

Problem Set 2

Applied Stats II

Due: February 19, 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 February 19, 2023. 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.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.

```
1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/
  datasets/climateSupport.RData?raw=true"))
3 data <- climateSupport
4 typeof(data$choice)
5 ##changing from integer to logical
6 data$choice<- as.logical(as.numeric(as.factor(data$choice))-1)
7
8 #additive model
9 model_add <- glm(choice ~ ., data = data, family = 'binomial')
10 print(summary(model_add))
11
12 ##Null model
13 nullMod <- glm(choice ~ 1, # 1 = fit an intercept only (i.e. sort of a "
  mean")
14               data = data,
15               family = "binomial")
16 ##Null v not
17 anova(nullMod, model_add, test = "Chisq")
18 anova(nullMod, model_add, test = "LRT")
19 ##exponential
20 exp(confint(model_add))
21 ##some extra
22 confMod <- data.frame(cbind(lower = exp(confint(model_add)[,1]),
23                             coefs = exp(coef(model_add)),
24                             upper = exp(confint(model_add)[,2]))))
25
26 # Then use this to make a plot
27 library(ggplot2)
28 ggplot(data = confMod, mapping = aes(x = row.names(confMod), y = coefs))
29   +
30   geom_point() +
31   geom_errorbar(aes(ymin = lower, ymax = upper), colour = "red") +
32   coord_flip() +
33   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

```

1 Output 1
2 Call:
3 glm(formula = choice ~ ., family = "binomial", data = data)
4
5 Deviance Residuals:
6 Min       1Q   Median       3Q      Max
7 -1.4259  -1.1480  -0.9444   1.1505   1.4298
8
9 Coefficients:
10 Estimate Std. Error z value Pr(>|z|)
11 (Intercept) -0.005665    0.021971  -0.258  0.796517
12 countries.L  0.458452    0.038101  12.033 < 2e-16 ***
13 countries.Q -0.009950    0.038056  -0.261  0.793741
14 sanctions.L -0.276332    0.043925  -6.291  3.15e-10 ***
15 sanctions.Q -0.181086    0.043963  -4.119  3.80e-05 ***
16 sanctions.C  0.150207    0.043992   3.414  0.000639 ***
17
18 stars not working so replaced by
19 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 , 1
20
21 (Dispersion parameter for binomial family taken to be 1)
22
23 Null deviance: 11783  on 8499  degrees of freedom
24 Residual deviance: 11568  on 8494  degrees of freedom
25 AIC: 11580
26
27 Number of Fisher Scoring iterations: 4
28 #####
29 Output 2
30 Analysis of Deviance Table
31
32 Model 1: choice ~ 1
33 Model 2: choice ~ countries + sanctions
34 Resid. Df Resid. Dev Df Deviance Pr(>Chi)
35 1      8499      11783
36 2      8494      11568    5    215.15 < 2.2e-16 ***
37
38 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 , 1
39 #####
40 Output 3
41 Confint
42           2.5 %    97.5 %
43 (Intercept) 0.9524387 1.0381058

```

```

44 countries.L 1.4679656 1.7044456
45 countries.Q 0.9189295 1.0667733
46 sanctions.L 0.6959419 0.8267142
47 sanctions.Q 0.7654570 0.9094241
48 sanctions.C 1.0661324 1.2667989

```

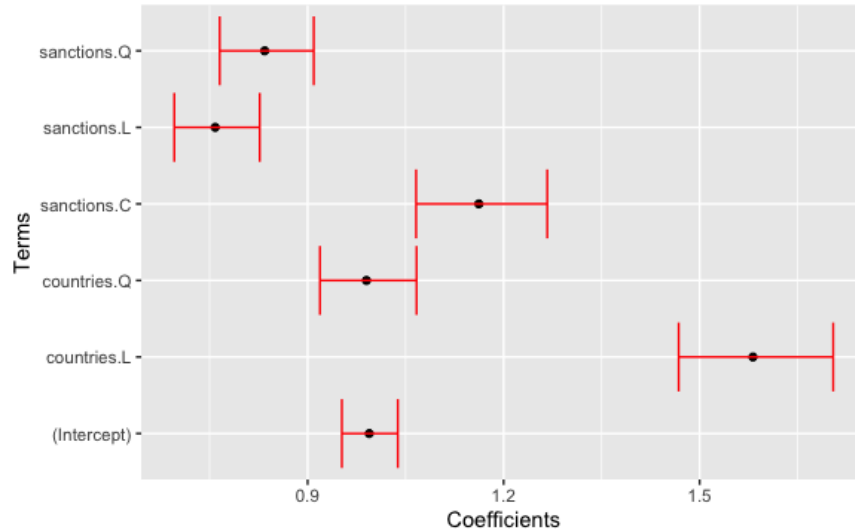


Figure 1: confidence intervals

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)

```

1 ##Question 2(a)
2 # Subset the data for policy with 160 countries
3 policy_160 <- filter(climateSupport, countries == "160 of 192")
4
5 # Fit logistic regression model
6 model160 <- glm(choice ~ sanctions, data = policy_160, family =
7   binomial)
8
9 # Estimate the change in odds when sanctions increase from 5% to 15%
10 odds_ratio <- exp(coef(model160)[2] * (15 - 5) / 100)
11
12 # Print the result
13 cat("Increasing sanctions from 5% to 15% increases the odds of
14   supporting the policy by a factor of", round(odds_ratio, 2), "\n"
15 )

```

```

1 Increasing sanctions from 5% to 15% increases the odds of supporting
2   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?

```

1 # Subset the data for policy with 80 countries
2 policy_80 <- filter(climateSupport, countries == "80 of 192")
3
4 # Fit logistic regression model
5 model80 <- glm(choice ~ sanctions, data = policy_80, family =
6   binomial)
7 summary(model80)

```

```

1
2 Call:
3 glm(formula = choice ~ sanctions, family = binomial, data = policy_
4   80)
5 Deviance Residuals:
6 Min       1Q   Median       3Q      Max
7 -1.315  -1.157  -1.033   1.135   1.329
8
9 Coefficients:
10 Estimate Std. Error z value Pr(>|z|)
11 (Intercept)  0.005309    0.038188   0.139 0.889433
12 sanctions.L -0.383502    0.074974  -5.115 3.13e-07 ***
13 sanctions.Q -0.258570    0.076377  -3.385 0.000711 ***
14 sanctions.C  0.144713    0.077755   1.861 0.062724 .
15 ---
16 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 , 1
17
18 (Dispersion parameter for binomial family taken to be 1)
19
20 Null deviance: 3874.6 on 2794 degrees of freedom
21 Residual deviance: 3834.4 on 2791 degrees of freedom
22 AIC: 3842.4
23
24 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.

```

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

```

```

1 Analysis of Deviance Table
2
3 Model 1: choice ~ countries + sanctions
4 Model 2: choice ~ countries * sanctions
5 Resid. Df Resid. Dev Df Deviance Pr(>Chi)

```

6	1	8494	11568			
7	2	8488	11562	6	6.2928	0.3912
8						

- 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.