

Problem Set 2

Applied Stats II- Melissa Campbell

Due: February 19, 2023

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in **R**, please include the code you used to get your answers. Please also include the **.R** file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub in **.pdf** form.
- This problem set is due before 23:59 on Sunday February 19, 2023. No late assignments will be accepted.

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.

- (a) Here is an additive model, based on the number of countries participating and the possible sanctions for non-compliance:

```

1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2023/blob/main/
  datasets/climateSupport.RData?raw=true"))
3
4 ls()
5 readRDS(file = "climateSupport.RData")
6 summary(climateSupport)
7 summary(con)
8 head(climateSupport)
9 tail(climateSupport)
10 colnames(climateSupport)
11
12 # Use ifelse() with as.logical()...
13 as.logical(ifelse(climateSupport$choice == "Supported", 1, 0))
14 as.numeric(climateSupport$choice)
15 class(climateSupport$sanctions)
16 as.numeric(climateSupport$countries)
17 as.numeric(climateSupport$sanctions)
18 class(climateSupport$countries)
19 class(climateSupport$sanctions)
20 ## a) Run the logit regression
21 mod <- glm(choice ~ .,
22            data = climateSupport,
23            family = "binomial")
24
25
26 summary(mod)
27

```

- Call: glm(formula = choice ~ ., family = "binomial", data = climateSupport)
 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 1e-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

[1] 4 The number of iterations is 4

```

1 ## Likelihood ratio test
2 # Create a null model
3 nullMod <- glm(choice ~ 1, # 1 = fit an intercept only )
4           data = climateSupport ,
5           family = "binomial")
6 Call:  glm(formula = choice ~ 1, family = "binomial", data =
         climateSupport)
7
8 Coefficients:
9 (Intercept)
10 -0.006588
11
12 Degrees of Freedom: 8499 Total (i.e. Null);  8499 Residual
13 Null Deviance:      11780
14 Residual Deviance: 11780  AIC: 11790
15
16 # Run an anova test on the model compared to the null model
17 anova(nullMod, mod, test = "Chisq")
18 anova(nullMod, mod, test = "LRT") # LRT is equivalent
19
20 ## Extracting confidence intervals (of the coefficients)
21
22 exp(confint(mod)) # transform to odds ratio using exp()
23
24 # data.frame of confidence intervals and coefficients. calculates the
    profile likelihood confidence interval
25 confMod <- data.frame(cbind(lower = exp(confint(mod)[,1]) ,
26                             coefs = exp(coef(mod)) ,
27                             upper = exp(confint(mod)[,2])) )
28 confMod
29           lower      coefs      upper
30 (Intercept) 0.9524387 0.9943507 1.0381058
31 countries.L 1.4679656 1.5816245 1.7044456
32 countries.Q 0.9189295 0.9900994 1.0667733
33 sanctions.L 0.6959419 0.7585609 0.8267142
34 sanctions.Q 0.7654570 0.8343637 0.9094241
35 sanctions.C 1.0661324 1.1620743 1.2667989
36 summary(confMod)
37           lower      coefs      upper
38 Min.      :0.6959   Min.      :0.7586   Min.      :0.8267
39 1st Qu.:0.8038   1st Qu.:0.8733   1st Qu.:0.9416

```

40	Median	:0.9357	Median	:0.9922	Median	:1.0524
41	Mean	:0.9778	Mean	:1.0535	Mean	:1.1354
42	3rd Qu.	:1.0377	3rd Qu.	:1.1201	3rd Qu.	:1.2168
43	Max.	:1.4680	Max.	:1.5816	Max.	:1.7044

2. The global null hypothesis is that neither the number of countries involved nor the imposition of sanctions have an effect on support or environmental policies or agreements in this study. The coefficient estimate in the output indicates the average change in the log odds of the response variable associated with a one unit increase in each predictor variable. Here, mod2 shows that a one unit increase in the predictor variable choice is associated with an average change of -0.276332 in the log odds of the response variable. The p-value associated with the z value for the countries L variable is 2e-16 . Since this value is less than .05, we would say it is a statistically significant predictor variable in the model. The p-value associated with the z value for the countries Q variable is 0.793741 . Since this value is more than .05, we can say it is not a statistically significant predictor variable in the model. The p-values associated with Sanctions L, Q and C are all below .05, and so are all statistically significant, being respectively 3.15e-10 ***, 3.80e-05 and 0.000639.

3. 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 2.5 (Intercept) 0.9524387 1.0381058 countries.L
1.4679656 1.7044456 countries.Q 0.9189295 1.0667733 sanctions.L 0.6959419 0.8267142
sanctions.Q 0.7654570 0.9094241 sanctions.C 1.0661324 1.2667989

```

1
2 # Using model.matrix to check for each unique level of sanctions, and
   each unique level of country participation
3
4 model.matrix( ~ unique(sanctions), data = climateSupport)
5 model.matrix( ~ unique(countries), data = climateSupport)
6
7 (Intercept) unique(countries).L unique(countries).Q
8 1          1          -7.850462e-17          -0.8164966
9 2          1           7.071068e-01           0.4082483
10 3          1          -7.071068e-01           0.4082483
11 attr(,"assign")
12 [1] 0 1 1
13 attr(,"contrasts")
14 attr(,"contrasts") "unique(countries)"
15 [1] "contr.poly"
16

```

4.

```

5: with(climateSupport , expand.grid(choice = unique(choice),
2:                               countries = unique(countries),
3:                               sanctions = unique(sanctions)))
4
5

```

Table 1: Climate Support

	choice	countries	sanctions (percent)
	Not supported	80 of 192	15
	Supported	80 of 192	15
	Not supported	160 of 192	15
	Supported	160 of 192	15
	Not supported	20 of 192	15
	Supported	20 of 192	15
	Not supported	80 of 192	None
	Supported	80 of 192	None
	Not supported	160 of 192	None
	Supported	160 of 192	None
	Not supported	20 of 192	None
6.	Supported	20 of 192	None
	Not supported	80 of 192	5
	Supported	80 of 192	5
	Not supported	160 of 192	5
	Supported	160 of 192	5
	Not supported	20 of 192	5
	Supported	20 of 192	5
	Not supported	80 of 192	20
	Supported	80 of 192	20
	Not supported	160 of 192	20
	Supported	160 of 192	20
	Not supported	20 of 192	20
	Supported	20 of 192	20

```

1 predicted_data <- with(climateSupport , expand.grid(choice = unique(choice
2:                               ),
3:                               countries = unique(
4:                               countries),
5:                               sanctions = unique(
6:                               sanctions)))
7
8 predicted_data <- cbind(predicted_data , predict(mod2,
9:                               newdata = predicted_data ,
10:                              type = "response",
11:                              se = TRUE))

```

```

9
10 #Fill out the confidence intervals and predicted probability
11 predicted_data <- within(predicted_data,
12                           {PredictedProb <- plogis(fit)
13                            LL <- plogis(fit - (1.96 * se.fit))
14                            UL <- plogis(fit + (1.96 * se.fit))
15                           })
16 predicted_data
17
18      choice countries sanctions      fit      se.fit residual.scale
19 1  Not supported  80 of 192      15% 0.4826196 0.01339475      1
20   0.6245423
21 2   Supported  80 of 192      15% 0.4826196 0.01339475      1
22   0.6245423
23 3  Not supported 160 of 192      15% 0.5603146 0.01320502      1
24   0.6424920
25 4   Supported 160 of 192      15% 0.5603146 0.01320502      1
26   0.6424920
27 5  Not supported  20 of 192      15% 0.3998931 0.01301632      1
28   0.6047759
29 6   Supported  20 of 192      15% 0.3998931 0.01301632      1
30   0.6047759
31 7  Not supported  80 of 192     None 0.5159191 0.01335758      1
32   0.6323008
33 8   Supported  80 of 192     None 0.5159191 0.01335758      1
34   0.6323008
35 9  Not supported 160 of 192     None 0.5928323 0.01307316      1
36   0.6498673
37 10 Supported 160 of 192     None 0.5928323 0.01307316      1
38   0.6498673
39 11 Not supported  20 of 192     None 0.4322534 0.01315368      1
40   0.6125478
41 12 Supported  20 of 192     None 0.4322534 0.01315368      1
42   0.6125478
43 13 Not supported  80 of 192      5% 0.5635428 0.01347950      1
44   0.6433566
45 14 Supported  80 of 192      5% 0.5635428 0.01347950      1
46   0.6433566
47 15 Not supported 160 of 192      5% 0.6381958 0.01242082      1
48   0.6598309
49 16 Supported 160 of 192      5% 0.6381958 0.01242082      1
50   0.6598309
51 17 Not supported  20 of 192      5% 0.4798090 0.01326799      1
52   0.6238247
53 18 Supported  20 of 192      5% 0.4798090 0.01326799      1
54   0.6238247
55 19 Not supported  80 of 192     20% 0.4403193 0.01312026      1
56   0.6144449
57 20 Supported  80 of 192     20% 0.4403193 0.01312026      1
58   0.6144449

```

39	21	Not supported	160 of 192	20%	0.5180228	0.01349899	1
		0.6328542					
40	22	Supported	160 of 192	20%	0.5180228	0.01349899	1
		0.6328542					
41	23	Not supported	20 of 192	20%	0.3598012	0.01249970	1
		0.5949099					
42	24	Supported	20 of 192	20%	0.3598012	0.01249970	1
		0.5949099					
43		LL PredictedProb					
44	1	0.6121517	0.6183663				
45	2	0.6121517	0.6183663				
46	3	0.6305164	0.6365253				
47	4	0.6305164	0.6365253				
48	5	0.5925172	0.5986620				
49	6	0.5925172	0.5986620				
50	7	0.6200448	0.6261930				
51	8	0.6200448	0.6261930				
52	9	0.6381189	0.6440148				
53	10	0.6381189	0.6440148				
54	11	0.6002417	0.6064116				
55	12	0.6002417	0.6064116				
56	13	0.6311429	0.6372719				
57	14	0.6311429	0.6372719				
58	15	0.6488188	0.6543455				
59	16	0.6488188	0.6543455				
60	17	0.6115432	0.6177028				
61	18	0.6115432	0.6177028				
62	19	0.6021912	0.6083351				
63	20	0.6021912	0.6083351				
64	21	0.6204750	0.6266853				
65	22	0.6204750	0.6266853				
66	23	0.5830488	0.5889923				
67	24	0.5830488	0.5889923				

7. If any of the explanatory variables are significant in this model, then:

- 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)
- The estimated probability would be 0.6598309
- What is the estimated probability that an individual will support a policy if there are 80 of 192 countries participating with no sanctions?
- The estimated probability would be 0.6323008
- Would the answers to 2a and 2b potentially change if we included the interaction term in this model? Why?
 - Perform a test to see if including an interaction is appropriate.

```

1 mod3 <- glm(choice ~ sanctions + countries + countries *
2   sanctions ,
3   data = climateSupport ,
4   family = "binomial")
5 mod3
6 Call:  glm(formula = choice ~ sanctions + countries + countries *
7   sanctions ,
8   family = "binomial" , data = climateSupport)
9
10 Coefficients:
11 (Intercept)          sanctions.L
12 sanctions.Q          -0.003809          -0.274221
13 -0.182289
14 sanctions.C          countries.L
15 countries.Q          0.153245          0.457140
16 -0.011167
17 sanctions.L:countries.L  sanctions.Q:countries.L  sanctions.C:
18 countries.L
19 -0.001754          -0.007622
20 0.095197
21 sanctions.L:countries.Q  sanctions.Q:countries.Q  sanctions.C:
22 countries.Q
23 0.133840          0.093425
24 0.010449
25
26 Degrees of Freedom: 8499 Total (i.e. Null); 8488 Residual
27 Null Deviance: 11780
28 Residual Deviance: 11560 AIC: 11590
29
30 mod4 <- glm(choice ~ sanctions + countries ,
31 + data = climateSupport ,
32 + family = "binomial")
33 > mod4
34
35 Call:  glm(formula = choice ~ sanctions + countries , family = "
36 binomial" ,
37 data = climateSupport)
38
39 Coefficients:
40 (Intercept)  sanctions.L  sanctions.Q  sanctions.C  countries.L
41 -0.005665    -0.276332    -0.181086     0.150207     0.458452
42 countries.Q
43 -0.009950
44
45 Degrees of Freedom: 8499 Total (i.e. Null); 8494 Residual
46 Null Deviance: 11780
47 Residual Deviance: 11570 AIC: 11580

```

- Without the interaction the AIC is 11580 and with it the AIC is 11590, so

they are very similar. This suggests that the interaction term is not needed in the model.