Problem Set 1

Philip Kruger
18328699
19/02/23

Question 1

The data set was loaded and and a summary of an additive glm was gotten with the following code:

```
load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/datasets/climateSupport.RData?raw=tru

mod <- glm(choice ~ ., # period functions as omnibus selector (kitchen sink additive model)
data = climateSupport,
family = "binomial")

summary(mod)
```

This output the following coefficients:

```
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
```

The intercept has coefficient -0.005665, this means that when somebody is asked if they support a climate policy which involves 20 countries and has no sanctions, the odds that the person will support it is $e^{-0.005665} =$ 0.994. This is the baseline odds ratio and is nearly 1 which means that the likelyhood of a person supporting this policy is 50/50.
For 80 countries the coefficient is 0.458452, this means that, all else being equal, the log odds of a person supporting the policy increases by 0.458452 over the baseline odds ratio. This gives an odds ratio of 1.573. This means that the presence of the policy applying to 80 countries is associated with an increase in the odds of a person saying they support it.
For 160 countries the coefficient is -0.009950, this means that, all else being equal, the log odds of a person supporting the policy increases by -0.009950 over the baseline odds ratio.
For 5 percent sanctions the coefficient is -0.276332, this means that, all else being equal, the log odds of a person supporting the policy increases by -0.276332 over the baseline odds ratio.
For 15 percent sanctions the coefficient is -0.181086, this means that, all else being equal, the log odds of a

person supporting the policy increases by -0.181086 over the baseline odds ratio.
For 20 percent sanctions the coefficient is 0.150207, this means that, all else being equal, the log odds of a person supporting the policy increases by 0.150207 over the baseline odds ratio.
Thus involving 80 countries and having sanctions of 20 percent are the only coefficients that increase the odds of a person supporting the policy over the baseline odds.

The global null hypothesis is that all coefficients (except intercept) are actually 0. This is checked with the following code:

```
nullMod <- glm(choice ~ 1, # 1 = fit an intercept only (i.e. sort of a "mean")
data = climateSupport,
family = "binomial")

#  Run an anova test on the model compared to the null model
anova(nullMod, mod, test = "Chisq")
```

which gives output:

```
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
```

We can see that the p-value is sufficiently small. As such we can reject the null hypothesis that all coefficients are zero.
For p-values of coefficients, all are *** significant except the intercept and 160 countries option.

2. a)
The Odds ratio in the two cases are calculated as:

$$\text{odds} = e^{intercept+countrycoeff+sanctionsCoeff}$$
$$\text{odds for 5 percent} = e^{-0.005665+-0.009950+-0.276332} = 0.7468$$
$$\text{odds for 15 percent} = e^{-0.005665+-0.009950+-0.181086} = 0.8214$$

This gives us an odds ratio of:

$$\text{odds ratio} = \frac{\frac{oddsodsupporting5percent}{oddsofnotsupporting5percent}}{\frac{oddsofsupporting15percent}{oddsofnotsupporting15percent}}$$
$$\text{odds ratio} = \frac{\frac{.7468}{1-.7468}}{\frac{.8214}{1-.8214}} = 0.6413$$

This means that the odds of somebody supporting the 5 percent sanctions is 0.6413 times greater than someone supporting 15 percent sanctions.

2. b)
To get the estimated probability from the coefficient we can use the formula:

$$\text{probability} = \frac{e^{intercept+80countirescoeff}}{1-e^{intercept+80countirescoeff}}$$
$$\text{probability} = \frac{e^{-.005665+0.458452}}{1+e^{-.005665+0.458452}} = 0.6113$$

This means that if an individual is asked if they support that policy the probability that they will respond affirmatively is 61 percent.

2. c)
To test if the interactive term would make a difference the following code was used to generate the models and compare them:

```
mod2 <- glm(choice ~ countries+sanctions, # additive model
data = climateSupport,
family = "binomial")

mod3 <- glm(choice ~ countries*sanctions, # interactive model
data = climateSupport,
family = "binomial")

anova(mod2, mod3, test = "Chisq") #comparing additive and interactive model
```

This generates the output:

```
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
```

Since the p-value is 0.3912, we are unable to reject the null hypothesis that the coefficients are the same. As such using a multiplicative model would not make a difference to any of the above results.