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.

```r
# Structure of the dataset
str(climateSupport)

# Summary statistics of the dataset
summary(climateSupport)

# Dimensions of the dataset
dim(climateSupport)

# View first few rows of the dataset
head(climateSupport)

# Fit logistic regression model with interaction term
model_interaction <- glm(choice ~ countries * sanctions, data =
  climateSupport, family = binomial)

# Summary of the model
summary(model_interaction)

```

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)             -0.003809   0.022006  -0.173 0.862583
countries.L              0.457140   0.038078  12.005  < 2e-16 ***
countries.Q             -0.011167   0.038152  -0.293 0.769750
sanctions.L             -0.274221   0.043953  -6.239 4.41e-10 ***
sanctions.Q             -0.182289   0.044011  -4.142 3.45e-05 ***
sanctions.C              0.153245   0.044069   3.477 0.000506 ***
countries.L:sanctions.L -0.001754   0.076700  -0.023 0.981755
countries.Q:sanctions.L  0.133840   0.075554   1.771 0.076484 .
countries.L:sanctions.Q -0.007622   0.076156  -0.100 0.920278
countries.Q:sanctions.Q  0.093425   0.076303   1.224 0.220806
countries.L:sanctions.C  0.095197   0.075608   1.259 0.208001
countries.Q:sanctions.C  0.010449   0.077046   0.136 0.892123
---
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 intercept term (-0.003809) represents the estimated log odds of supporting the policy when all predictors are at their reference levels.

The coefficients for the "countries" variable indicate how the odds of supporting the policy change with different levels of participation. Specifically, the "countries.L" coefficient (0.457140) suggests that for each unit increase in the "countries" variable (moving from the reference level to the next), the log odds of supporting the policy increase by approximately 0.457.

The coefficients for the "sanctions" variable indicate how the odds of supporting the policy change with different levels of sanctions. For example, the coefficient for "sanctions.L" (-0.274221) suggests that for each unit increase in sanctions (moving from "5" to "15"), the log odds of supporting the policy decrease by approximately 0.274.

None of the interaction terms appear to be statistically significant at conventional levels (i.e., p-value ¿ 0.05), indicating that there's no strong evidence to suggest that the relationship between "countries" and "sanctions" varies significantly across different levels of the other predictor.

```r
# Global null hypothesis and p-value
null_hypothesis <- "There is no relationship between the predictors and the likelihood of an individual supporting the policy."
p_value <- summary(model_interaction)$coefficients[1, "Pr(>|z|)"]

# Print the global null hypothesis and p-value
cat("Global Null Hypothesis:", null_hypothesis, "\n")
cat("p-value:", p_value, "\n")

```

H0: "There is no relationship between the predictors and the likelihood of an individual supporting the policy."

The p-value associated with this hypothesis is approximately 0.8626, suggesting that there is no significant evidence to reject this null hypothesis.

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)

```
# Extract coefficient for the interaction term between countries and sanctions
interaction_coef <- coef(model_interaction)["countries.L:sanctions.L"]

# Interpretation of the coefficient
cat("The coefficient for the interaction term between countries and sanctions is:", interaction_coef, "\n\n")
cat("Interpretation:\n")
cat("For each one-unit increase in sanctions (from 5% to 15%), the odds of supporting the policy\n")
cat("are multiplied by exp(", interaction_coef, ") = ", exp(interaction_coef), "\n\n")

```

   For each one unit increase in sanctions, from 5 to 15 percentage, the odds of supporting the policy are multiplied by exp -0.001754016 = 0.9982475.

   This implies that increasing sanctions from 5 to 15 percentage is associated with a slight decrease in the odds of supporting the policy. However, the effect is minimal, indicating that the change in odds is not substantial.

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

```
# Define the values of countries and sanctions
countries_value <- "80 of 192"
sanctions_value <- "None"

# Predict the probability using the logistic regression model
predicted_probability <- predict(model_interaction,
newdata = data.frame(countries = countries_value, sanctions = sanctions_value),
type = "response")

# Print the estimated probability
cat("The estimated probability that an individual will support the policy if there are",
countries_value, "countries participating with no sanctions is:", predicted_probability, "\n")

```

   The estimated probability that an individual will support the policy if there are 80 of 192 countries participating with no sanctions is: 0.5252101

(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.

```
# Fit logistic regression model without the interaction term
model_no_interaction <- glm(choice ~ countries + sanctions, data =
    climateSupport, family = binomial)

# Perform likelihood ratio test
lrt <- anova(model_no_interaction, model_interaction, test = "Chisq")

# Print the results
print(lrt)

```

```
Analysis of Deviance Table

Model 1: choice ~ countries + sanctions
Model 2: choice ~ countries * sanctions
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8494      11568
2      8488      11562  6   6.2928   0.3912
```

In this case, the p-value (Pr(¿Chi)) is 0.3912, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis that the model without the interaction term is sufficient. This suggests that including the interaction term does not significantly improve the model's fit.