Problem Set 2 - Robert Baker

Applied Stats II

2024-02-17

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.

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

# Explore
head(climateSupport)

# Summarise
summary(climateSupport)

# Change baseline category for sanctions and countries variables
climateSupport$sanctions <- relevel(factor(climateSupport$sanctions, ordered = F),
ref = "5%")
climateSupport$countries <- relevel(factor(climateSupport$countries, ordered = F),
ref = "20 of 192")

# Fit logistic regression model to predict choice variable based on countries & sanctions
climate_logit <- glm(choice ~ countries + sanctions, data = climateSupport,
family = binomial(link = "logit"))

# Display a summary of the logistic regression model
summary(climate_logit)
```

```r
# Fit null logistic regression model to compare with full model using LRT
null_model <- glm(choice ~ 1, data = climateSupport, family = binomial(link = "logit"))

# Perform likelihood ratio test to compare the null and full models
likelihood_ratio_test <- anova(null_model, climate_logit, test = "Chisq")

# Print results of LRT
print(likelihood_ratio_test)
```

```
Coefficients:
                    Estimate Std. Error z value
(Intercept)          -0.0808     0.0532   -1.52
countries80 of 192    0.3364     0.0538    6.25
countries160 of 192   0.6483     0.0539   12.03
sanctionsNone        -0.1919     0.0622   -3.09
sanctions15%         -0.3251     0.0622   -5.22
sanctions20%         -0.4954     0.0623   -7.95

                    Pr(>|z|)
(Intercept)            0.128
countries80 of 192   4.1e-10 ***
countries160 of 192  < 2e-16 ***
sanctionsNone          0.002 **
sanctions15%         1.8e-07 ***
sanctions20%         1.8e-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


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   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First, we fitted an additive model and utilised Rs summary() output to extract the deviance from both the null and full models for our likelihood ratio test. We can also calculate this by hand by fitting a model without covariates. Both approaches give the same result with a p-value close to zero (below our critical threshold of 0.05). This indicates we reject the null hypothesis that neither variable enhances our model fit.

Regarding the model coefficients, the estimates suggest that the likelihood of supporting a policy increases with the number of countries participating and decreases with higher levels of sanctions. All coefficients exhibit statistical significance, with very low p-values. The analysis of deviance table further supports the models significance, with a p-value much lower than our critical threshold of 0.05.

As we can see this model required 4 iterations to find the maximum likelihood estimates.

```
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).

From the coefficient output, we observe that the coefficient for "sanctions15%" is -0.32.
This indicates that compared to the reference category "sanctionsNone", increasing
sanctions from 5% to 15% reduces the odds of an individual
supporting the policy by approximately e -0.32 times, on average, while
holding other variables constant.

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

```r
# Option 1: Predicting probability "by hand"
exp_coefs <- exp(coef(climate_logit)[1] +
                   coef(climate_logit)[2] * 1 +
                   coef(climate_logit)[3] * 0 +
                   coef(climate_logit)[4] * 1 +
                   coef(climate_logit)[5] * 0 +
                   coef(climate_logit)[6] * 0)

predicted_prob <- round(exp_coefs / (1 + exp_coefs), 2)

# Print
print(predicted_prob)

# Option 2: Predict probability using logistic regression model
predicted_prob <- predict(
  climate_logit,
  newdata = data.frame(countries = "80 of 192",
                       sanctions = "None"),
  type = "response"
)

# Rounding
rounded_predicted_prob <- round(predicted_prob, 2)

# Print
print(rounded_predicted_prob)
```

We see above the estimated probability that an individual will support the policy, given the condition of 80 out of 192 countries participating with no sanctions, is 0.52 or 52%.

(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.

```r
# Fit logistic regression model without interaction term
model_main_effects <- glm(choice ~ countries + sanctions,
```

```r
data = climateSupport, family = binomial(link = "logit"))

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

# Perform likelihood ratio test using Anova function
lrt <- anova(model_main_effects, model_interaction, test = "LRT")

# Print
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     6.29     0.39
```

Yes, the answers to 2a and 2b could potentially change as including an interaction term in the model allows for the estimation of both the intercept and the slope for two fitted lines, each representing a different policy type with varying numbers of participating countries (20 of 192; 80 of 192; 160 of 192).

The analysis indicated that the inclusion of interaction terms in the model, which assesses the impact of the number of participating countries by level of sanctions, does not yield statistically significant results. We observe that none of the interactive effects were statistically different from zero, suggesting a lack of discernible influence of the interaction between the number of participating countries and sanction levels on the dependent variable.

A likelihood ratio test comparing the additive model to the interactive model results in a p-value of 0.39, failing to reach the critical threshold of 0.05. As a result, we fail to reject the null hypothesis as there is insufficient evidence to justify the inclusion of the interaction term in the model.

Therefore, the analysis indicates that the number of participating countries and sanction levels, when considered together, do not significantly influence the likelihood of individuals supporting the policy.