Problem Set 3

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 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.

   In the model below, the output shows the logit regression coefficients relative to the reference category. Under 'REG', the 1.012 suggests that, holding everything else constant, for a one unit change (i.e. 0 to 1) in democracy/not, the logit coefficient for 'negative' relative to 'no change' (the reference category) will go up by that amount, 1.012. In other words, if the country becomes a democracy, the chances of their GDP going down from year t-1 to t are higher compared to not changing or going up. This is also true when holding all else constant for the logit coefficient for 'positive' relative to 'no change', and the effect is bigger and significant at a smaller alpha level. Holding all else constant there doesn't appear to be a statistically significant effect for a change in the ratio of oil in exports, although the logit regression coefficients seem sizeable.

Table 1: Unordered Multinomial Logit Model (Model 1)

	Dependent variable:	
	negative	positive
	(1)	(2)
REG	1.012*	1.405**
	(0.550)	(0.548)
OIL	0.575	0.362
	(0.745)	(0.742)
Constant	3.488***	4.212***
	(0.231)	(0.229)
Akaike Inf. Crit.	4,780.264	4,780.264
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Relative Risk Ratios Model 1

	Dependent variable:	
	negative	positive
	(1)	(2)
REG	2.752*	4.075**
	(0.550)	(0.548)
OIL	1.776	1.436
	(0.745)	(0.742)
Constant	32.717***	67.503***
	(0.231)	(0.229)
Akaike Inf. Crit.	4,780.264	4,780.264
Note:	*p<0.1; **p<0.05; ***p<0.01	

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

Ans 2:

```
    > summary(ord.log)
Call:
polr(formula = GDPWdiff ~ REG + OIL, data = gdpChange, Hess = TRUE)

Coefficients:
      Value Std. Error t value
REG  0.4172     0.07499   5.563
OIL -0.1953     0.11505  -1.698

Intercepts:
                    Value     Std. Error t value
noChange|negative   -4.8708    0.2028    -24.0142
negative|positive   -0.6912    0.0474    -14.5690

Residual Deviance: 4771.979
AIC: 4779.979

```

The output shows that for countries who are democratic, holding all else constant, the log odds of having no change in their GDP (versus negative or positive change) is actually 0.4172 points lower than countries who are not democratic. So the formulation for the categories of no change and for negative change, soley by democracy/non-democracy, become $logit(P(Y_iNoC)) = -4.8708 - 0.4172x1_i$ and $logit(P(Y_iNeg)) = -0.6912 - 0.4172x1_i$ respectively. Holding all else constant, for countries who exceeded

the 50 percent ratio of oil to total exports, the log odds of having no change in their GDP (versus negative or positive change) is actually 0.1953 points higher than countries who are not democratic. Adding this to the formulations for the categories of no change and for negative change, we get $logit(P(Y_i NoC)) = -4.8708 - 0.4172x1_i + 0.1953x2_i$ and $logit(P(Y_i Neg)) = -0.6912 - 0.4172x1_i + 0.1953x2_i$ respectively.

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: Poisson Model

	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

On average, all else held constant, there is no statistically significant effect on the log odds of a PAN candidate visiting a swing district, however we do observe a small, non-significant change (-0.081), suggesting, counterintuitively, that PAN candidates may

visit competitive districts less than 'safe seats'.

Table 4: Robust Estimates and CIs

	Estimate	Robust SE	Pr(>|z|)	LowLim	UppLim
(Intercept)	-3.810	0.285	0	-4.369	-3.251
competitive.district	-0.081	0.260	0.755	-0.592	0.429
marginality.06	-2.080	0.187	0	-2.447	-1.713
PAN.governor.06	-0.312	0.255	0.221	-0.811	0.188

Robust estimates for the p-value and confidence intervals suggest that actually, the status of having a PAN governor, all else held constant, on average does not have a statistically significant effect on the log odds of visits from PAN candidates pre-election.

Table 5: Chi-Square Difference of Means

Statistic	N	Mean	St. Dev.	Min	Max
Resid. Df	2	2,404.500	2.121	2,403	2,406
Resid. Dev	2	1,232.562	341.263	991.253	1,473.872
Df	1	3.000		3	3
Deviance	1	482.619		482.619	482.619
Pr(>Chi)	1	0.000		0	0

After completing an ANOVA for goodness of fit of the poisson model as compared to our null model, our Chi-square test statistic is 482.619, and its p-value is minimal ($< 2.2e-16 * **$ in ANOVA output).

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

Table 6: Poisson Model Reproduction

	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

On average, and holding all other variables constant, a one unit-change in marginality for a district (i.e. an increase in poverty score), is associated with a 2.080 decrease in the log odds of receiving a visit from the PAN presidential candidate. This decrease is significant at the alpha level of 0.01. Somewhat suprising, this result could indicate potentially a preference for the PAN candidate to visit wealthy districts.

As the above robust estimates for the confidence intervals showed, the PAN.governor.06 coefficient is not statistically significant with this higher criteria. However, to interpret as though it was statistically significant: if all else was held constant at their means or baseline categories (i.e. a safe seat district with average marginality), on average the impact of having a PAN governor on the log odds of receiving visits from the PAN presidential candidate is a decrease of 0.312. Potentially this could indicate a complacency among PAN presidential candidates towards districts that have already elected PAN governors.

(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).

$y_i = -3.81023 - 0.08135(X1[competitive]) - 2.08014(X2[marginality]) - 0.31158(X3[Pangov])$
$= -3.81023 - 0.08135 - 0.31158$
$= -4.20316$
$e^{-4.20316} = 0.01494826558884$

Estimated mean visits is 0.015.