSupport, expand.grid(choice = unique(choice),
                                                   sanctions = unique(sanctions),
```

```
                                                    countries = unique(countries)))


predicted_data <- cbind(predicted_data, predict(reg_3,
                                                newdata = predicted_data,
                                                type = "response",
                                                se = TRUE))


predicted_data <- within(predicted_data,
                         {
                           PredictedProb <- plogis(fit)
                           LL <- plogis(fit - (1.96 * se.fit))
                           UL <- plogis(fit + (1.96 * se.fit))
                         })

Summary(predicted_data)
```

(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.

```

reg_null1 <- glm(as.factor(choice) ~ 1, data = climateSupport, family = "binomial")
anova(reg_null1, reg, test = "Chisq")
```

Interaction effects indicate that a third variable influences the relationship between an independent and dependent variable. This type of effect makes the model more complex, so I would therefore think it would change the model. Interctions mean outcome variable are treated as the expected probability of success, which is unreliable for hypothsis testing and maybe cause an increase in the varience in the model. Resulting in overfittinng.