

# Problem Set 2

Aryan Goyal

February 2024

First, I load in the data and inspect it:

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

#inspecting the data
head(climateSupport)
summary(climateSupport)
```

## 1 Question 1

Fit an additive model. Provide the summary output, the global null hypothesis, and p-value. Please describe the results and provide a conclusion.

I first releval the variables and pick the reference category. Then I use the `glm()` function in R to run an additive, logistic regression model:

```
additive <- glm(choice ~ countries + sanctions, data= climateSupport,
               family = binomial(link="logit"))
summary(additive)
```

Table 1:

	<i>Dependent variable:</i>
	choice
countries80 of 192	0.336*** (0.054)
countries160 of 192	0.648*** (0.054)
sanctionsNone	-0.192*** (0.062)
sanctions15%	-0.325*** (0.062)
sanctions20%	-0.495*** (0.062)
Constant	-0.081 (0.053)
Observations	8,500
Log Likelihood	-5,784.130
Akaike Inf. Crit.	11,580.260
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

I use the `anova()` function in R to run a null hypothesis test and get a p-value:

```
nullModel <- glm(choice~1, data=climateSupport,
                  family = binomial(link = "logit"))
anova(nullModel, additive, test="LRT")
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Null Model	8499	11783.41			
Additive Model	8494	11568.26	5	215.15	0.0000

The p-value is below our critical threshold of 0.05 which means we can reject our null hypothesis and find support for the alternate hypothesis that our additive model with variables for countries and sanctions explains the variation in our outcome variable better than the null model.

## 2 Question 2

### 2.1 2a)

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)

Answer: As we have set our reference category to 5%, I interpret the coefficient for the variable 'sanctions15%'. Hence, increasing sanctions from 5% to 15%, on average, leads to a decrease of 0.325 in the log odds that an individual will support the policy, for all potential number of countries (20 of 192; 80 of 192; 160 of 192). This coefficient is highly significant at the 0.01 level.

### 2.2 2b)

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

I use the `predict()` function to answer this:

```
predict(additive, newdata = data.frame(countries="80 of 192",
sanctions="None"), type="response")
```

The estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions is 0.52.

### 2.3 2c)

Would the answers to 2a and 2b potentially change if we included the interaction term in this model? Why?

Answer: Yes, the answers could potentially change if there is a meaningful interaction between 'countries' and 'sanctions'. This could change the interpretation of the coefficients and predictions. Unlike the additive model where other variables are held constant, the interaction term allows for the possibility that the effect of one variable is dependent on the level of another variable. As shown in the Table below, none of the interaction term coefficients are significant at the critical threshold of 0.05 which suggests that they do not provide improve the model fit.

```
#Fitting an interactive model
```

```
interactive <- glm(choice ~ countries * sanctions, data = climateSupport,
family = binomial(link="logit"))
summary(interactive)
```

Table 2:

	<i>Dependent variable:</i>
	choice
countries80 of 192	0.470*** (0.109)
countries160 of 192	0.743*** (0.106)
sanctionsNone	-0.122 (0.105)
sanctions15%	-0.219** (0.107)
sanctions20%	-0.374*** (0.107)
countries80 of 192:sanctionsNone	-0.095 (0.152)
countries160 of 192:sanctionsNone	-0.130 (0.151)
countries80 of 192:sanctions15%	-0.147 (0.154)
countries160 of 192:sanctions15%	-0.182 (0.151)
countries80 of 192:sanctions20%	-0.292* (0.153)
countries160 of 192:sanctions20%	-0.073 (0.152)
Constant	-0.153** (0.073)
Observations	8,500
Log Likelihood	-5,780.983
Akaike Inf. Crit.	11,585.970
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Perform a test to see if including an interaction is appropriate  
 Answer: To determine if including an interaction term is appropriate, I conduct a likelihood ratio test:

```
anova2 <- anova(additive,interactive, test="LRT")
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Additive	8494	11568.26			
Interactive	8488	11561.97	6	6.29	0.3912

After running the likelihood-ratio test (LRT), we get a p-value of 0.3912 which is not below the critical threshold of 0.05, hence we cannot reject our null hypothesis that the the model without the interaction term (Additive model) is just as good as the model with the interaction term (Interactive model). As the LRT is not significant, we cannot conclude that the model with the interaction term provides a better fit to the data than the additive model. Hence, including an interaction term does not provide value for this model.