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.

The R script for additive model, summary output and p-value is:

```r
# load data
load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/datasets/climateSupport.RData?raw=true"))

# Creating an unordered list of 'countries' variable
countries <- factor(climateSupport$countries, levels = c("20 of 192", "80 of 192", "160 of 192"), ordered = FALSE)

# Creating an unordered list of 'sanctions' variable
sanctions <- factor(climateSupport$sanctions, levels = c("None", "5%", "15%", "20%"), ordered = FALSE)

# Change baseline category for the "countries" variable
countries <- relevel(climateSupport$countries, ref = "20 of 192")

# Change baseline category for the "sanctions" variable
sanctions <- relevel(climateSupport$sanctions, ref = "None")

# Fit logistic regression model
model <- glm(choice ~ countries + sanctions, data = climateSupport, family = binomial(link = "logit"))

# Summary output
summary(model)

# creating a baseline model using intercept
null_model <- glm(choice~1, data = climateSupport, family = binomial(link = "logit"))

# comparing both the models using likelihood ratio
anova(null_model, model, test = "LRT")
```

Model Outputs:				
Coefficients	Estimate	Std. Error	z value	Pr(¿—z—)
(Intercept)	-0.08081	0.05316	-1.520	0.12848
countries80 of 192	0.33636	0.05380	6.252	4.05e-10 ***
countries160 of 192	0.64835	0.05388	12.033	¡ 2e-16 ***
sanctionsNone	-0.19186	0.06216	-3.086	0.00203 **
sanctions15%	-0.32510	0.06224	-5.224	1.76e-07 ***
sanctions20%	-0.49542	0.06228	-7.955	1.79e-15 ***

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 p-values below the benchmark of 5% show that both the explanatory variables **countries** and **sanctions** has signifcant impact on the response variable **choice**.

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 R script is:

```r
# Coefficient for sanctions at 15%
sanctions_15 <- -0.32510

# Calculate the odds ratio
odds_15 <- exp(sanctions_15)

# Print the odds ratio
print(odds_15)
```

   The probability of 15% sanctions is estimated to be approx. 72%.

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

The R script for probability estimation is:

```r
# Coefficients from our model
intercept <- -0.08081
countries_80 <- 0.33636
sanctions_none <- -0.19186

# Calculate the log odds for the specific condition: 80 of 192
    countries with no sanctions
log_odds <- intercept + countries_80*1 + sanctions_none*1

# Convert log odds to probability
probability <- 1 / (1 + exp(-log_odds))

# Print the estimated probability
print(probability)
```

The probability of 80 countries out of 192 countries particpating with no sanctions is estimated to be 51.6%.

(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.

The R script for interaction model is:

```r
# Fit model with interaction term
model_interaction <- glm(choice ~ countries + sanctions + countries:
    sanctions,
                         data = climateSupport,
                         family = binomial(link = "logit"))

# Use ANOVA to compare the models
anova_result <- anova(model, model_interaction, test = "LRT")
```

Model 1: choice countries + sanctions

Model 2: choice countries + sanctions + countries:sanctions

Resid. Df Resid. Dev Df Deviance Pr(¿Chi)

1 8494 11568

2 8488 11562 6 6.2928 0.3912

The p-value above the benchmark of 5% shows that the interaction term of **countries** and **sanctions** does not have a signifcant impact on the response variable **choice**.