

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 `GDPWdiff` as the output and "no change" as the reference category, including the estimated cut-off points and coefficients.

First I create cut-off points in the data at -0.000001 and 0 to determine negative, no change and positive effect intervals.

```
1 breaks <- c(-Inf, -0.000001, 0, Inf)
2
3 # Define labels for the intervals
4 labels <- c("negative", "no change", "positive")
```

I create a new column to store these results. I then re-level the data in order to use the no change category as the reference.

```
1 gdp_data$GDPWdiff_cat <- cut(gdp_data$GDPWdiff, breaks = breaks, labels =
  labels, include.lowest = TRUE)
1 gdp_data$GDPWdiff_cat <- relevel(gdp_data$GDPWdiff_cat, ref = "no change")
```

Finally, I run a multinomial logit model.

```
1 multinom_model1 <- multinom(GDPWdiff_cat ~ REG + OIL, data = gdp_data)
```

This produces the following results:

Table 1:		
	<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
Note:	*p<0.1; **p<0.05; ***p<0.01	

The coefficient for the constant for the negative category shows that when all independent variables are zero, the log odds of being in the negative category compared to the no change category are approximately 3.805.

The coefficient for the constant for the positive category shows that when all independent variables are zero, the log odds of being in the positive category compared to the no change category are approximately 4.534.

The coefficient for regime for the negative category shows that, when all else is held constant, a move from non-democracy to democracy is associated with an estimated increase of 1.379 in the log-odds of being in the negative category in comparison to the reference category.

The coefficient for regime for the positive category shows that, when all else is held constant, a move from non-democracy to democracy is associated with an estimated increase of 1.769 in the log-odds of being in the positive category in comparison to the reference category.

The coefficient for fuel exports for the negative category shows that, when all else is held constant, a change from 0 to 1 is associated with an estimated increase of 4.784 in the log-odds of being in the negative category in comparison to the reference category.

The coefficient for fuel exports for the positive category shows that, when all else is held constant, a change from 0 to 1 is associated with an estimated increase of 4.576 in the log-odds of being in the positive category in comparison to the reference category.

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

Table 2:

<i>Dependent variable:</i>	
GDPWdiff_cat	
REG	0.410*** (0.075)
OIL	-0.179 (0.115)
Observations	3,721

Note: *p<0.1; **p<0.05; ***p<0.01

The estimated cutoff points are the same here: -0.000001 and 0.

I ran an ordered multinomial logit:

```
1 polr_model1 <- polr(GDPWdiff_cat ~ REG + OIL, data = gdp_data, Hess= TRUE
  )
```

The coefficient for regime shows that, holding all else equal, for a move from non-democracy to democracy the log odds of moving to a higher category are estimated to increase by 0.410.

The coefficient for fuel exports shows that, holding all else equal, for a move from 0 to 1 the log odds of moving to a higher category are estimated to decrease by 0.179.

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.

I ran a Poisson regression:

```
1 poisson_model1 <- glm(PAN.visits.06 ~ competitive.district + marginality
  .06 + PAN.governor.06, data = mexico_elections, family = poisson)
```

The coefficient for `competitive.district` is -0.081, suggesting that PAN candidates are less likely to visit swing districts more. However, the P-value is 0.6336. It is greater than the typical significance level of 0.05, which means we fail to reject the null hypothesis. There is not enough evidence to conclude that PAN candidates visit swing districts more.

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

The coefficient for `marginality` is -2.080, meaning that an increase in `marginality` by one unit decreases the expected number of visits by a multiplicative factor of $e^{-2.080} = 0.125$

The coefficient for PAN Governor is -0.312, meaning that a move from no PAN governor to PAN governor decreases the expected number of visits by a multiplicative factor of $e^{-0.312} = 0.732$

Table 3:

	<i>Dependent variable:</i>
	PAN.visits.06
competitive.district	−0.081 (0.6336)
marginality.06	−2.080*** ($< 2 \times 10^{-16}$)
PAN.governor.06	−0.312* (0.617)
Constant	−3.810*** ($< 2 \times 10^{-16}$)
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

Variable	Test Statistic
competitive.district	-0.477
marginality.06	-17.728
PAN.governor.06	-1.869
Constant	-17.156

Table 4: Test statistic for Regression Model

- (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`).

I ran this calculation in R:

```
1 lambda <- exp(-3.81023+(-0.08135)*1+(-2.08014)*0+(-0.31158)*1)
```

This returned a result of 0.015 visits. The estimated mean number of visits in this district is 0.015