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 forwhich 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.

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

```
1  ######################
2  # 1
3  ######################
4  #Categorising GDPWdiff
5  gdp_data$GDPWdiff <- ifelse(gdp_data$GDPWdiff>0,"positive",
6                              ifelse(gdp_data$GDPWdiff<0,"negative","no
      change"))
7  gdp_data$GDPWdiff <- as.factor(gdp_data$GDPWdiff)
8  gdp_data$GDPWdiff <- relevel(gdp_data$GDPWdiff, ref = "no change")
9
10 # Running model
11 mul_mod <- multinom(GDPWdiff~REG+OIL, data=gdp_data)
12 summary(mul_mod)
13 stargazer(mul_mod)
```

Table 1:

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

According to the results,we can write the functions:

$$ln(\frac{P_{negative}}{P_{nochange}}) = 3.805 + 1.379REG + 4.784OIL \tag{1}$$

$$ln(\frac{P_{positive}}{P_{nochange}}) = 4.534 + 1.769REG + 4.576OIL \tag{2}$$

(1)
Intercept (3.805): This represents the log-odds of a country experiencing a negative change in GDP compared to no change when both REG and OIL are zero.

Coefficient of REG(1.379): For every one-unit increase in the REG variable, the log-odds of a country experiencing a negative change in GDP compared to no change increase by 1.379, holding all other variables constant. However, the p-value of coefficient is during 0.05 to 0.1.

Coefficient of OIL have no significance.

(2)
Intercept (4.534): This represents the log-odds of a country experiencing a positive change in GDP compared to no change when both REG and OIL are zero.

Coefficient of REG (1.769): For every one-unit increase in the REG variable, the log-odds of a country experiencing a positive change in GDP compared to no change increase by 1.769, holding all other variables constant.

Coefficient of OIL have no significance.

```
1  #######################
2  # 2
3  #######################
4  # Since I set the reference category in the previous question,
5  # but normally "negative" would be the default reference, I have adjusted
      it.
6  gdp_data$GDPWdiff <- relevel(gdp_data$GDPWdiff,ref = "negative","no
      change","positive")
7  # Running model
8  model <- polr(GDPWdiff ~ REG + OIL, data = gdp_data)
9  summary(model)
10 stargazer(model)
```

According to the results,we can write the functions:

$$ln(\frac{P_{negative}}{P_{nochange}}) = -0.7312 + 0.398REG - 0.199OIL \tag{3}$$

$$ln(\frac{P_{positive}}{P_{nochange}}) = -0.7105 + 0.398REG - 0.199OIL \tag{4}$$

Table 2:

	Dependent variable:
	GDPWdiff
REG	0.398***
	(0.075)
OIL	−0.199*
	(0.116)
Observations	3,721
Note:	*p<0.1; **p<0.05; ***p<0.01

(3)
Intercept (-0.7312): This represents the log-odds of a country experiencing
a negative change in GDP compared to no change when both REG and OIL are zero.

Coefficient of REG (0.398): For every one-unit increase in the REG variable, the log-odds of
a country experiencing a negative change in GDP compared to no change increase by 0.398,
holding all other variables constant.

Coefficient of OIL(-0.199): For every one-unit increase in the OIL variable,the log-odds of a
country experiencing a negative change in GDP compared to no change decrease by 0.199,
holding all other variables constant. However, the p-value of this coefficient is during from
0.1 to 0.05.

(4)
Intercept (-0.7105): This represents the log-odds of a country experiencing a positive change
in GDP compared to no change when both REG and OIL are zero.

Coefficient of REG (0.398): For every one-unit increase in the REG variable, the log-odds of
a country experiencing a positive change in GDP compared to no change increase by 0.398,
holding all other variables constant.

Coefficient of OIL (-0.199): For every one-unit increase in the OIL variable, the log-odds of
a country experiencing a positive change in GDP compared to no change decrease by 0.199,
holding all other variables constant.However, the p-value of this coefficient is during from
0.1 to 0.05, which not very significant.

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  #a
3  #######################
4  # wrangle
5  mexico_elections <- within (mexico_elections , {
6  PAN. governor.06<- as.logical (PAN. governor.06)
7  competitive.district<- as.logical (competitive.district)
8  })
9
10 # EDA
11 str (mexico_elections)
12 summary (mexico_elections)
13
14 with (mexico_elections ,
15      list (mean(PAN. visits.06), var(PAN. visits.06)))
16
17 # Running model
18 mod.ps <- glm(PAN. visits.06~ PAN. governor.06+competitive.district+
        marginality.06,
19              data = mexico_elections , family = poisson)
20 summary (mod.ps)
21 stargazer (mod.ps)
```

Due to the p-value of coefficient of competitive.district don not have enough distance to zero, so it means no significance.
There is no evidence that PAN presidential candidates visit swing districts more.

Table 3:

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

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

Coefficient of PAN.governor.06(-0.312): For every one-unit increase in the PAN.governor.06 variable, the log-odds of PAN visit decrease by 0.312, holding all other variables constant.
Due to p-value of coefficient of PAN.governor.06 don not have enough distance to zero,so the PAN.governor.06 is a not strongly reliable and negative predictor of PAN visit.

Coefficient of marginality.06(-2.080): For every one-unit increase in the marginality.06 variable, the log-odds of PAN visit decrease by 2.080, holding all other variables constant

(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`). coefficients.

Accororading to the table 3:

$$ln(\lambda) = -3.810 - 0.312 PAN.governor.06 - 0.081 competitive.district - 2.080 marginality.06 \tag{5}$$

$$\lambda = exp(-3.810 - 0.312 * 1 - 0.081 * 1 - 2.080 * 0) = exp(-4.203) \tag{6}$$