

# Problem Set 3

Tianxin Zhang 21352232 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.

```

1 data <- read.csv("C:/Users/Caesar/Documents/GitHub/StatsII_Spring2023/
  datasets/gdpChange.csv", header = TRUE)
2
3
4
5 # set the level for outcome variable
6
7 data$GDPWdiff_reo <- ifelse(data$GDPWdiff == 0, "no change", ifelse(data$
  GDPWdiff < 0, "negative", "positive"))
8
9 data$GDPWdiff_reo <- as.factor(data$GDPWdiff_reo)
10
11 data$GDPWdiff_reo <- relevel(data$GDPWdiff_reo, ref = "no change")
12
13
14 # run unordered multinomial regression
15
16 mult.log <- multinom(GDPWdiff_reo ~ REG + OIL, data = data)
17 summary(mult.log)
18 stargazer(mult.log, title="mult.log Results")

```

Table 1: mult.log Results

	<i>Dependent variable:</i>	
	negative	positive
	(1)	(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	

Interpretation:

Negative vs. no change model:

Coefficient for REG: Holding other covariates constant, for every one unit change in regime type, from non-democracy to democracy, we expect to see an increase in the log-odds of GDP change negatively vs. no change in GDP by 1.379 on average.

Coefficient for OIL: Holding other covariates constant, for every one unit change in oil exportation, from "the average ratio of fuel exports to total exports in 1984-86 did not exceeded 50%" to "the average ratio of fuel exports to total exports in 1984-86 exceeded 50%", we expect to see an increase in the log-odds of GDP change negatively vs. no change in GDP by 4.784 on average.

Intercept: Holding all covariates constant as 0, the log odds of GDP change negatively vs. no change in GDP is 3.805 on average.

Positive vs. no change model:

Coefficient for REG: Holding other covariates constant, for every one unit change in regime type, from non-democracy to democracy, we expect to see an increase in the log-odds of GDP change positively vs. no change in GDP by 1.769 on average.

Coefficient for OIL: Holding other covariates constant, for every one unit change in oil exportation, from "the average ratio of fuel exports to total exports in 1984-86 did not exceeded 50%" to "the average ratio of fuel exports to total exports in 1984-86 exceeded 50%", we expect to see an increase in the log-odds of GDP change positively vs. no change in GDP by 4.576 on average.

Intercept: Holding all covariates constant as 0, the log odds of GDP change positively vs. no change in GDP is 4.534 on average.

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

```
1 # run ordered multinomial regression
2 data$GDPWdiff_reo <- factor(data$GDPWdiff_reo , levels = c("negative", "
  no change", "positive"))
3 ord.log <- polr(GDPWdiff_reo ~ REG + OIL, data = data, Hess = TRUE)
4 summary(ord.log)
5 stargazer(ord.log, title="ord.log Results")
```

Interpretation:

Coefficient for Regime type: Holding other covariates constant, one unit change in the regime type, from the category of non-democracy to democracy, we expect to see an

Table 2: ord.log Results

<i>Dependent variable:</i>	
GDPWdiff_reo	
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	

increase in the log odds of change in GDP in the positive direction by 0.398 on average.

Coefficient for OIL: Holding other covariates constant, for every one unit change in oil exportation, from "the average ratio of fuel exports to total exports in 1984-86 did not exceeded 50%" to "the average ratio of fuel exports to total exports in 1984-86 exceeded 50%", we expect to see an decrease in the log-odds of GDP change in a positive direction by 0.199 on average.

Intercept (negative—no change -0.7312): Holding all covariates constant as 0, the cut-off points of the log odds between negative change in GDP and no change in GDP in this model is -0.7312.

Intercept (no change—positive -0.7105): Holding all covariates constant as 0, the cutoff points of the log odds between no change in GDP and positive change in GDP in this model 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.

Table 3: mod.ps Results

	<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

```

1 # a)
2
3 data2 <- read.csv("C:/Users/Caesar/Documents/GitHub/StatsII_Spring2023/
4 datasets/MexicoMuniData.csv", header = TRUE)
5
6 mod.ps <- glm(PAN.visits.06 ~ competitive.district + marginality.06 + PAN
7 .governor.06, data = data2, family = "poisson")
8 result <- summary(mod.ps)
9 stargazer(mod.ps, title="mod.ps Results")
10
11 cfs <- coef(mod.ps)
12 cfs[2] # coefficient for competitive.district
13
14 se_cd <- result$coefficients[2, 2]
15 se_cd # standard error of competitive.district

```

```

15
16 ## I will conduct a Z test to test individual coefficient for the
    variable Competitive.district
17 ## I set 95% as our confidence interval
18 ## H0: B1 = 0
19 ## Ha: B0 not equals to 0
20 ## Z test score:
21 z = (cfs[2]-0)/se_cd
22 z # -0.4766106
23
24 # two-tail , p-value
25 p = (1-pnorm(abs(z)))*2
26 p
27 # p value: 0.6336394, which is larger than 5%. Therefore, we cannot reject
    the null hypothesis that
28 # the coefficient for the competitive.district equals to 0.

```

Therefore, there is no sufficient evidence that PAN presidential candidate will visit swing districts more.

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

1. Interpretation of the coefficient for `marginality.06`: Holding all other covariates constant, for a one unit change in `marginality.06`, from the category of not-marginal to the category of marginal, the logs of expected counts of visit is expected to decrease by 2.080.

2. Interpretation of the coefficient for `PAN.governor.06`: Holding all other covariates constant, for a one unit change in `PAN.governor.06`, from the category of not PAN.governor to the category of PAN.governor, the logs of expected counts of visit is expected to decrease by 0.312.

(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 # c)
2
3
4 pred <- data.frame(competitive.district = 1,
5                     marginality.06 = 0,
6                     PAN.governor.06 = 1)
7 predict(mod.ps, newdata= pred, type = "response")
8
9 # 0.01494818
10
11

```

```
12
13 exp(cfs[1] + cfs[2] * 1 + cfs[3] * 0 + cfs[4] * 1)
14
15 ## 0.01494818
16
17 ## In conclusion, the estimated mean number of visit is 0.
```