

# Problem Set 3

## Applied Stats II

Due: March 26, 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 March 26, 2023. 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.

Table 1: Multinomial Regression

	<i>Dependent variable:</i>	
	Negative (1)	Positive (2)
REG	1.379* (0.769)	1.769** (0.767)
OIL	4.784 (6.885)	4.576 (6.885)
Constant	3.805*** (0.271)	4.534*** (0.269)
Akaike Inf. Crit.	4,690.770	4,690.770
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Coefficient Probabilities

	(Intercept)	REG	OIL
Negative	0.978	0.799	0.992
Positive	0.989	0.854	0.990

Table 3: p-values

	(Intercept)	REG	OIL
Negative	0	0.073	0.487
Positive	0	0.021	0.506

The cutoff for the log odds of no change to negative GDP difference is 3.80 (SE = 0.27), which is average effect on log odds of negative GDP difference of a country

being both a non-democracy and non- major oil exporter on the log og negative GDP difference. The cutoff for log odds for no change to positive GDP difference is 4.533 (SE = 2.6).

Holding oil exporter status constant and with no change as the reference, moving from non- to democracy increases the log odds of negative GDP difference by an average of 3.80 (SE = 0.76). The average effect of being a democracy on the log odds of positive GDP difference is 4.533 (SE = 76). Holding regime type constant, moving from non- to major oil exporter increases the log of negative GDP difference by an average of 4.78 (SE = 6.88). The average of effect of being a major oil exporter on the log odds of positive GDP difference is 4.57 (SE = 6.88) However, at the 95% confidence level, only for the association between regime type and positive GDP difference can we reject the null hypothesis that this relationship is random ( $z = 2.03$ ,  $p = 0.02$ ).

Converting to probabilities, when holding a countries oil exporter status constant, democracies have 79.8 % greater chance of experiencing negative economic growth compared to no change over non-democracies, and an 85.5 % chance they experiencing positive economic growth. Democracies are thus 5.7% more likely than non-democracies to experience positive economic growth compared to no change, than they are negative economic growth compared to no change. When holding regime type constant, major oil exporters are 99.1% more likely than non-oil exporters to experience negative growth compared to no change, and 98.9% more likely to experience positive growth. Thus, when controlling for regime type, major oil exporters are 0.2 % more likely to experience negative growth over positive growth in relation to no change.'

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

Table 4: Ordered Logistic Regression

	Value	Std. Error	t value	p value	p value.1
REG	0.398	0.075	5.300	0.00000	0.00000
OIL	-0.199	0.116	-1.717	0.086	0.086
Negative No Change	-0.731	0.048	-15.360	0	0
No Change Positive	-0.710	0.048	-14.955	0	0

odds of negative to no change in GDP difference when a country is both a non-major oil exporter and a non-democracy and is -.7312 (SE = 0.04). The average effect of a country being a non-major oil exporter and a non-democracy on the log logs odds no change to positive is 0.71 ( SE = 0.04)

Holding oil export status constant, moving from democracy to non-democracy increases the log of GDP difference average of 0.39 (SE = 0.04) from negative to positive. Holding regime type constant, moving from non- to major oil exporter reduces the log of ordered

GDP difference -0.19. However, in ordered logistic, neither of these relationships are significant, meaning we cannot reject the null hypothesis that either association is random.

Converting to probabilities, holding whether oil exporter status constant, democracies are 59% more likely than non-democracies to have positive GDP difference. Holding regime type constant, countries which are major oil exporters are 45% less like to have positive GDP growth than non-major oil exporters.

## 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. (see pg 5 for table)

When controlling for the other information the model, a district going from non-competitive to competitive reduces log count of districts visited by an average -0.08, suggesting competition is associated with a reduction in candidate visits. A p value  $\geq 0.05$  means that cannot reject the null hypothesis that this relationship is by chance ( $z = 0.477$ ,  $p = 0.63$ )

- (b) Interpret the `marginality.06` and `PAN.governor.06` coefficients.

When controlling for the other information the model, each unit increase in the poverty level reduces the log of the count of PAN presidential visits by -2.08 ( $z = -17.72$ ,  $p \leq 0.001$ ). Having a PAN governor reduces the log of the presidential visit count by -0.31. However, the association is not significant ( $z = 1.89$ ,  $p \geq 0.05$ ), this we cannot dismiss the null hypothesis that the association is by chance.

- (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`).

A hypothetical competitive district with average poverty and a PAN governor would get an average of 0 visits from the winning candidate (see pg 6 for code).

```
1 (exp(coeff[1] + coeff[2] + coeff[4]))
```

Table 5: Poisson Regression for Presidential Candidate Visits

	<i>Dependent variable:</i>
	PAN.visits.06
competitive.district	−0.081 (0.171)
marginality.06	−2.080*** (0.117)
PAN.governor.06	−0.312* (0.167)
Constant	−3.810*** (0.222)
Observations	2,407
Log Likelihood	−645.606
Akaike Inf. Crit.	1,299.213
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01