

# Problem Set 3

## Applied Stats II

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 GD PWdiff as the output and "no change" as the reference category, including the estimated cutoff points and coefficients.

```

1 #creating factor variables
2 data[data$GDPWdiff == 0, "GDPfactor"] <- "no change"
3 data[data$GDPWdiff > 0, "GDPfactor"] <- "positive"
4 data[data$GDPWdiff < 0, "GDPfactor"] <- "negative"
5
6
7 # re-leveling data
8 data$GDPfactor<- relevel(as.factor(data$GDPfactor), ref = "no change")
9
10 # running the unordered regression
11
12 unor_mul_regr <- multinom(GDPfactor ~ REG + OIL, data = data)
13
14 summary(unor_mul_regr)
]

```

Table 1: Unordered 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	

To interpret this multinomial regression (unordered) we have to recognise that each regression is going from "no change" to "negative" and "no change" to "positive".

- First, we can start with the intercept of 'no change' to 'negatively'- When we hold all covariates (REG and OIL) constant at 0, we can assume that the log odds of 'negative' on average is 3.805
- For 'no change' to 'negative' for the coefficient REG, if we are holding all other covariates constant, we could assume, for every one unit change from non-democracy to democracy there would be an increase in the log odds of GDP of 'no change' to 'negative' of 1.379, on average.
- If we then look at the OIL coefficient, while holding all other covariates constant, we can assume for every one unit change in the exportation of oil we could expect to see an increase in the log odds of GDP change by 4.784 on average of 'no change' to 'negative'

Next, looking at 'no change' to 'positive'.

- First, we can start with the intercept of 'no change' to 'positive'- When we hold all covariates (REG and OIL) constant at 0, we can assume that the log odds of 'positive' on average is 4.534
- For 'no change' to 'positive' for the coefficient REG, if we are holding all other covariates constant, we could assume, for every one unit change from non-democracy to democracy there would be an increase in the log odds of GDP of 'no change' to 'positive' of 1.7969, on average.
- If we then look at the OIL coefficient, while holding all other covariates constant, we can assume for every one unit change in the exportation of oil we could expect to see an increase in the log odds of GDP change by 4.576 on average of 'no change' to 'positive'

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

```

1 # reordering
2
3 data$GDPfactor <- relevel(data$GDPfactor, ref = "negative")
4
5 # run the ordered regression
6
7 order_regression <- polr(GDPfactor ~ REG + OIL, data = data)
8 summary(order_regression)
9 # Intercepts negative —> no change aprox -0.7312...Intercepts no change
   —> positive aprox -0.7105

```

]

Table 2: Ordered Multinomial Regression

	<i>Dependent variable:</i>
	GDPfactor
REG	0.398*** (0.075)
OIL	-0.199* (0.116)
Observations	3,721
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Once we have reordered and rerun the logit regression, we can interpret again. Order as in..negative —¿ no change —¿ positive

- Looking at the REG coefficient, while holding all other covariates constant, with one unit change in REG from non-democracy to democracy we can see an increase of log odds in GDP of 0.398 on average
- The coefficient OIL can be interpreted as holding all other covariates constant, with one unit change in oil exporation we could see a decrease in log odds by 0.199 on average
- We can then look at both intercepts, one from 'negative' to ' no change' and 'no change' to ' positive' . The intercept for 'negative' to ' no change' is -0.7312, this can be interpreted as ,when we hold all covariates constant at 0, the point at which negative switches to no change is at this point as well. While then the 'no change' to ' positive' intercept is -0.7105, so when we hold all covariates constant at 0 the point at where the change happens is at -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
2 poisson <- glm(PAN.visits.06 ~
3               competitive.district + marginality.06 + PAN.governor.06,
4               data = mexico_elections, family = poisson)
5
6 summary(poisson)
]

```

Table 3: Mexican Poisson Regression

	<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

If we are to look at our competitive.district coefficients we can see that our z-value is -0.477 which is not a larger enough to reject our null hypotheses that PAN presidential candidates do not visit swing districts more. We can also see from the p-value 0.6336 is larger than 0.05, also showing that our test statistic is not large enough.

Table 4: Coefficients and Model Summary

Variable	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.81023	0.22209	-17.156	<2e-16 ***
competitive.district	-0.08135	0.17069	-0.477	0.6336
marginality.06	-2.08014	0.11734	-17.728	<2e-16 ***
PAN.governor.06	-0.31158	0.16673	-1.869	0.0617 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for poisson family taken to be 1)				
Null deviance	1473.87	on 2406 degrees of freedom		
Residual deviance	991.25	on 2403 degrees of freedom		
AIC	1299.2			

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

- For the coefficient `marginality.06` we can interpret that while holding all other covariates constant, a one unit change from not-marginal to marginal, the logs change in expected visits to decrease by 2.080 which also can translate to, the expected number of visits of aprox 0.125 when we that -2.080 and put it to e, (e-2.080).
- For the `PAN.governor.06` coefficients, while holding all other covariates constants, a one unit change in `PAN.governor.06` or not `PAN.governor` to `PAN.governor`, the logs counts expects decreases by 0.312. Also could translate to the aprox 0.732 number of visits if there was to be a PAN governor compared to not having a PANa governor.

(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 exp(coef(poisson)[1] + coef(poisson)[2] * 1 + coef(poisson)[3] * 0 + coef
  (poisson)[4] * 1)#from slides
2 # 0.01494818
3 # OR
4 prediction <- data.frame(competitive.district = 1,
5                           marginality.06 = 0,
6                           PAN.governor.06 = 1)
7 predict(poisson, newdata = prediction, type = "response")
8 # 0.0194818

```

] We could predict that an estimated number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive, an average poverty level, as well as a PAN governor, would be arpx 0.014