

# Problem Set 3

Han Li

Due: March 24, 2024

## 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 March 24, 2024. No late assignments will be accepted.

## Question 1

We are interested in how governments' management of public resources impacts economic prosperity. Our data come from Alvarez, Cheibub, Limongi, and Przeworski (1996) and is labelled `gdpChange.csv` on GitHub. The dataset covers 135 countries observed between 1950 or the year of independence or the first year for which data on economic growth are available ("entry year"), and 1990 or the last year for which data on economic growth are available ("exit year"). The unit of analysis is a particular country during a particular year, for a total  $> 3,500$  observations.

- Response variable:
  - `GDPWdiff`: Difference in GDP between year  $t$  and  $t-1$ . Possible categories include: "positive", "negative", or "no change"
- Explanatory variables:
  - `REG`: 1=Democracy; 0=Non-Democracy
  - `OIL`: 1=if the average ratio of fuel exports to total exports in 1984-86 exceeded 50%; 0= otherwise

Please answer the following questions:

1. Construct and interpret an unordered multinomial logit with `GDPWdiff` as the output and "no change" as the reference category, including the estimated cutoff points and coefficients.

```

1 gdp_data <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/gdpChange.csv", stringsAsFactors = F)
2 #create depvar
3 gdp_data[gdp_data$GDPWdiff==0, "GDPWdiff_cat"] <- "no change"
4 gdp_data[gdp_data$GDPWdiff>0, "GDPWdiff_cat"] <- "positive"
5 gdp_data[gdp_data$GDPWdiff<0, "GDPWdiff_cat"] <- "negative"
6 gdp_data$GDPWdiffcat <- relevel(as.factor(gdp_data$GDPWdiff_cat), ref="no change")
7 #(1)
8 gdp_mult <- multinom(GDPWdiff_cat ~ REG + OIL, data=gdp_data)
9 summary(gdp_mult)

1 summary(gdp_mult)
2 Call:multinom(formula = GDPWdiff_cat ~ REG + OIL, data = gdp_data)
3 Coefficients: (Intercept) REG OIL
4 negative 3.805370 1.379282 4.783968
5 positive 4.533759 1.769007 4.576321
6 Std. Errors: (Intercept) REG OIL
7 negative 0.2706832 0.7686958 6.885366
8 positive 0.2692006 0.7670366 6.885097
9 Residual Deviance: 4678.77
10 AIC: 4690.77
11

```

Interpretation: the unordered multinomial logit model treat the change of the GDP in categories (positive, negative and no change) as two separate regression lines with no change as the base. Concretely, compared to no change, the log odds for a negative change in gdp from t to t-1 on average is about 3.8 when it's a non democratic country with less or equal than 0.5 share of fuel exports in 1984-86. On average keeping Oil constant, when the country is demoratic, the log odds for a negative change increases by 1.378. Keeping regime constant, when the country 's average ratio for feul exports exceeds 0.5 between 84-86 the log odds for the gdp change to be negative will increase 4.78. On the other hand, the log odds for a positive change for a country with both REG and OIL being 0 from no change is 4.53. Keeping OIL constant, shifting the regime to democracy will increase the log odds for positive GDP change to by 1.77, and when keeping regime constant, change the ratio for Oil variable to 1 from 0 will increase the log odds for the change for GDP from no change to positive by 4.57 on average.

2. Construct and interpret an ordered multinomial logit with `GDPWdiff` as the outcome variable, including the estimated cutoff points and coefficients.

```

1 #(2)
2 gdp_data$GDPWdiff_cat <- relevel(gdp_data$GDPWdiff_cat, ref="negative") #
  order ascending
3
4 gdp_ord <- polr(GDPWdiff_cat ~ REG + OIL, data=gdp_data)
5 summary(ordered_logit)

1 Call:polr(formula = GDPWdiff_cat ~ REG + OIL, data = gdp_data)
2 Coefficients:
3      Value Std. Error    t value
4 REG  0.3985    0.07518    5.300
5 OIL -0.1987    0.11572   -1.717
6 Intercepts:
7
8      Value      Std. Error t value
9 negative|no change  -0.7312    0.0476  -15.3597
10 no change|positive -0.7105    0.0475  -14.9554
11 Residual Deviance: 4687.689
12 AIC: 4695.689

```

To interpret the estimated coefficients, when REG change from 0 to 1 when holding OIL constant, the log odds for the GDP change to go up a category (i.e. from negative to no change, or from no change to positive) will increase by 0.3985; likewise when keeping REG constant and change OIL from 0 to 1 will decrease the log odds for GDP upward category change by 0.1987 on average. For the estimated cutoffs (intercept), when the REG and OIL are both 0, on average the log odds for the GDP category change from negative to no change is -0.7312m and the average log odds for the GDP category change for no change to positive is -0.7105.

## Question 2

Consider the data set `MexicoMuniData.csv`, which includes municipal-level information from Mexico. The outcome of interest is the number of times the winning PAN presidential candidate in 2006 (`PAN.visits.06`) visited a district leading up to the 2009 federal elections, which is a count. Our main predictor of interest is whether the district was highly contested, or whether it was not (the PAN or their opponents have electoral security) in the previous federal elections during 2000 (`competitive.district`), which is binary (1=close/swing district, 0="safe seat"). We also include `marginality.06` (a measure of poverty) and `PAN.governor.06` (a dummy for whether the state has a PAN-affiliated governor) as additional control variables.

- (a) Run a Poisson regression because the outcome is a count variable. Is there evidence that PAN presidential candidates visit swing districts more? Provide a test statistic and p-value.

```

1 mexico_elections <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/
  StatsII_Spring2024/main/datasets/MexicoMuniData.csv")
2 # poisson model

```

```

3 mpoisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
  PAN.governor.06, data = mexico_elections, family=poisson)
4 summary(mpoisson)
5 dispersiontest(mpoisson) #can't reject null as pvalue >0.05, no need for
  ZIP

```

```

1 Call:
2 glm(formula = PAN.visits.06 ~ competitive.district + marginality.06 +
  PAN.governor.06, family = poisson, data = mexico_elections)
3 Coefficients:              Estimate Std. Error z value Pr(>|z|)
4 (Intercept)        -3.81023      0.22209  -17.156  <2e-16 ***
5 competitive.district -0.08135      0.17069   -0.477   0.6336
6 marginality.06      -2.08014      0.11734  -17.728  <2e-16 ***
7 PAN.governor.06     -0.31158      0.16673   -1.869   0.0617 .
8
9 (Dispersion parameter for poisson family taken to be 1)
10 Null deviance: 1473.87  on 2406  degrees of freedom
11 Residual deviance: 991.25  on 2403  degrees of freedom
12 AIC: 1299.2 Number of Fisher Scoring iterations: 7
13

```

```

1 Overdispersion
2 testdata:  mpoisson
3 z = 1.0668, p-value = 0.143
4 alternative hypothesis: true dispersion is greater than 1
5 sample estimates: dispersion      2.09834
6

```

From the summary we can see the competitive.district p value is 0.6336, which is  $>0.05$  so we do not have enough evidence to reject the null hypothesis that there is no association between competitive.district and thenumber of visits from the winning PAN presidential candidate in 2006. We can also see from overdispersion test that the p-value is  $>0.05$ , so we have not found enough evidence to reject the null hypothesis that the true dispersion is smaller or equal to 1. Hence we can keep the poisson model without needing to run the zero inflated model.

- (b) From the above summary, we can interpret that for marginality.06, holding PAN governor and competitive district constant, a one-unit increase in poverty leads to a change in the expected counts of visits by a multiplicative factor of about 0.1249,  $\exp(-2.08014)$ , on average. For PAN governor, holding marginality.06 and competitive district constant, having a PAN governor changes the expected count of visits by a multiplicative factor of about 0.7323,  $\exp(-0.31158)$ , on average.
- (c) Provide the estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive (competitive.district=1), had an average poverty level (marginality.06 = 0), and a PAN governor (PAN.governor.06=1).

```

1 mexico_elections <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/
  StatsII_Spring2024/main/datasets/MexicoMuniData.csv")

```

```
2 # poisson model
3 mpoisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
  PAN.governor.06, data = mexico_elections, family=poisson)
4 summary(mpoisson)
```

The predicted value is 0.0149, so the expected counts of visits from the winning PAN candidate are 0.0149 for a competitive district with an average poverty level and a PAN governor.