Problem Set 2

Applied Stats II

Due: February 28, 2022

Kathryn Glen

17331061

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]

Basic admin code at the beginning:

```
# remove objects
rm(list=ls())
# detach all libraries
detachAllPackages <- function() {
  basic.packages <- c("package:stats", "package:graphics", "package:grDevices
", "package:utils", "package:datasets", "package:methods", "package:base")
  package.list <- search()[ifelse(unlist(gregexpr("package:", search()))==1,
TRUE, FALSE)]
  package.list <- setdiff(package.list, basic.packages)
  if (length(package.list)>0)  for (package in package.list) detach(package,
character.only=TRUE)
}
detachAllPackages()

# load libraries
pkgTest <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[,  "Package"])]
```

```
  if (length(new.pkg))
    install.packages(new.pkg,  dependencies = TRUE)
  sapply(pkg,  require,  character.only = TRUE)
}

# here is where you load any necessary packages
# ex: stringr
# lapply(c("stringr"),  pkgTest)

lapply(c("rstudioapi",
         "mgcv",
         "gam",
         "ggplot2"),  pkgTest)
```

```
## Loading required package: rstudioapi
```

```
## Loading required package: mgcv
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-36. For overview type 'help("mgcv-package")'.
```

```
## Loading required package: gam
```

```
## Loading required package: splines
```

```
## Loading required package: foreach
```

```
## Warning: package 'foreach' was built under R version 4.1.2
```

```
## Loaded gam 1.20
```

```
##
## Attaching package: 'gam'
```

```
## The following objects are masked from 'package:mgcv':
##
##     gam, gam.control, gam.fit, s
```

```
## Loading required package: ggplot2
```

```
## [[1]]
## rstudioapi
##       TRUE
##
## [[2]]
## mgcv
## TRUE
##
## [[3]]
##  gam
## TRUE
##
```

```
## [[4]]
## ggplot2
##    TRUE

# set wd for current folder
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

Problem 1

```
# load data
load(url("https://github.com/ASDS-TCD/StatsII_Spring2022/blob/main/datasets/climateSupport.RData?raw=true"))
ClimSup <- climateSupport

View(ClimSup)
```

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.

Run some analysis of the data to better understand the variables:

```
unique(ClimSup$choice)

## [1] Not supported Supported
## Levels: Not supported Supported

# 1 = support, 0 = not support

unique(ClimSup$countries)

## [1] 80 of 192  160 of 192 20 of 192
## Levels: 20 of 192 < 80 of 192 < 160 of 192

# Countries = 3 levels : Levels: 20 of 192 < 80 of 192 < 160 of 192

unique(ClimSup$sanctions)

## [1] 15%  None 5%   20%
## Levels: None < 5% < 15% < 20%

# Sanctions =  3 Levels: None < 5% < 15% < 20%
```

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

Do GLM, binomial family:

```
glmChoice <- glm(ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions, fami
ly=binomial(logit))
summary(glmChoice)
```

```
##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions,
##     family = binomial(logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4259  -1.1480  -0.9444   1.1505   1.4298
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -0.005665   0.021971  -0.258 0.796517
## ClimSup$countries.L  0.458452   0.038101  12.033  < 2e-16 ***
## ClimSup$countries.Q -0.009950   0.038056  -0.261 0.793741
## ClimSup$sanctions.L -0.276332   0.043925  -6.291 3.15e-10 ***
## ClimSup$sanctions.Q -0.181086   0.043963  -4.119 3.80e-05 ***
## ClimSup$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
```

Problem: output including .L, .Q and .C at the end - indicating Linear, Quadratic and Cubic.

I do not understand how to interpret these in this format.

```
glmChoice <- glm(ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions, fami
ly=binomial(logit))
```

Stackexchange says the problem is that they are ordered, I need to unorder the variables by hand so I can see the labels on the factors.

Suggested code to fix: my.variable <- factor(my.variable, ordered=FALSE)

```
ClimSup$countries <-  factor(ClimSup$countries, ordered = FALSE)
ClimSup$sanctions <- factor(ClimSup$sanctions, ordered = FALSE)
```

Run glmChoice again.

```
glmChoice <- glm(ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions, fami
ly=binomial(logit))
summary(glmChoice)

##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions,
##     family = binomial(logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4259  -1.1480  -0.9444   1.1505   1.4298
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -0.27266    0.05360  -5.087 3.64e-07 ***
## ClimSup$countries80 of 192    0.33636    0.05380   6.252 4.05e-10 ***
## ClimSup$countries160 of 192   0.64835    0.05388  12.033  < 2e-16 ***
## ClimSup$sanctions5%           0.19186    0.06216   3.086  0.00203 **
## ClimSup$sanctions15%         -0.13325    0.06208  -2.146  0.03183 *
## ClimSup$sanctions20%         -0.30356    0.06209  -4.889 1.01e-06 ***
## ---
## 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
```

And they are now labelled so I can understand what it is trying to say.

Global Null Hypothesis

I tried your code, but you didn't show where you got glm.reduced and glm.full from?

anova(glm.reduced, glm.full, test = "Chisq")

So instead found this online

Significance for the overall model:

Null deviance: 11783 on 8499 degrees of freedom

Residual deviance: 11568 on 8494 degrees of freedom

```
1-pchisq(11783-11568, 8499-8494)
```

```
## [1] 0
```

But I only get this: 0

Attempt to create a glm null model by only having Choice (intercept)

```
glmnull <- glm(ClimSup$choice ~ 1, family=binomial(logit))
```

Running anova()

```
anova(glmnull, glmChoice, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: ClimSup$choice ~ 1
## Model 2: ClimSup$choice ~ ClimSup$countries + ClimSup$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
```

INTERPRETATION:

P-value is below 0.05. At least one predictor is reliable in the the model.

P-Value

I don't know if you mean something specific about providing "the" p-value. But all I can think of is including p-value in my discussion of the coefficients

```
summary(glmChoice)
```

```
##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions,
##     family = binomial(logit))
##
## 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)                  -0.27266    0.05360  -5.087 3.64e-07 ***
## ClimSup$countries80 of 192    0.33636    0.05380   6.252 4.05e-10 ***
## ClimSup$countries160 of 192   0.64835    0.05388  12.033  < 2e-16 ***
## ClimSup$sanctions5%           0.19186    0.06216   3.086  0.00203 **
## ClimSup$sanctions15%         -0.13325    0.06208  -2.146  0.03183 *
## ClimSup$sanctions20%         -0.30356    0.06209  -4.889 1.01e-06 ***
## ---
## 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
```

Note : (1 means support)

```
exp(-0.27266)
```

```
## [1] 0.7613516
```

In the lectures you didn't indicate how to interpret the negative factors as percentages, so I initially referred to e-0.27266 = 0.7613516 = 76%. However Dunteman & Ho indicates that you should take the difference from 0 as the percentage, making 0.7613516 = 100-76 = 24%.

20 of 192 countries participating is associated with a decrease in likelihood of an individual supporting a policy by e-0.27266 = 0.7613516 = 24%.

```
exp(0.33636)
```

```
## [1] 1.399843
```

```
exp(0.64835)
```

```
## [1] 1.912383
```

80 of 192 and 160 of 192 countries participating is associated with an increase in likelihood of an individual supporting a policy by e0.33636(= 1.399843 = 40%) and e0.64835 (= 1.912383 = 91%) respectively.

```
exp(0.19186)
```

```
## [1] 1.211501
```

5% sanctions is associated with an increase in likelihood of an individual supporting a policy by e0.19186 = 1.211501 = 21% increase.

```
exp(-0.13325)
```

```
## [1] 0.8752463
```

```
exp(-0.30356)
```

```
## [1] 0.7381856
```

15% and 20% sanctions is associated with a decrease in likelihood of an individual supporting a policy by e-0.13325 (= 0.8752463 = 12%) and e-0.30356 (= 0.7381856 = 26%) respectively.

All of them are significant as their p-value is above .05.

Conclusion:

Small numbers of participating countries in a policy means individuals will not support while higher numbers of participating countries leads to the likelihood of more widespread support.

In contrast, individuals will be more likely to support a policy if the sanctions are lower, while higher sanctions is associated with a lower likelihood of individual support.

```
glmChoice$coefficients
```

```
##                  (Intercept)  ClimSup$countries80 of 192
##                   -0.2726631                   0.3363609
## ClimSup$countries160 of 192         ClimSup$sanctions5%
##                    0.6483497                   0.1918553
##        ClimSup$sanctions15%        ClimSup$sanctions20%
##                   -0.1332475                  -0.3035641
```

(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)

Need to do an interaction to see how country participation and sanctions interact

```
glmint <- glm(ClimSup$choice ~ ClimSup$countries * ClimSup$sanctions, family=
binomial(logit))
summary(glmint)

##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries * ClimSup$sanctions,
##     family = binomial(logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4359  -1.1570  -0.9632   1.1349   1.4079
##
## Coefficients:
##                                                 Estimate Std. Error z val
ue
## (Intercept)                                     -0.27469    0.07534  -3.6
46
## ClimSup$countries80 of 192                       0.37562    0.10627   3.5
35
## ClimSup$countries160 of 192                      0.61266    0.10801   5.6
72
## ClimSup$sanctions5%                              0.12179    0.10518   1.1
58
## ClimSup$sanctions15%                            -0.09687    0.10822  -0.8
95
## ClimSup$sanctions20%                            -0.25260    0.10806  -2.3
38
## ClimSup$countries80 of 192:ClimSup$sanctions5%   0.09471    0.15232   0.6
22
## ClimSup$countries160 of 192:ClimSup$sanctions5%  0.13009    0.15103   0.8
61
## ClimSup$countries80 of 192:ClimSup$sanctions15% -0.05229    0.15167  -0.3
45
## ClimSup$countries160 of 192:ClimSup$sanctions15% -0.05165    0.15267  -0.3
38
## ClimSup$countries80 of 192:ClimSup$sanctions20% -0.19721    0.15104  -1.3
06
## ClimSup$countries160 of 192:ClimSup$sanctions20%  0.05688    0.15367   0.3
70
##                                                 Pr(>|z|)
## (Intercept)                                     0.000267 ***
## ClimSup$countries80 of 192                      0.000408 ***
## ClimSup$countries160 of 192                     1.41e-08 ***
## ClimSup$sanctions5%                             0.246909
## ClimSup$sanctions15%                            0.370723
## ClimSup$sanctions20%                            0.019412 *
## ClimSup$countries80 of 192:ClimSup$sanctions5%  0.534071
## ClimSup$countries160 of 192:ClimSup$sanctions5% 0.389063
## ClimSup$countries80 of 192:ClimSup$sanctions15% 0.730262
```

```
## ClimSup$countries160 of 192:ClimSup$sanctions15% 0.735136
## ClimSup$countries80 of 192:ClimSup$sanctions20%  0.191675
## ClimSup$countries160 of 192:ClimSup$sanctions20% 0.711279
## ---
## 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: 11562  on 8488  degrees of freedom
## AIC: 11586
##
## Number of Fisher Scoring iterations: 4
```

ClimSup$countries160 of 192:ClimSup$sanctions5% 0.13009
ClimSup$countries160 of 192:ClimSup$sanctions15% -0.05165

```
exp(0.13009) # 1.138931 = 13% increase

## [1] 1.138931

exp(-0.05165) # 0.9496612 = 6% decrease

## [1] 0.9496612
```

Increasing sanctions from 5% to 15% for the policy in which nearly all countries participated is associated with a shift from the likelihood of an individual supporting a policy increasing by 13% at 5% to the likelihood of an individual supporting the policy decreasing by 6% at 15%.

(b) For the policy in which very few countries participate [20 of 192], how does in- creasing sanctions from 5% to 15% change the odds that an individual will support the policy? (Interpretation of a coefficient)

ClimSup$sanctions5$sanctions15% -0.09687

```
exp(0.12179) # 1.129517 # 13%

## [1] 1.129517

exp(-0.09687) # 0.907674 # 9%

## [1] 0.907674
```

Increasing sanctions from 5% to 15% for the policy in which very few countries participated is associated with a shift from the likelihood of an individual supporting a policy increasing by 13% at 5% to the likelihood of an individual supporting the policy decreasing by 9% at 15%.

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

```
exp(0.37562) # 1.455894
```

```
## [1] 1.455894
```

Probability - increases by 46%

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

```
summary(glmChoice)
```

```
##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions,
##     family = binomial(logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4259  -1.1480  -0.9444   1.1505   1.4298
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -0.27266    0.05360  -5.087 3.64e-07 ***
## ClimSup$countries80 of 192    0.33636    0.05380   6.252 4.05e-10 ***
## ClimSup$countries160 of 192   0.64835    0.05388  12.033  < 2e-16 ***
## ClimSup$sanctions5%           0.19186    0.06216   3.086  0.00203 **
## ClimSup$sanctions15%         -0.13325    0.06208  -2.146  0.03183 *
## ClimSup$sanctions20%         -0.30356    0.06209  -4.889 1.01e-06 ***
## ---
## 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
```

```
summary(glmint)
```

```
##
## Call:
## glm(formula = ClimSup$choice ~ ClimSup$countries * ClimSup$sanctions,
##     family = binomial(logit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4359  -1.1570  -0.9632   1.1349   1.4079
##
## Coefficients:
##                                                 Estimate Std. Error z val
ue
## (Intercept)                                     -0.27469    0.07534  -3.6
46
## ClimSup$countries80 of 192                       0.37562    0.10627   3.5
35
## ClimSup$countries160 of 192                      0.61266    0.10801   5.6
72
## ClimSup$sanctions5%                              0.12179    0.10518   1.1
58
## ClimSup$sanctions15%                            -0.09687    0.10822  -0.8
95
## ClimSup$sanctions20%                            -0.25260    0.10806  -2.3
38
## ClimSup$countries80 of 192:ClimSup$sanctions5%   0.09471    0.15232   0.6
22
## ClimSup$countries160 of 192:ClimSup$sanctions5%  0.13009    0.15103   0.8
61
## ClimSup$countries80 of 192:ClimSup$sanctions15% -0.05229    0.15167  -0.3
45
## ClimSup$countries160 of 192:ClimSup$sanctions15% -0.05165    0.15267  -0.3
38
## ClimSup$countries80 of 192:ClimSup$sanctions20% -0.19721    0.15104  -1.3
06
## ClimSup$countries160 of 192:ClimSup$sanctions20%  0.05688    0.15367   0.3
70
##                                                 Pr(>|z|)
## (Intercept)                                     0.000267 ***
## ClimSup$countries80 of 192                      0.000408 ***
## ClimSup$countries160 of 192                     1.41e-08 ***
## ClimSup$sanctions5%                             0.246909
## ClimSup$sanctions15%                            0.370723
## ClimSup$sanctions20%                            0.019412 *
## ClimSup$countries80 of 192:ClimSup$sanctions5%  0.534071
## ClimSup$countries160 of 192:ClimSup$sanctions5% 0.389063
## ClimSup$countries80 of 192:ClimSup$sanctions15% 0.730262
## ClimSup$countries160 of 192:ClimSup$sanctions15% 0.735136
## ClimSup$countries80 of 192:ClimSup$sanctions20% 0.191675
## ClimSup$countries160 of 192:ClimSup$sanctions20% 0.711279
## ---
```

```
## 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: 11562  on 8488  degrees of freedom
## AIC: 11586
##
## Number of Fisher Scoring iterations: 4
```

I don't understand this question. I already have included the interaction term, it is what my answers are based on so they couldn't change.

I don't know how I would have answered the first two questions if I hadn't included the interaction term.

It is also confusing because it refers to 2a and 2b while these questions are only 1a and 1b. Maybe it is from a previous draft of the problem set?

• Perform a test to see if including an interaction is appropriate.

Compare the additive and multiplicative models using anova to judge whether the deviance is large enough to judge whether using an interaction would give us different information and provide a better fit to the data.

```
anova(glmChoice, glmint, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: ClimSup$choice ~ ClimSup$countries + ClimSup$sanctions
## Model 2: ClimSup$choice ~ ClimSup$countries * ClimSup$sanctions
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      8494      11568
## 2      8488      11562  6   6.2928   0.3912
```

The deviance is 6.2928, which is large enough a difference between the slopes of the two models that I think using an interaction is appropriate.