Problem Set 3

Kaley Burg

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.

   - First I loaded the data, subset it, and then created two factor variables, one for the unordered model and one for the ordered model:

```
# load data
gdp_data <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/gdpChange.csv", stringsAsFactors = F)


#subsetting
gdp_data_sub <- gdp_data[,c("GDPWdiff", 'REG', "OIL")]


#creating an ordered factor
gdp_data_sub$GDPWdiff_cut <- ifelse(gdp_data_sub$GDPWdiff < 0, "Negative",
                                    ifelse(gdp_data_sub$GDPWdiff == 0, "NoChange", "Positive"))


gdp_data_sub$GDPWdiff_ord <- factor(gdp_data_sub$GDPWdiff_cut,
                                    levels = c("Negative", "NoChange", "Positive"),
                                    ordered = T)


#creating an unordered factor
gdp_data_sub$GDPWdiff_unord <- factor(gdp_data_sub$GDPWdiff_cut,
                                    levels = c("Negative", "NoChange", "Positive"),
                                    ordered = F)
```

   - Then I changed my reference category for the unordered model to "no change"

```
gdp_data_sub$GDPWdiff_unord <- relevel(gdp_data_sub$GDPWdiff_unord, ref = "NoChange")
```

   - Then I ran the unordered multinomial logit, results shown in table 1:

```
unord_reg <- multinom(GDPWdiff_unord ~ REG + OIL, data = gdp_data_sub)
```

   - Next I converted the log odds to the odds ratios, shown in table 2:

```
unord_OR <- exp(coef(unord_sum))
```

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	

Table 2:

	(Intercept)	REG	OIL
Negative	44.942	3.972	119.578
Positive	93.108	5.865	97.156

- The **log-odds interpretation** for these results is as follows:
  - **Negative Intercept:** For Non-Democracies with an average fuel export of less than 50%, the log odds of having a negative GDP difference between year t and t-1 is **3.805**. This is a statistically significant relationship.
    - This uses a cutoff point of < 0 GDP to define this intercept
  - **Positive Intercept:** For Non-Democracies with an average fuel export of less than 50%, the log odds of having a positive GDP difference between year t and t-1 is **4.534**. This is a statistically significant relationship.
    - This uses a cutoff point of > 0 GDP to define this intercept
  - **Negative OIL:** A change from less than 50% of total fuel exports to more than 50% of total fuel exports is associated, on average, with a **4.784** increase in the log odds of changing from a result of no change in GDP to a negative change in GDP, holding democracy constant. This is not a statistically significant relationship.

- **Positive OIL:** A change from less than 50% of total fuel exports to more than 50% of total fuel exports is associated, on average, with a **4.576** increase in the log odds of changing from a result of no change in GDP to a positive change in GDP, holding democracy constant. This is not a statistically significant relationship.
- **Negative REG:** A change from a non-democracy to a democracy is associated, on average, with a **3.805** increase in the log odds of changing from a result of no change in GDP to a negative change in GDP, holding fuel exports. This is a statistically significant relationship.
- **Positive REG:** A change from a non-democracy to a democracy is associated, on average, with a **4.534** increase in the log odds of changing from a result of no change in GDP to a positive change in GDP, holding fuel exports. This is a statistically significant relationship.

- The **odds ratio** interpretation for these results is as follows:
  - **Negative Intercept:** The baseline odds ratio can be represented as $e^{\hat{\beta}_0} = e^{3.805} = 44.942$ meaning that 44.942 is the estimated baseline odds of having a negative GDP change compared to having no change.
  - 44.942 indicates that the odds of a country having a negative change in GDP is approximately 44.942 times higher than the odds of having no change.
  - **Positive Intercept:** The baseline odds ratio can be represented as $e^{\hat{\beta}_0} = e^{4.534} = 93.108$ meaning that 93.108 is the estimated baseline odds of having a positive GDP change compared to having no change.
  - 93.108 indicates that the odds of a country having a positive change in GDP is approximately 93.108 times higher than the odds of having no change.
  - **Negative OIL:** A change from less than 50% of total fuel exports to more than 50% of total fuel exports increases the odds of having a negative GDP change by a multiplicative factor of 119.578; it increases the odds by $\approx$ 11857.8%, controlling for regime type.
  - **Positive OIL:** A change from less than 50% of total fuel exports to more than 50% of total fuel exports increases the odds of having a positive GDP change by a multiplicative factor of 97.156; it increases the odds by $\approx$ 9615.6%, controlling for regime type.
  - **Negative REG:** A change a non-democracy to a democracy increases the odds of of having a negative GDP change by a multiplicative factor of 3.972; it increases the odds by $\approx$ 297.2%, controlling for fuel exports.
  - **Positive REG:** A change a non-democracy to a democracy increases the odds of of having a positive GDP change by a multiplicative factor of 5.865; it increases the odds by $\approx$ 486.5%, controlling for fuel exports.

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

   - First I ran the ordered multinomial logit, results shown in table 3:

```
ord_log <- polr(GDPWdiff_ord ~ REG + OIL, data = gdp_data_sub, Hess =
    TRUE)
```

   - Then I converted the results into odd ratios, shown in table 4:

Table 3: Ordinal Logistic Regression Model

	Dependent variable:
	GDPWdiff_ord
REG	0.398***
	(0.075)
OIL	−0.199*
	(0.116)
Observations	3,721
Intercepts:	
Negative—NoChange	-0.7312***
	(0.0476)
NoChange—Positive	-0.7105***
	(0.0475)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4:

	OR	2.5 %	97.5 %
REG	1.490	1.286	1.727
OIL	0.820	0.655	1.031

   - The **log-odds interpretation** for these results is as follows:
     - For countries that are a democracy, compared to non-democracies, the log odds of having a higher GDP change is 0.398, holding fuel exports constant

  - For countries that have an average ratio of fuel exports higher than 50%, compared to those that don't, the log odds of having a higher GDP change is -0.199, holding regime type constant
  - The **cutoff points/intercepts are:** -0.7312 for transitioning from negative change in GDP to no change in GDP, and -0.7105 for transitioning from no change in GDP to positive change.
    * The cutoff points are the cut points or transitions along newly created ordinal latent variable.
    * What this tells us is that the points of transition along the x axis is a small change between those categories on the ordinal dimension; in other words, it wouldn't take very long to go from one category to the next. These units are largely uninterpretable though

- The **odds ratio interpretation is:**
  - For a country that is a democracy, the odds of having a higher GDP change is 1.490 higher compared to a country that is a non-democracy, holding constant fuel exports.
  - For a country that has an average ratio of fuel exports higher than 50%, the odds of having a higher GDP change is 0.820 compared to a country with fuel exports lower than 50%, holding regime type constant.

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.

- First I loaded the dataset and ran the Poisson regression, results shown in table 5:

```
mexico_elections <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsII_Spring2024/main/datasets/MexicoMuniData.csv")


```

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

- Then I converted the results from the log-odds to the odds ratio, results shown in table 6:

```r
poisson_OR <- exp(cbind(OR = coef(poisson_mod), confint(poisson_mod)))
```

Table 5: Poisson Regression Model

	Coefficient	Standard Error	Z-score	p-value
competitive.district	-0.081	0.171	-0.477	0.6336
marginality.06	-2.080^{***}	0.117	-17.728	$< 2e-16$
PAN.governor.06	-0.312^{*}	0.167	-1.869	0.0617
(Constant)	-3.810^{***}	0.222	-17.156	$< 2e-16$
Observations	2,407			
Log Likelihood	-645.606			
Akaike Inf. Crit.	1,299.213			
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 6:

	OR	2.5 %	97.5 %
(Intercept)	0.022	0.014	0.034
competitive.district	0.922	0.666	1.302
marginality.06	0.125	0.099	0.156
PAN.governor.06	0.732	0.523	1.007

- **There is not evidence that PAN presidential candidates visit swing districts more**. The β coefficient is -0.081, indicating that a close/swing district, compared to a safe seat, is associated with, on average, a -0.081 change in log counts of PAN presidential candidate visits. The test statistic (z-score) is **-0.477** and the p-value is **0.6336**, which is therefore not significant.
- The full interpretation is: there is a decrease in the log counts by 0.081 going from a non competitive to competitive district (safe to swing state), this result is not statistically significant. Furthermore, we cannot reject the null hypothesis stating that the coefficient for competitive district is 0. In other words, we have no evidence for the claim that PAN presidential candidates visit competitive close/swing districts more than districts labeled as safe.

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

- **marginality.06 interpretation**: For every 1 unit increase in marginality, there is, on average, a 2.08 decrease in the log counts of the PAN presidential candidate visiting a a district
- This is associated with a multiplicative change of 0.125 in the expected number of visits
- **PAN.governor.06 interpretation:** when a state has a PAN-affiliated governor, compared to states without pan affiliated governors, there is an, on average, 0.312 decrease in the log counts of the PAN presidential candidate visiting a district
- This is associated with a multiplicative change of 0.732 in the expected number of visits

(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 calculated this with the code from the week 9 slides:

```
coeffs <- coefficients(poisson_mod)

estimated_mean <- exp(coeffs[1] + coeffs[2]*1 + coeffs[3]*0 + coeffs
    [4]*1)
```

- The estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district that was competitive, had an average poverty level and a PAN governor was **0.01494818**