Applied Stats II: Problem Set 3

Luna Goldstein

Due: March 28, 2022

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 class on Monday March 28, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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.

   R code:

```
data$GDPWdiff <- relevel(data$GDPWdiff, ref= "no change")
mult.log <- multinom(GDPWdiff ~ REG + OIL, data = data)
```

Table 1: The coefficients

	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	

   Interpreting our coefficients:

   REG: The countries with a negative GDP difference, moving from a non-democracy to democracy,increases GDP growth by a log odds of 1.379282.

   REG: The countries with a positive GDP difference, moving from non-democracy to democracy, increases GDP growth by a log odds of 1.769007.

   OIL: The countries with a negative GDP difference, moving from less than 5o% to exceeding the 5o% ratio of fuel exports, increases GDP growth by a log odds of 1.379282.

   OIL: The countries with a positive GDP difference, moving from less than 5o% to exceeding the 5o% ratio of fuel exports, increases GDP growth by a log odds of 1.769007.

Not all the estimated coefficients are significant.

Table 2: The P-values

	REG	OIL
negative	0.073	0.487
positive	0.021	0.506

From our first and second table it is clear that the intercept's p-value is significant and so is the relationship between the explanatory variable (REG) and countries with a positive GDP difference. However, the remaining variables have p-values which are not significant.

Table 3: Confidence interval

	REG	OIL
negative	0.073	0.487
positive	0.021	0.506

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

   R code:

```r
ord.log <- polr(GDPWdiff ~ REG + OIL, data = data, Hess = TRUE)
```

Table 4:

	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

Interpreting our coefficients:

Moving from a non-democracy to a democracy increases the GDP growth by a log offs of 0.398.

Moving from less than 5o% to exceeding the 5o% ratio of fuel exports, decreases GDP growth by a log odds of 0.199.

It is important to note that not all the estimated coefficients are significant. Only the explanatory variable REG has a significant p-value.

Table 5: Confidence interval

	2.5%	97.5%
REG	0.2516	0.5464
OIL	−0.4237	0.03019

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.

1. 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.

   R code:

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

Table 5:

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

We do not reject the null hypothesis that states that the PAN presidential candidates visit swing districts more.

log count: -0.08135

p-value: 0.6336

test statistic: -0.477

2. Interpret the `marginality.06` and `PAN.governor.06` coefficients.

   If poverty in the district were to increase by one unit, the PAN government visitations would be expected to decrease by a log count 2.08, while holding all other variables in the model constant.

   If the district went from not having a PAN affiliated governor to having one, the PAN government visitations would be expected to decrease by a log count 0.31, while holding all other variables in the model constant.

3. 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).

   R code:

```
visits <- exp(coeff[1] + coeff[2]*1 + coeff[3]*0 + coeff[4]*1)
```

Estimated mean number of visits from the winning PAN presidential candidate = 0.01494818