Applied Stats II Problem Set 2

Qi Liu

17/02/2026

Setup (key data preparation steps)

The dataset is loaded from `climateSupport.RData`. I recode the response to a binary indicator and set factor baselines to match the assignment design (`countries = 20 of 192`, `sanctions = None`). The full script uses a robust path search to run on any computer; here I show the essential transformations.

```
1   # key preparation steps (see PS02.R for robust path loading)
2   # load("climateSupport.RData")  # full script loads safely from file path
3
4   # recode response: Supported -> 1, Not supported -> 0
5   choice_chr <- tolower(trimws(as.character(climateSupport$choice)))
6   climateSupport$choice01 <- dplyr::case_when(
7   choice_chr == "supported" ~ 1L,
8   choice_chr == "not supported" ~ 0L,
9   TRUE ~ NA_integer_
10  )
11
12  # ensure factors + set reference categories
13  climateSupport$countries <- factor(as.character(climateSupport$countries),
      ordered = FALSE)
14  climateSupport$sanctions <- factor(as.character(climateSupport$sanctions),
      ordered = FALSE)
15  climateSupport$countries <- relevel(climateSupport$countries, ref = "20 of
      192")
16  climateSupport$sanctions <- relevel(climateSupport$sanctions, ref = "None"
      )
17
```

Question 1

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

Model. I estimate an additive logistic regression model with reference categories `countries`

= 20 of 192 and `sanctions = None`:

$$\text{logit}\{P(\texttt{choice} = 1)\} = \beta_0 + \beta_{80}\mathbb{I}(\texttt{countries} = 80/192) + \beta_{160}\mathbb{I}(\texttt{countries} = 160/192) + \gamma_5\mathbb{I}(\texttt{sanctions} = 5$$

Global null hypothesis.

$$H_0 : \beta_{80} = \beta_{160} = \gamma_5 = \gamma_{15} = \gamma_{20} = 0.$$

I test this using a likelihood ratio test comparing the intercept-only model to the additive model.

R code (key steps).

```
# Q1: additive model + global LR test
m_null <- glm(choice01 ~ 1, data = climateSupport, family = binomial(link = "logit"))
m_add  <- glm(choice01 ~ countries + sanctions, data = climateSupport, family = binomial(link = "logit"))

summary(m_add)
anova(m_null, m_add, test = "Chisq")
```

Summary output.

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)          -0.27266    0.05360   -5.087 3.64e-07 ***
countries160 of 192   0.64835    0.05388   12.033  < 2e-16 ***
countries80 of 192    0.33636    0.05380    6.252 4.05e-10 ***
sanctions15%         -0.13325    0.06208   -2.146  0.03183 *
sanctions20%         -0.30356    0.06209   -4.889 1.01e-06 ***
sanctions5%           0.19186    0.06216    3.086  0.00203 **

Null deviance: 11783  on 8499  degrees of freedom
Residual deviance: 11568  on 8494  degrees of freedom
AIC: 11580
```

Global LR test output.

```
Analysis of Deviance Table

Model 1: choice01 ~ 1
Model 2: choice01 ~ countries + sanctions
Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      8499      11783
2      8494      11568  5   215.15 < 2.2e-16 ***
```

Interpretation and conclusion. The number of participating countries is strongly and positively associated with support. Relative to 20/192, both 80/192 ($\hat{\beta}_{80} = 0.336$, $p < 0.001$)

and $160/192$ ($\hat{\beta}_{160} = 0.648$, $p < 0.001$) increase the log-odds of support. Sanctions also matter: compared to no sanctions, 5% sanctions increase support ($\hat{\gamma}_5 = 0.192$, $p = 0.002$), while 15% and 20% sanctions decrease support ($\hat{\gamma}_{15} = -0.133$, $p = 0.032$; $\hat{\gamma}_{20} = -0.304$, $p < 0.001$). The global LR test yields $\chi^2(5) = 215.15$ with $p < 2.2 \times 10^{-16}$, so I reject H_0 and conclude the explanatory variables significantly improve model fit.

Question 2

(a) For nearly all countries participating (160/192), how does increasing sanctions from 5% to 15% change the odds of support? (b) For few countries participating (20/192), how does increasing sanctions from 5% to 15% change the odds of support? (c) What is the estimated probability of support if there are 80/192 participating with no sanctions?

Key idea. In an additive model (no interaction), the effect of sanctions does not depend on the level of `countries`. Therefore, the 15% vs 5% odds ratio is the same for parts (a) and (b).

R code (key steps).

```
1    # Q2a/Q2b: OR(15% vs 5%) = exp(beta_15 - beta_5)
2    b <- coef(m_add)
3    OR_15_vs_5 <- exp(b["sanctions15%"] - b["sanctions5%"])
4    OR_15_vs_5
5
6    # Q2c: predicted probability at countries=80/192, sanctions=None
7    newdat <- data.frame(
8    countries = factor("80 of 192", levels = levels(climateSupport$countries))
,
9    sanctions = factor("None", levels = levels(climateSupport$sanctions))
10   )
11   p_hat <- predict(m_add, newdata = newdat, type = "response")
12   p_hat
13
```

(a) and (b) Odds ratio output.

$$OR(15\% \text{ vs } 5\%) = \exp(\gamma_{15} - \gamma_5) = 0.7225.$$

```
1    OR(15% vs 5%) = 0.7224531
2
```

Thus, holding everything else constant, increasing sanctions from 5% to 15% multiplies the odds of supporting the policy by 0.7225, i.e., the odds decrease by about $1 - 0.7225 = 0.2775$ (27.8%). Because the model is additive, the same odds ratio applies when `countries = 20 of 192`.

(c) Predicted probability output.

```
1    Predicted probability = 0.5159191
2
```

Therefore, the estimated probability of support under `countries = 80 of 192` and `sanctions = None` is approximately 0.516.

Question 3

Would the answers to 2(a) and 2(b) potentially change if we included an interaction term in this model? Why? Perform a test to see if including an interaction is appropriate.

Explanation. Yes, potentially. If we include an interaction term `countries × sanctions`, the effect of sanctions is allowed to vary across different levels of participating countries. In that case, the odds ratio for moving from 5% to 15% sanctions could differ when `countries = 160/192` versus `countries = 20/192`.

R code (key steps).

```
# Q3: interaction model + LR test
m_int <- glm(choice01 ~ countries * sanctions, data = climateSupport,
    family = binomial(link = "logit"))
anova(m_add, m_int, test = "Chisq")
```

LR test output (additive vs interaction).

```
Analysis of Deviance Table

Model 1: choice01 ~ countries + sanctions
Model 2: choice01 ~ countries * sanctions
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       8494      11568
2       8488      11562  6    6.2928    0.3912
```

Since $p = 0.3912 > 0.05$, we fail to reject the null hypothesis that the interaction terms are jointly zero. There is no strong evidence that the effect of sanctions differs by the number of participating countries, so the additive model is adequate.