Problem Set 2

Applied Stats II

Due: February 28, 2022

Name: Gareth Moen

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.

```r
#### Part 1 ####

# View data structure
summary(clsu)
str(clsu)

# Logit regression, additive model
reg <- glm(choice ~ countries + sanctions,
           data = clsu,
           family = binomial) #link = "logit"

summary(reg)

# Exponentiate coefficients to get odds
reg_exp <- exp(coef(reg))
# Visualise the results
stargazer(reg_exp, type = "text")

# When the participating countries is '20 of 192' and sanctions are 'none',
# the odds of someone supporting a given policy = 0.994.

# Global null hypothesis
reg_null <- glm(choice ~ 1, data = clsu, family = "binomial") # 1 = fit an
# intercept only (i.e. sort of a "mean")
anov_res <- anova(reg_null, reg, test = "Chisq")

# Another way to view the results
stargazer(anov_res, type = "text")

# p value < 0.01, we can conclude that at least one predictor is reliable in
# model. That is to say that the number of countries participating, and the
# size of the sanctions for countries for not following the agreement have an
# influence on whether people support the policy or not
```

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)

(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)

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

(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.

```r

#### Part 2 ####
# 2a
# Predicting values from dataset
predicted_data <- with(clsu, expand.grid(countries = unique(countries),
                                         sanctions = unique(sanctions)))

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

# Confidence intervals and predicted probability
predicted_data <- within(predicted_data,
                         {
                           PredictedProb <- plogis(fit)
                           LL <- plogis(fit - (1.96 * se.fit))
                           UL <- plogis(fit + (1.96 * se.fit))
                         })

# Plot the estimates and confidence intervals
ggplot(data = predicted_data, mapping = aes(x = row.names(predicted_data), y = PredictedProb)) +
  geom_point() +
  geom_errorbar(aes(ymin = LL, ymax = UL), colour = "red") +
  labs(
    x = "Variable Values - Additive Model",
    y = "Predicted Probabilities"
    )

#Interpretation of coefficients
View(predicted_data)
```

```

# 160 of 192 & 5% > PredictedProb of 0.6543
# 160 of 192 & 15% > PredictedProb of 0.6365
# When the participating countries is '160 of 192' and sanctions increase from
# '5%' to 15% the odds of someone supporting a given policy drops by 2.8%

# 2b
# 20 of 192 & 5% > PredictedProb of 0.6177
# 20 of 192 & 15% > PredictedProb of 0.5987
# When the participating countries is '20 of 192' and sanctions increase from
# '5%' to 15% the odds of someone supporting a given policy drops by 3.2%

# 2c
# When the participating countries is '20 of 192' and there are no sanctions,
# the estimated probability of someone supporting a policy is 0.6064

## 2d
# It seems likely that there would be an interaction between the level of
# sanctions and the number of countries participating in a policy so including
# the interaction terms seems like a good idea. Let's see what happens.

# Null hypothesis
# H0:  # of participating countries | sanctions (none) =  # of participating
# countries | sanctions (5%) =  # of participating countries | sanctions (15%)
# =  # of participating countries | sanctions (20%)

# Let's run a new glm with the interaction
reg_2 <- glm(choice ~ countries * sanctions,
             data = clsu,
             family = binomial) #link = "logit"

summary(reg_2)

# Global null hypothesis 2
reg_nul_2 <- glm(choice ~ 1, data = clsu, family = "binomial")
anov_res_2 <- anova(reg_nul_2, reg_2, test = "Chisq")

# p value < 0.01, we can conclude that at least one predictor is also reliable
# in this interactive model.
# That is to say that there is an interaction happening between the
```

```r
variables

# Exponentiate coefficients to get odds
reg_2_exp <- exp(coef(reg_2))
# Visualise the results
stargazer(reg_2_exp, type = "text")

# Predicting new values from the dataset with interaction
predicted_data2 <- with(clsu, expand.grid(countries = unique(countries),
                                          sanctions = unique(sanctions)))

predicted_data2 <- cbind(predicted_data2, predict(reg_2,
                                                  newdata = predicted_data2,
                                                  type = "response",
                                                  se = TRUE))

# Confidence intervals and predicted probability
predicted_data2 <- within(predicted_data2,
                          {
                            PredictedProb <- plogis(fit)
                            LL <- plogis(fit - (1.96 * se.fit))
                            UL <- plogis(fit + (1.96 * se.fit))
                          })

# Plot of estimates and confidence intervals
ggplot(data = predicted_data2, mapping = aes(x = row.names(predicted_data2), y = PredictedProb)) +
  geom_point() +
  geom_errorbar(aes(ymin = LL, ymax = UL), colour = "blue") +
  labs(
    x = "Variable Values - Interactive Model",
    y = "Predicted Probabilities"
  )

# 160 of 192 & 5% > PredictedProb of 0.6555
# 160 of 192 & 15% > PredictedProb of 0.6335
# When the participating countries is '160 of 192' and sanctions increase from
# '5%' to 15% the odds of someone supporting a given policy drops by 3.5%,] a
# difference of +0.7% from the additive model

# 20 of 192 & 5% > PredictedProb of 0.6135
# 20 of 192 & 15% > PredictedProb of 0.6006
# When the participating countries is '160 of 192' and sanctions increase from
# '5%' to 15% the odds of someone supporting a given policy drops by 2.1%, a
```

```
# difference of 1.1% from the additive model

# Based on these results there seems to be a difference in the
    results of
# both models, although the difference is only 3.5% at most
```