

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 cutoff points and coefficients.

```
1
2 # Create a categorical variable for GDP difference
3 gdp_data$GDPWdiff_Cat <- cut(gdp_data$GDPWdiff,
4                             breaks = c(-Inf, -0.01, 0.01, Inf),
5                             labels = c("negative", "no change", "positive"
6                                     ))
7 # the multinomial logit model
8 multinom_model <- multinom(GDPWdiff_Cat ~ REG + OIL, data = gdp_data, ref
9                             = "no change")
10 # Display the summary of the model
11 summary(multinom_model)
```

Coefficients:

	(Intercept)	REG	OIL
no change	-3.8011902	-1.351703	-7.9240683
positive	0.7284081	0.389905	-0.2076511

Std. Errors:

	(Intercept)	REG	OIL
no change	0.27014596	0.75825317	32.9772055
positive	0.04789662	0.07552484	0.1158094

Residual Deviance: 4678.728

AIC: 4690.728

The intercepts reflect the cutoff point at which the probability(log odd) of being in one category shifts to another when all predictors are at their reference levels. Because the no change category is the reference, there are no intercepts or coefficients estimated for moving into no change from no change. The intercept or cutoff point for positive is 0.7284081: the log-odds of a positive change in GDP from no change when all predictor variables (REG and OIL) are 0. The REG positive coefficient is 0.389905: when holding OIL constant, moving from a non-democracy (REG = 0) to a democracy (REG = 1) along with an increase in the log-odds of a positive GDP change from no change by 0.39. The OIL positive coefficient is -0.2076511, when holding REG constant, countries with a high ratio of fuel exports(OIL = 1) have a 0.21 log-odds decrease when a "positive" GDP changes from "no change".

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

```
1 ordered_model <- polr(GDPWdiff_Cat ~ REG + OIL, data = gdp_data, Hess=
  TRUE)
2
3 summary(ordered_model)
```

Coefficients:

	Value	Std. Error	t value
REG	0.3985	0.07518	5.300
OIL	-0.1987	0.11572	-1.717

Intercepts:

	Value	Std. Error	t value
negative no change	-0.7312	0.0476	-15.3597
no change positive	-0.7105	0.0475	-14.9554

Residual Deviance: 4687.689

AIC: 4695.689

The intercept(the cutoff point) for the change from negative to no change is -0.7312, indicating the baseline log-odds of being in no change from negative when all predictors are 0.

The intercept(the cutoff point) from no change to positive is -0.7105, the baseline log-odds of being positive from no change when all predictors are 0.

The coefficient for REG 0.3985, when holding all else constant, an increase in the REG variable (moving from a non-democracy to a democracy) is associated with an increase in the log-odds of being in a higher category of GDP difference by 0.3985

The coefficient for OIL -0.1987, which stands for an increase in OIL (increasing fuel exports) is associated with a decrease in the log-odds of being in a higher category of GDP difference by 0.1987, when holding all else 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.

```
1 # Constructing the Poisson regression model
2 poisson_model <- glm(PAN.visits.06 ~ competitive.district + marginality
   .06 + PAN.governor.06,
3                       family = poisson(link = "log"),
4                       data = mexico_elections)
5
6 summary(poisson_model)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.81023	0.22209	-17.156	<2e-16 ***
competitive.district	-0.08135	0.17069	-0.477	0.6336
marginality.06	-2.08014	0.11734	-17.728	<2e-16 ***
PAN.governor.06	-0.31158	0.16673	-1.869	0.0617 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1473.87 on 2406 degrees of freedom
 Residual deviance: 991.25 on 2403 degrees of freedom
 AIC: 1299.2

There is no evidence to suggest that PAN presidential candidates visited swing districts more than other districts. The coefficient for competitive. the district is not statistically significant, with test statistic -0.477, with a p-value of 0.6336 above the 0.05 threshold, when controlling for poverty levels and whether a state had a PAN-affiliated governor, being a competitive district did not significantly affect the number of visits by PAN candidates.

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

marginality.06 coefficient is -2.08014 shows a negative relationship between the marginality (poverty level) of a district and the log count of visits. This means that in higher levels of poverty with fewer visits from the PAN presidential candidate, a one-unit increase in marginality leads to decrease the expected count of visits by $e^{-2.08014}$ when holding other variables constant.

PAN.governor.06 coefficient -0.31158 suggests a negative relationship between having a PAN-affiliated governor and the log count of visits, with a PAN governor associated with fewer visits, a one-unit increase in marginality leads to a decrease in the expected count of visits by $e^{-0.31158}$ when holding other variables constant.

- (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`).

```
1 # the estimate log count
2 log_count <- predict(poisson_model, newdata = data.frame(competitive.
   district=1, marginality.06=0, PAN.governor.06=1), type = "link")
3
4 #Convert to the original count
5 count_predicted <- exp(log_count)
6
7 print(count_predicted)
8 #0.01494818
9 #The estimated mean number of visits was calculated by putting the
   provided values into the regression equation and then using the
   exponential function to convert the log count for the number of visits
   in a Poisson regression model back to number of visit
```