

Answer Key: Problem Set 2

Applied Stats II

Jeffrey Ziegler

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 18, 2024. 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.

We'll first fit our additive model, and use the `summary()` output given by R to get the deviance from the null and full model for our likelihood ratio test. We can also do this manually, which you can see below, by running a model with no covariates.

```
1 # load data
2 load(url("https://github.com/ASDS-TCD/StatsII_Spring2024/blob/main/
  datasets/climateSupport.RData?raw=true"))
3 # change baseline category to answer Qs
4 climateSupport$sanctions <- relevel(factor(climateSupport$sanctions,
  ordered = F), ref="5%")
5 climateSupport$countries <- relevel(factor(climateSupport$countries,
  ordered = F), ref="20 of 192")
6
7 # run additive model
8 climate_logit <- glm(choice ~ countries+sanctions, data=climateSupport,
  family = binomial(link = "logit"))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.08081	0.05316	-1.520	0.12848
countries80 of 192	0.33636	0.05380	6.252	4.05e-10 ***
countries160 of 192	0.64835	0.05388	12.033	< 2e-16 ***
sanctionsNone	-0.19186	0.06216	-3.086	0.00203 **
sanctions15%	-0.32510	0.06224	-5.224	1.76e-07 ***
sanctions20%	-0.49542	0.06228	-7.955	1.79e-15 ***

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

```
1 # (1a)
2 # option #1: use info from summary() output
3 # set up hypothesis test yourself & reference chi-sq distribution
4 pchisq(11783-11568, 5, lower.tail = F)
```

```
[1] 1.749304e-44
```

```
1 # option #2: manually run null model & anova
2 null_logit <- glm(choice~1, data=climateSupport, family = binomial(link =
  "logit"))
3 anova(null_logit, climate_logit, test="LRT")
```

```
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 ***

Either way we go about it, we get a p-value of nearly zero which is below our critical threshold of 0.05, suggesting that we can reject the null hypothesis that neither variable increases our model fit (i.e., the explained variation of our outcome is reliably better with the two variables included in our model).

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)*

Column 1 in Table 1 displays the estimated coefficients for the additive model that we ran in part #1. Remember, this is an additive model, so for all potential policies (20 of 192; 80 of 192; 160 of 192) going from our new baseline category (5%) to 15% in sanctions, on average, results in a decrease of 0.33 in the log odds that a participant will prefer a given policy.

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

```
1 # (2c)
2 # option #1: predict function
3 round(predict(climate_logit, newdata = data.frame(countries="80 of
  192", sanctions="None"), type="response"), 2)
```

```
0.52
```

```
1 # option #2: predicted probability "by hand"
2 exp_coefs <- exp(coef(climate_logit)[1]+(coef(climate_logit)[2]*1)+(
  coef(climate_logit)[3]*0)+(coef(climate_logit)[4]*1)+(coef(
  climate_logit)[5]*0)+(coef(climate_logit)[6]*0))
3 round(exp_coefs/(1+exp_coefs), 2)
```

```
0.52
```

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

The answer to 2a and 2b could potentially change because we would be estimating the slope, as well as the intercept, for two fitted lines, one for each policy type with different #s of participating countries [20 of 192; 80 of 192; 160 of 192]. Let's run it in R and see if there's any substantive difference.

```
1 # (2d)
2 # estimate interactive model
3 climate_logit_interact <- glm(choice ~ countries * sanctions, data =
  climateSupport, family = binomial(link = "logit"))
```

We can see, however, in Column 2 of Table 1 that none of the interactive effects are statistically different from zero. This is our first indication that there is not a differentiable effect of the # of participating countries by level of sanctions.

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

```
1 # compare to additive model using LRT
2 anova(climate_logit, climate_logit_interact, test="LRT")
```

	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

We get a p-value that is not below our critical threshold of 0.05, so we cannot reject the null hypothesis that the interactive model does not improve our fit over the additive model (i.e., there does not appear to be sufficient evidence to justify an interactive effect of the # of participating countries by sanction level).

Table 1: Additive and interactive model of `choice ~ countries+sanctions`

	Model 1	Model 2
(Intercept)	−0.08 (0.05)	−0.15* (0.07)
countries80 of 192	0.34*** (0.05)	0.47*** (0.11)
countries160 of 192	0.65*** (0.05)	0.74*** (0.11)
sanctionsNone	−0.19** (0.06)	−0.12 (0.11)
sanctions15%	−0.33*** (0.06)	−0.22* (0.11)
sanctions20%	−0.50*** (0.06)	−0.37*** (0.11)
countries80 of 192:sanctionsNone		−0.09 (0.15)
countries160 of 192:sanctionsNone		−0.13 (0.15)
countries80 of 192:sanctions15%		−0.15 (0.15)
countries160 of 192:sanctions15%		−0.18 (0.15)
countries80 of 192:sanctions20%		−0.29 (0.15)
countries160 of 192:sanctions20%		−0.07 (0.15)
AIC	11580.26	11585.97
BIC	11622.55	11670.54
Log Likelihood	−5784.13	−5780.98
Deviance	11568.26	11561.97
N	8500	8500

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$