Problem Set 2

Doireanna Craven

Due: February 28, 2022

Instructions

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.

   First, let's find out how the data is stored:

```r
str(data)
# chr [1:2] "climateSupport" ".Random.seed" These are the objects held
# in the RData file

str(climateSupport)
# 'data.frame': 8500 obs. of  3 variables:
# $ choice    : Factor w/ 2 levels "Not supported",..: 1 1 1 1 1 1 2 1 2 1 ...
# $ countries: Ord.factor w/ 3 levels "20 of 192"<"80 of 192"<..: 2 3 3 2 3 1
# $ sanctions: Ord.factor w/ 4 levels "None"<"5%"<"15%"<..: 3 3 1 3 2 3 2 4 2

# Three columns, variables are of the data structure type 'factors'
# Check the order of factors
table(climateSupport$countries)
table(climateSupport$sanctions)
```

   The inputs are stored as factors, which is helpful. One option is to use `acm.disjonctif(climateSupport)` to expand the table, otherwise I have had to un-order the factor before running the GLM.

```r
# Make factors unordered
climateSupport$countries <- factor(climateSupport$countries, ordered=FALSE)
climateSupport$sanctions <- factor(climateSupport$sanctions, ordered=FALSE)

# Rum GLM
glm <- glm(choice ~ ., data = climateSupport, family = binomial(link = "logit"))
summary(glm)
```

   Baseline categories are 20 countries, and no sanctions, (default is the first factors that appear in `table()`).

Before we interpret the output of the GLM, we can try to discount the global null hypothesis. The null hypothesis is that there is no relationship between the response variable and any of the input variables. To test this hypothesis we need a reduced model.

```
# Reduced model
glm_red <- glm(choice ~ 1, data = climateSupport, family = binomial(link = "logit"))
sum_glm_red <- summary(glm_red)

test_null_hyp <- anova(glm_red, glm, test = "Chisq")
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	8499	11783.41	NA	NA	NA
2	8494	11568.26	5	215.15	1.624583e-44

The anova output gives a p-value in the region of 10^{-44} therefore we can reject the null hypothesis at pretty much any level of alpha.

Now to interpret the results of the GLM:

	Estimate	Std. Error	z value	Pr(> | z|)
(Intercept)	-0.273	0.054	-5.087	0.00000
countries80 of 192	0.336	0.054	6.252	0
countries160 of 192	0.648	0.054	12.033	0
sanctions5%	0.192	0.062	3.086	0.002
sanctions15%	-0.133	0.062	-2.146	0.032
sanctions20%	-0.304	0.062	-4.889	0.00000

All results following are given to 3 significant figures.
Odds = $e^{logodds}$.
Probability = odds/(1 + odds).

The probabilities below are calculated as follows, i.e.:
Baseline odds:

```
exp(sum_glm$coefficients[1,1])/(1 + exp(sum_glm$coefficients[1,1]))
```

80 countries, no sanctions:

```
bl = sum_glm$coefficients[1,1]
exp(bl + sum_glm$coefficients[2,1])/(1 + exp(bl + sum_glm$coefficients
    [2,1]))
```

When 20 countries are included with zero sanctions (baseline), the odds that an individual supports a policy are $e^{-0.273} = 0.761$. The probability is 0.432.

When 80 countries are included the baseline odds are multiplied by a factor of $e^{0.336} = 1.40$ to 3 s.f. The probability that an individual supports a policy (80 countries, zero sanctions) is 0.516.

When 160 countries are included the baseline odds are multiplied by a factor of $e^{0.648} = 1.91$. The probability that an individual supports a policy (160 countries, zero sanctions) is 0.593.

When sanctions are introduced at 5% the baseline odds are multiplied by a factor of $e^{0.192} = 1.21$. The probability that an individual supports a policy (20 countries, 5% sanctions) is 0.480.

When sanctions are introduced at 15% the baseline odds are multiplied by a factor of $e^{-0.133} = 0.875$. The probability that an individual supports a policy (20 countries, 15% sanctions) is 0.400.

When sanctions are introduced at 20% the baseline odds are multiplied by a factor of $e^{-0.304} = 0.738$. The probability that an individual supports a policy (20 countries, 20% sanctions) is 0.360.

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)

   There are two ways to find this answer. From the GLM generated above, we can bring the log odds at 5% sanctions down to baseline level of zero sanctions by subtracting 0.192, and then factor in 15% sanctions by subtracting 0.133. (The effect of switching sanctions in this model does not depend on the number of countries.) The overall change in the odds is a multiplication by a factor of $e^{-0.325}$ = 0.72, i.e. a 28% reduction in odds.

Another way to answer this would be to generate another GLM with baseline 160 countries and sanctions 5% by re-levelling the factors, and finding the coefficient corresponding to 15% sanctions. I need practice with R so I'm going to use this method.

```r
# How does baseline countries = 160 and sanctions 5% behave?
climateSupport2 <- copy(climateSupport)
climateSupport2$countries <- relevel(climateSupport2$countries, ref =
    "160 of 192")
climateSupport2$sanctions <- relevel(climateSupport2$sanctions, ref =
    "5%")
glm_2 <- glm(choice ~ ., data = climateSupport2, family = binomial(
    link = "logit"))
sum_glm_2 <- summary(glm_2)
```

	Estimate	Std. Error	z value	Pr(> | z|)
(Intercept)	0.568	0.054	10.551	0
countries20 of 192	-0.648	0.054	-12.033	0
countries80 of 192	-0.312	0.054	-5.792	0
sanctionsNone	-0.192	0.062	-3.086	0.002
sanctions15%	$\boxed{-0.325}$	0.062	-5.224	0.00000
sanctions20%	-0.495	0.062	-7.955	0

This gives the same change in odds as calculated above when sanctions are increased from 5% to 15%: a multiplication by a factor of $e^{-0.325} = 0.72$.

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

This change is independent of the number of countries therefore it will match the answer to the last question. The overall change in the odds is a multiplication by a factor of $e^{-0.325} = 0.72$, i.e. a 28% reduction in odds. To confirm:

```
# How does baseline countries = 20 and sanctions 5% behave?
climateSupport3 <- copy(climateSupport)
climateSupport3$countries <- relevel(climateSupport3$countries, ref = "20 of 192")
climateSupport3$sanctions <- relevel(climateSupport3$sanctions, ref = "5%")
glm_3 <- glm(choice ~ ., data = climateSupport3, family = binomial(link = "logit"))
sum_glm_3 <- summary(glm_3)
```

	Estimate	Std. Error	z value	Pr(> | z|)
(Intercept)	-0.081	0.053	-1.520	0.128
countries80 of 192	0.336	0.054	6.252	0
countries160 of 192	0.648	0.054	12.033	0
sanctionsNone	-0.192	0.062	-3.086	0.002
sanctions15%	$\boxed{-0.325}$	0.062	-5.224	0.00000
sanctions20%	-0.495	0.062	-7.955	0

Confirmed.

(c) What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions? This can be calculated from the original GLM by adding the intercept to the coefficient for 80 countries.

Log-odds = - 0.273 + 0.336 = 0.064.

Odds = $e^{0.064}$ = 1.07.

```
sum <- sum_glm$coefficients[1,1] + sum_glm$coefficients[2,1]
exp(sum)/(1 + exp(sum))
```

Probability = 0.516.

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

Potentially yes, if the interaction terms prove to be significant. As the model stands, changes in sanctions do not vary depending on how many countries are included.

```r
# Include interaction terms.
glm_int <- glm(choice ~ countries + sanctions + countries*sanctions,
    data = climateSupport, family = binomial(link = "logit") )
summ_glm_int <- summary(glm_int)
```

	Estimate	Std. Error	z value	Pr(> | z|)
(Intercept)	-0.275	0.075	-3.646	0.0003
countries80 of 192	0.376	0.106	3.535	0.0004
countries160 of 192	0.613	0.108	5.672	0
sanctions5%	0.122	0.105	1.158	0.247
sanctions15%	-0.097	0.108	-0.895	0.371
sanctions20%	-0.253	0.108	-2.338	0.019
countries80 of 192:sanctions5%	0.095	0.152	0.622	0.534
countries160 of 192:sanctions5%	0.130	0.151	0.861	0.389
countries80 of 192:sanctions15%	-0.052	0.152	-0.345	0.730
countries160 of 192:sanctions15%	-0.052	0.153	-0.338	0.735
countries80 of 192:sanctions20%	-0.197	0.151	-1.306	0.192
countries160 of 192:sanctions20%	0.057	0.154	0.370	0.711

None of the interactive terms are statistically significant.

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

```r
test_int <- anova(glm, glm_int, test = "Chisq")
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	8494	11568.26	NA	NA	NA
2	8488	11561.97	6	6.292831	0.391199

The p-value is 0.39, so we cannot reject the null hypothesis that there is no difference between the additive and interactive models, i.e., the interactive terms have not added anything.