

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: "", "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.

First, I load and wrangle data a little to create my categorical variable, making sure to set "No change" as the reference.

```
1 # Data
2 gdp <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/gdpChange.csv", stringsAsFactors = F)
3
4 # Creating three-level categorical outcome:
5 gdp <- gdp %>%
6   mutate(GDPWdiff_cat = ifelse(GDPWdiff < 0, "Decrease",
7                                 ifelse(GDPWdiff == 0, "No_change",
8                                         "Increase")),
9   GDPWdiff_cat = factor(GDPWdiff_cat, levels = c("No_change", "Increase", "Decrease")))
```

Then I run the unordered model and extract and report the results, also converting to odds ratios:

```
1 ## Unordered version:
2 unordered_logit <- multinom(GDPWdiff_cat ~ REG + OIL, data = gdp, ref = "No_change")
3 summary(unordered_logit)
4
5 # Turning to OR:
6 odds_ratios <- exp(coef(unordered_logit))
7 print(round(odds_ratios, 2))
8
9 # Reporting:
10 stargazer(unordered_logit, type = "latex")
11
12 odds_ratios_df <- as.data.frame(round(odds_ratios, 2))
13 odds_ratios_df$Outcome <- rownames(odds_ratios_df)
14 stargazer(odds_ratios_df, type = "latex", title = "Odds Ratios from Model 1",
15           summary = FALSE,
16           rownames = FALSE)
```

This yields the following results:

Table 1:		
	<i>Dependent variable:</i>	
	Increase	Decrease
	(1)	(2)
REG	1.800** (0.770)	1.400* (0.770)
OIL	4.600 (6.900)	4.800 (6.900)
Constant	4.500*** (0.270)	3.800*** (0.270)
Akaike Inf. Crit.	4,691.000	4,691.000
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Odds Ratios from Model 1			
(Intercept)	REG	OIL	Outcome
93.110	5.870	97.160	Increase
44.940	3.970	119.580	Decrease

Interpreting the results from the first table:

- A transition from a non-democracy to a democracy is linked, on average, with a 3.805 increase in the log odds of transitioning from no change in GDP to a decrease in GDP, while controlling for fuel exports. This relationship is statistically significant.
- A transition from a non-democracy to a democracy is associated, on average, with a 4.534 increase in the log odds of transitioning from no change in GDP to an increase in GDP, while controlling for fuel exports. This relationship is statistically significant.
- A shift from less than 50% to more than 50% of total fuel exports is, on average, linked with a 4.784 increase in the log odds of transitioning from no change in

GDP to a decrease change in GDP, while controlling for regime type. However, this relationship is not statistically significant.

- A shift from less than 50% to more than 50% of total fuel exports is, on average, associated with a 4.576 increase in the log odds of transitioning from no change in GDP to an increase change in GDP, while controlling for regime type. However, this relationship is not statistically significant.
- For non-democracies with an average fuel export of less than 50%, the log odds of experiencing a decrease in GDP is 3.805. This relationship is statistically significant.
- For non-democracies with an average fuel export of less than 50%, the log odds of experiencing an increase in GDP is 4.534. This relationship is statistically significant.

And interpreting the results from the conversions to ORs:

- Transitioning from a non-democracy to a democracy increases the odds of experiencing a decrease in GDP by approximately 297.2%, while controlling for fuel exports.
- Transitioning from a non-democracy to a democracy increases the odds of experiencing an increase in GDP by approximately 486.5%, while controlling for fuel exports.
- Shifting from less than 50% to more than 50% of total fuel exports increases the odds of experiencing a decrease in GDP by approximately 11857.8%, while controlling for regime type.
- Shifting from less than 50% to more than 50% of total fuel exports increases the odds of experiencing an increase in GDP by approximately 9615.6%, while controlling for regime type.
- The estimated baseline odds of having a decrease in GDP compared to no change is approximately 44.942 times larger.
- The estimated baseline odds of having an increase in GDP compared to no change is approximately 93.108 times larger.

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

I re-level my variable for the ordered model:

```
1 ## Ordered version:
2 gdp <- gdp %>%
3   mutate(GDPWdiff_cat = factor(GDPWdiff_cat, levels = c
4                                 ("Decrease", "No_change", "Increase"),
5                                 ordered = TRUE))
```

Then I run the model and extract the results, also converting to odds ratios:

```
1 # Running
2 ordered_logit <- polr(GDPWdiff_cat ~ REG + OIL, data = gdp, Hess=TRUE)
3 summary(ordered_logit)
4
5 # Turning to OR:
6 odds_ratios <- exp(coef(ordered_logit))
7 print(round(odds_ratios, 2))
8
9 # Reporting:
10 stargazer(ordered_logit, type = "latex")
11 odds_ratios_df <- as.data.frame(round(odds_ratios, 2))
12 odds_ratios_df$Outcome <- rownames(odds_ratios_df)
13 stargazer(odds_ratios_df, type = "latex", title = "Odds Ratios from Model
14           1",
15           summary = FALSE,
16           rownames = FALSE)
```

This yields the following:

Table 3:	
	<i>Dependent variable:</i>
	GDPWdiff_cat
REG	0.398*** (0.075)
OIL	-0.199* (0.116)
Intercepts:	
Decrease—No_change	-0.7312 (0.0476)
No_change—Increase	-0.7105 (0.0475)
Observations	3,721
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4: Odds Ratios from Model 2

Odds Ratio	Outcome
1.490	REG
0.820	OIL

From Table 3, we see that:

- For democracies, as opposed to non-democracies, the log odds of having a higher GDP chance is 0.398, while controlling for fuel exports.
- For countries with a fuel exports ratio larger than 50%, as opposed with those with ratios smaller than 50%, the log odds of having a higher GDP chance is -0.199, while controlling for regime type.
- Finally, the cutoff points are estimated to be -0.7312 for transitioning from a decrease in GDP to no change, and -0.7105 for transitioning from no change to an increase in GDP. These points are thresholds that set the boundaries between the different categories of our outcome variable, and they tell us that the transition between categories on the ordinal dimension is relatively small, and that the

change from one category to the next is not very substantial, so it wouldn't take much movement along the latent variable to shift from one category to another.

From Table 4, we see that:

- For democracies, as opposed to non-democracies, the likelihood of having a larger GDP change is 1.49 times higher, while controlling for fuel exports.
- For countries whose ratio of fuel exports is larger than 50%, the likelihood of having a larger GDP is 0.82 times that of countries with fuel exports ratios smaller than 50%, while controlling for regime type.

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.

Reading the data and running the regression:

```
1 # Data
2 mex <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/MexicoMuniData.csv")
3
4 poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
5               PAN.governor.06, data = mex, family = "poisson")
6 summary(poisson)
7
8 # Reporting
9 stargazer(poisson, type = "latex")
```


This yields the following results:

Table 5:	
	<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

From Table 5, we see that the estimated coefficient for competitive district is -0.081. Additionally, our test statistic (z score in this case) is -0.477, with an associated p-value of 0.6336. Thus, the estimated coefficient tells us that a competitive district, compared to districts where the PAN is safe, is associated with an average change in the log counts of PAN candidate visits of -0.081. The sign of this coefficient is the opposite of what we would expect (we would expect more visits in competitive as opposed to safe districts), but the coefficient is quite small and not statistically significant. The associated p-value of 0.6336 means we cannot reject the null hypothesis that the coefficient for competitive district is 0, and we have found no evidence that PAN presidential candidates visit swing districts more.

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

- The `marginality` coefficient of -2.08 means that a 1-unit increase in `marginality` is, on average, associated with a 2.08 decrease in the log counts of the PAN presidential candidate visits. So, visits would seem to decrease as poverty increases.
- The `PAN.governor` coefficient of -0.312 means that having a PAN governor, as opposed to having a governor from another party, is associated with, on average, a 0.312 decrease in the log counts of the PAN presidential candidate visits. This means that there are less expected visits for states where the governor is from PAN.

(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`).

Calculating the number of visits:

```
1 # To get the visits:
2 log_count_est <- coef(poisson)[1] +
3   coef(poisson)["competitive.district"]*1 +
4   coef(poisson)["marginality.06"]*0 +
5   coef(poisson)["PAN.governor.06"]*1
6
7 # Converting log to count:
8 expected_visits <- exp(log_count_est)
9 print(expected_visits)
```

The estimated average number of visits is **0.015**.