# Problem Set 2

Applied Stats II

Due: February 19, 2023

## Instructions

- This problem set is due before 23:59 on Sunday February 19, 2023. No late assignments will be accepted.

Code in `PS2_ImeldaFinn.R`

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.

- 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]

Read in the data and modified `choice` variable:

```
load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/datasets/climateSupport.RData?raw=true"))
# choice = 1,2
# countries = 1, 2, 3
# sanctions = 1, 2, 3, 4

# get a version of the dataset with the response variable coded as
# True = supported
```

```
# False = not supported
cs <- climateSupport
cs$choice <- as.logical(as.numeric(cs$choice)-1)

summary(cs)
...

   choice              countries       sanctions
Mode :logical     20 of 192 :2865    None:2119
FALSE:4264        80 of 192 :2795    5%   :2133
TRUE :4236        160 of 192:2840    15%  :2111
                                     20%  :2137

```

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.

   ```
   mod
   ```

   (a) Summary output,

   ```

   Call:
   glm(formula = choice ~ ., family = binomial(link = "logit"),
       data = cs)

   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 ***
   ---
   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

   ```

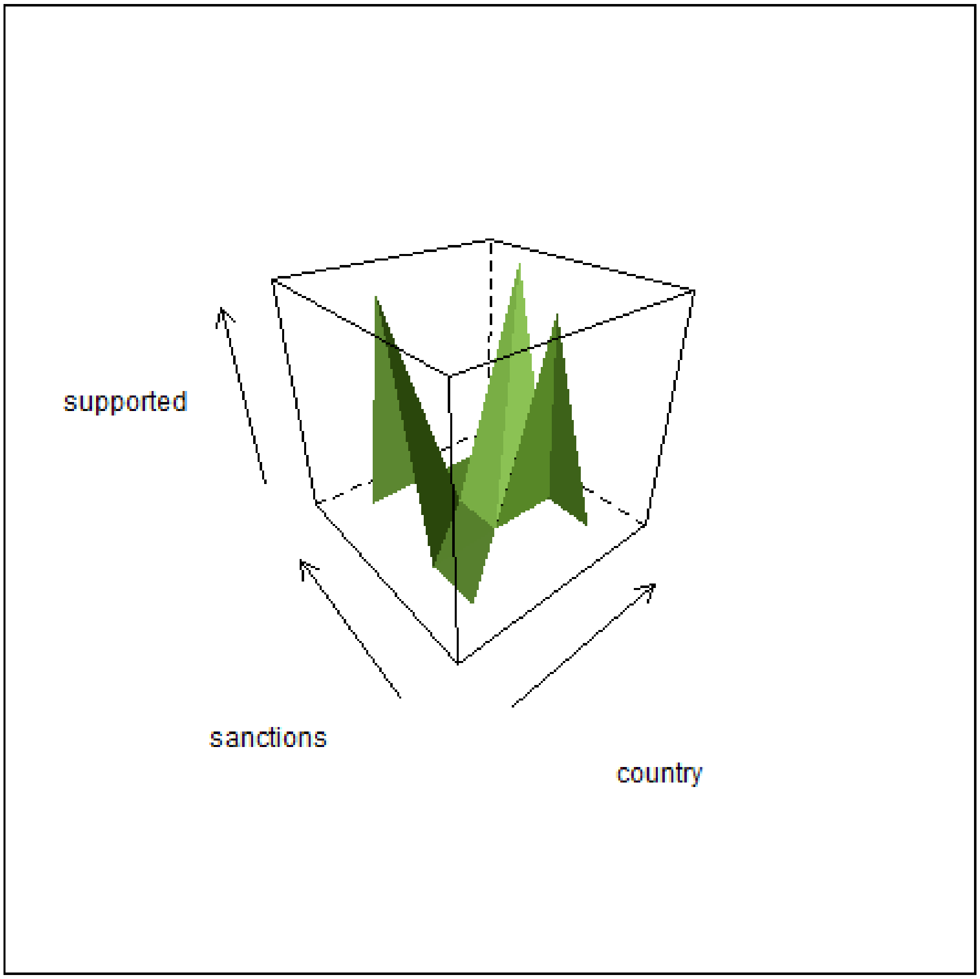


Figure 1: plot of additive (glm) model

(b) The global null hypothesis and $p$-value.

$H_0$: the explanatory variables have no effect on the likelihood of an individual supporting a policy

$H_a$: one or more of the explanatory variables have some effect on the likelihood of an individual supporting a policy

$\alpha = 0.05$

The data was modelled with no explanatory variables (*choice*˜1). The comparison of the two models is shown in 1

```
null_mod <- glm(choice ~1, family = binomial(link="logit"), data = cs
    )
```

A test was run to compare the deviances of the two models.

```
anova_null <- anova(null_mod, mod, test = "LRT")
```

The results are shown in 2. The $\chi^2$ statistic $= 11783 - 11568 = 215.15$. The assocated p-value with 5 degrees of freedom is $2.2 \times 10^{-16}$.

As the p-value is below $\alpha$ we reject the null hypothesis. The evidence does not support the assumption that none of the explanatory variables have any effect on our response variable `choice`. We expect that one or more of our explanatory variables will have a statistically significant effect on the probability of a policy being supported.

Table 1:

| | *Dependent variable:* | |
|---|---|---|
| | choice | |
| | *logistic* | |
| | (1) | (2) |
| countries: 80 of 192 | 0.458*** | |
| | (0.038) | |
| countries: 160 of 192 | −0.010 | |
| | (0.038) | |
| sanctions: 5% | −0.276*** | |
| | (0.044) | |
| sanctions: 5% | −0.181*** | |
| | (0.044) | |
| sanctions: 5% | 0.150*** | |
| | (0.044) | |
| Constant | −0.006 | −0.007 |
| | (0.022) | (0.022) |
| Observations | 8,500 | 8,500 |
| Log Likelihood | −5,784.130 | −5,891.705 |
| Akaike Inf. Crit. | 11,580.260 | 11,785.410 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

Table 2:

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Resid. Df | 2 | 8,496.500 | 3.536 | 8,494 | 8,499 |
| Resid. Dev | 2 | 11,675.830 | 152.134 | 11,568.260 | 11,783.410 |
| Df | 1 | 5.000 | | 5 | 5 |
| Deviance | 1 | 215.150 | | 215.150 | 215.150 |
| Pr(>Chi) | 1 | 0.000 | | 0 | 0 |

(c) Please describe the results and provide a conclusion.

When 20 out of 192 countries are included and there are no sanctions (base case), then the expected odds of a participant agreeing with a policy are $e^{-0.005665} = 0.994351$

A one unit increase in $X_k$ increases the odds of supporting a policy by a multiplicative factor of $e^{\beta_k}$

When 20 out of 192 countries are included and there are sanctions of 5%, the logodds change by $e^{-0.276332} = 0.758561$ compared to the base

$logit(p) = -0.005665 + -0.276332$

When 80 out of 192 countries are included and there are no sanctions, the logodds change by $e^{-0.458453} = 0.632261$ compared to the base

$logit(p) = -0.005665 - 0.458453$

When 80 out of 192 countries are included and there are sanctions of 5%, then:

$logit(p) = -0.005665 + -0.276332 - 0.458453 =$

$e^{logit(p)} = 0.4768993$

ie the odds reduce by 48.0%.

The predicted probabilities, and confidence intervals, are in Table 3

The estimates for $\beta_k$ are all significant at $p = 0.01$ except for 'countries: 160 of 192' (`countries.Q`), ie there is a predicted -0.1 change in *logit* going from 80 to 160 countries, but it is not statistically significant.

Table 3:

| | countries | sanctions | fit | se.fit | residual.scale | UL | LL | PredictedProb |
|---|---|---|---|---|---|---|---|---|
| 1 | 80 of 192 | 15% | 0.483 | 0.013 | 1 | 0.625 | 0.612 | 0.618 |
| 2 | 160 of 192 | 15% | 0.560 | 0.013 | 1 | 0.642 | 0.631 | 0.637 |
| 3 | 20 of 192 | 15% | 0.400 | 0.013 | 1 | 0.605 | 0.593 | 0.599 |
| 4 | 80 of 192 | None | 0.516 | 0.013 | 1 | 0.632 | 0.620 | 0.626 |
| 5 | 160 of 192 | None | 0.593 | 0.013 | 1 | 0.650 | 0.638 | 0.644 |
| 6 | 20 of 192 | None | 0.432 | 0.013 | 1 | 0.613 | 0.600 | 0.606 |
| 7 | 80 of 192 | 5% | 0.564 | 0.013 | 1 | 0.643 | 0.631 | 0.637 |
| 8 | 160 of 192 | 5% | 0.638 | 0.012 | 1 | 0.660 | 0.649 | 0.654 |
| 9 | 20 of 192 | 5% | 0.480 | 0.013 | 1 | 0.624 | 0.612 | 0.618 |
| 10 | 80 of 192 | 20% | 0.440 | 0.013 | 1 | 0.614 | 0.602 | 0.608 |
| 11 | 160 of 192 | 20% | 0.518 | 0.013 | 1 | 0.633 | 0.620 | 0.627 |
| 12 | 20 of 192 | 20% | 0.360 | 0.012 | 1 | 0.595 | 0.583 | 0.589 |

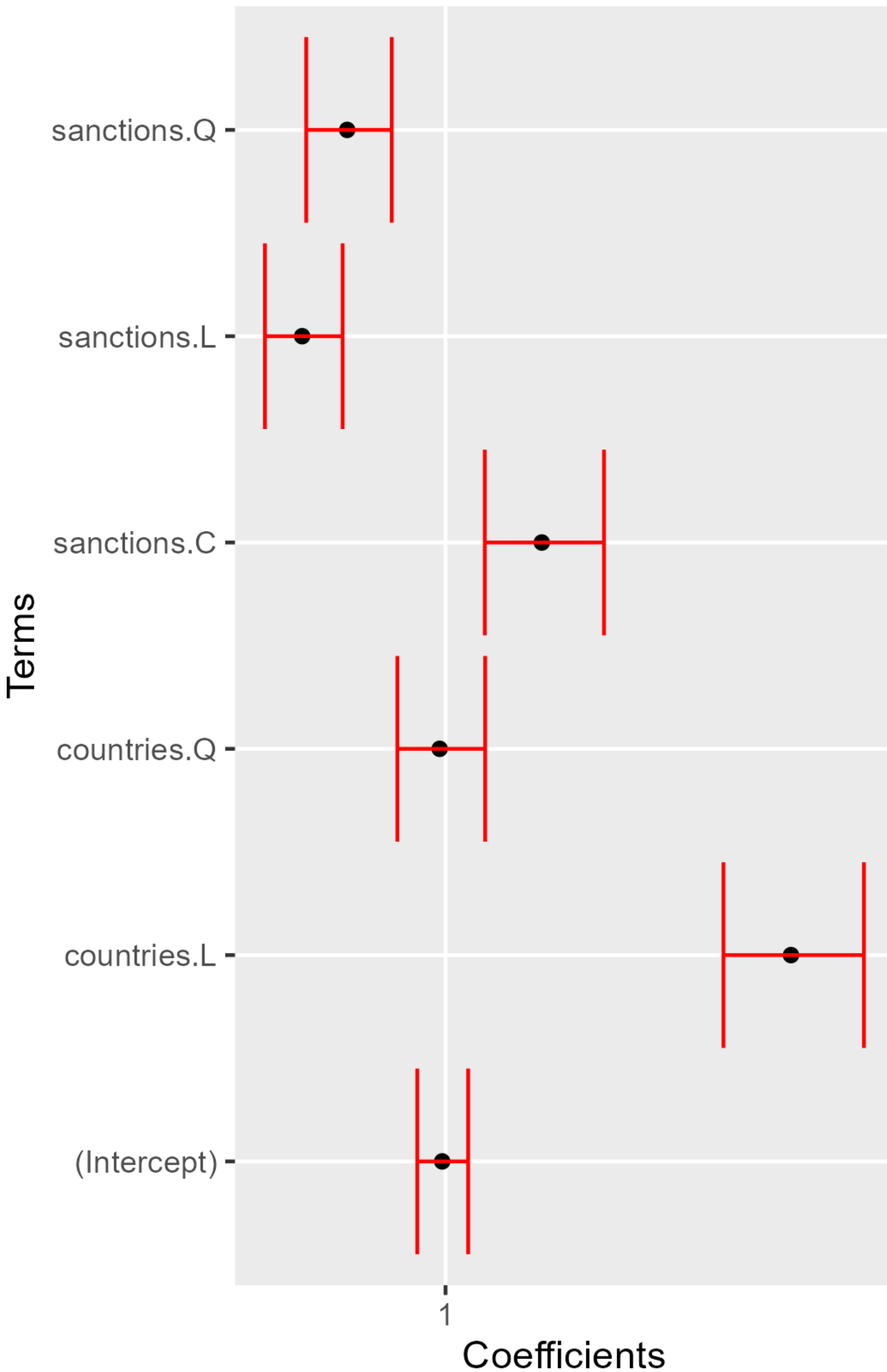


Figure 2: coefficients of additive model

It took 4 iterations to find the maximum likelihood estimates.

The log likelihood is -5,784.130

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

   (c) Including an interaction term would potentially change the results in 2a and 2b. The values for the coefficients would potentially be different (eg $beta_k$) and we would have to include the constituent coefficient values in calculating the value of the logit.

      - A model was run on the data, with an interaction between `countries` and `sanctions`, and an ANOVA/$\chi_2$ test was run. The results are shown in Tables 4 and 5.
        The test statistic of 6.2928, with 6 degrees of freedom, lead to a p-value of 0.3912. Therefore we cannot reject the null hypothesis that the two models are the same, ie we do not conclude that an interaction term is appropriate.

```r
int_mod <- glm(choice ~countries + sanctions + countries *
    sanctions,
                 family = binomial(link="logit"), data = cs)

anova_int <- anova(mod, int_mod, test= "LRT")
```

```r
    int_mod <- glm(choice ~countries + sanctions + countries *
    sanctions,
                 family = binomial(link="logit"), data = cs)

    anova_int <- anova(mod, int_mod, test= "LRT")

```

Table 4:

| | Dependent variable: | |
|---|---|---|
| | choice | |
| | *logistic* | |
| | (1) | (2) |
| countries: 80 of 192 | 0.458*** | 0.457*** |
| | (0.038) | (0.038) |
| countries: 160 of 192 | −0.010 | −0.011 |
| | (0.038) | (0.038) |
| sanctions: 5% | −0.276*** | −0.274*** |
| | (0.044) | (0.044) |
| sanctions: 5% | −0.181*** | −0.182*** |
| | (0.044) | (0.044) |
| sanctions: 5% | 0.150*** | 0.153*** |
| | (0.044) | (0.044) |
| countries.L:sanctions.L | | −0.002 |
| | | (0.077) |
| countries.Q:sanctions.L | | 0.134* |
| | | (0.076) |
| countries.L:sanctions.Q | | −0.008 |
| | | (0.076) |
| countries.Q:sanctions.Q | | 0.093 |
| | | (0.076) |
| countries.L:sanctions.C | | 0.095 |
| | | (0.076) |
| countries.Q:sanctions.C | | 0.010 |
| | | (0.077) |
| Constant | −0.006 | −0.004 |
| | (0.022) | (0.022) |
| Observations | 8,500 | 8,500 |
| Log Likelihood | −5,784.130 | −5,780.983 |
| Akaike Inf. Crit. | 11,580.260 | 11,585.970 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

Table 5: ANOVA additive vs Interactive

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Resid. Df | 2 | 8,491.000 | 4.243 | 8,488 | 8,494 |
| Resid. Dev | 2 | 11,565.110 | 4.450 | 11,561.970 | 11,568.260 |
| Df | 1 | 6.000 | | 6 | 6 |
| Deviance | 1 | 6.293 | | 6.293 | 6.293 |
| Pr(>Chi) | 1 | 0.391 | | 0.391 | 0.391 |