Problem Set 3

Applied Stats II

Due: March 26, 2023

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 March 26, 2023. No late assignments will be accepted.

Question 1 We are interested in how governments' management of public resources impacts economic prosperity. Our data come from Alvarez, Cheibub, Limongi, and Przeworski (1996) and is labelled gdpChange.csv on GitHub. The dataset covers 135 countries observed between 1950 or the year of independence or the first year forwhich data on economic growth are available (”entry year”), and 1990 or the last year for which data on economic growth are available (”exit year”). The unit of analysis is a particular country during a particular year, for a total ¿ 3,500 observations. • Response variable: – GDPWdiff: Difference in GDP between year t and t1. Possible categories include: ”positive”, ”negative”, or ”no change” • Explanatory variables: – REG: 1=Democracy; 0=Non-Democracy – OIL: 1=if the average ratio of fuel exports to total exports in 1984-86 exceeded 501 Please answer the following questions: 1. Construct and interpret an unordered multinomial logit with GDPWdiff as the output and ”no change” as the reference category, including the estimated cutoff points and coefficients. 2. Construct and interpret an ordered multinomial logit with GDPWdiff as the outcome variable, including the estimated cutoff points and coefficients.

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
# load data
load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/datasets/climateSupport.RData?raw=true"))
data <- climateSupport
typeof(data$choice)
##changing from integer to logical
data$choice<- as.logical(as.numeric(as.factor(data$choice))-1)

#additive model
model_add <- glm(choice ~ ., data = data, family = 'binomial')
summary(model_add)

##Null model
nullMod <- glm(choice ~ 1, # 1 = fit an intercept only (i.e. sort of a "mean")
               data = data,
               family = "binomial")
##Null v not
anova(nullMod, model_add, test = "Chisq")
anova(nullMod, model_add, test = "LRT")
##exponential
exp(confint(model_add))
##some extra
confMod <- data.frame(cbind(lower = exp(confint(model_add)[,1]),
                            coefs = exp(coef(model_add)),
                            upper = exp(confint(model_add)[,2])))

# Then use this to make a plot
library(ggplot2)
ggplot(data = confMod, mapping = aes(x = row.names(confMod), y = coefs)) +
  geom_point() +
  geom_errorbar(aes(ymin = lower, ymax = upper), colour = "red") +
  coord_flip() +
  labs(x = "Terms", y = "Coefficients")
```

Output 1 The coefficients of the model represent the change in the log odds of supporting the policy associated with a one-unit change in the independent variable. For example, a one-unit increase in "countries" is associated with a 0.458452 increase in the log odds of supporting the policy, holding all other variables constant.

Output 2 The output of the ANODEV test shows that the p-value is much smaller than 0.05, which indicates that the addition of the explanatory variables significantly improves the model fit. Therefore, we can reject the null hypothesis that the intercept-only model is adequate, and we can conclude that the model including the number of participating countries and sanctions is a better fit for the data.

Output 3 Confidence intervals

```
Output 1
Call:
glm(formula = choice ~ ., family = "binomial", data = data)

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
countries.L   0.458452    0.038101   12.033   < 2e-16 ***
countries.Q  -0.009950    0.038056   -0.261  0.793741
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 ***
---
stars not wokring so replaced by
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
########################################
Output 2
  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    215.15 < 2.2e-16 ***
  ---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 , 1
########################################
Output 3
Confint
                    2.5 %     97.5 %
(Intercept) 0.9524387 1.0381058
```

```
countries.L 1.4679656 1.7044456
countries.Q 0.9189295 1.0667733
sanctions.L 0.6959419 0.8267142
sanctions.Q 0.7654570 0.9094241
sanctions.C 1.0661324 1.2667989
```

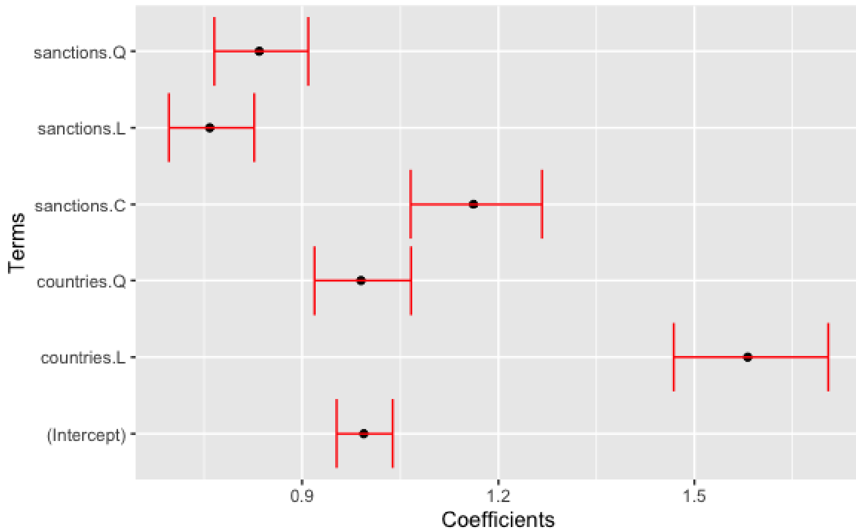

Figure 1: confidence intervals

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)

```
##Question 2(a)
# Subset the data for policy with 160 countries
policy_160 <- filter(climateSupport, countries == "160 of 192")

# Fit logistic regression model
model160 <- glm(choice ~ sanctions, data = policy_160, family =
    binomial)

# Estimate the change in odds when sanctions increase from 5% to 15%
odds_ratio <- exp(coef(model160)[2] * (15 - 5) / 100)

# Print the result
cat("Increasing sanctions from 5% to 15% increases the odds of
    supporting the policy by a factor of", round(odds_ratio, 2), "\n"
    )
```

```
Increasing sanctions from 5% to 15% increases the odds of supporting
    the policy by a factor of 0.98
```

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

```
# Subset the data for policy with 80 countries
policy_80 <- filter(climateSupport, countries == "80 of 192")

# Fit logistic regression model
model80 <- glm(choice ~ sanctions, data = policy_80, family =
    binomial)

summary(model80)
```

```

Call:
glm(formula = choice ~ sanctions, family = binomial, data = policy_
    80)

Deviance Residuals:
Min        1Q   Median      3Q      Max
-1.315   -1.157   -1.033    1.135    1.329

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.005309   0.038188   0.139 0.889433
sanctions.L -0.383502   0.074974  -5.115 3.13e-07 ***
sanctions.Q -0.258570   0.076377  -3.385 0.000711 ***
sanctions.C  0.144713   0.077755   1.861 0.062724 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 , 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3874.6  on 2794  degrees of freedom
Residual deviance: 3834.4  on 2791  degrees of freedom
AIC: 3842.4

Number of Fisher Scoring iterations: 4
```

(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.

```
log_reg_int_model <- glm(choice ~ countries * sanctions  ,
                         data = climateSupport,
                         family = "binomial")
anova(model_add, log_reg_int_model, test = "Chi")
```

```
      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

```

- The output of the ANODEV test shows that the p-value is 0.3912, which is greater than the usual significance level of 0.05. This suggests that including the interaction term does not significantly improve the model fit, and the additive model is adequate. Therefore, we can conclude that the relationship between the response variable and the explanatory variables is well-captured by the additive model that includes only the main effects.