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.

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

```
#create factors
data_gdp$OIL <- factor(data_gdp$OIL,
levels = c(0,1),
labels = c("Exports < 50%", "Exports > 50%"))
tail(data_gdp$OIL)

data_gdp$GDPWdiff_cat <- cut(data_gdp$GDPWdiff, breaks = c(-Inf, -1, 1,
  Inf),
labels = c("Negative", "No Change", "Positive"))

levels(data_gdp$GDPWdiff_cat)

#relevel
data_gdp$GDPWdiff_cat <- relevel(data_gdp$GDPWdiff_cat, ref = "No
  Change")

#(a)
multi_model <- multinom(GDPWdiff_cat ~ REG + OIL, data = data_gdp)
summary(multi_model)
```

Interpretation: A one unit increase in oil exports (oil exports ¿ 50 percent GDP) is associated with an increase of 0.57 in the log odds of negative GDP growth vs no change in a given country-year. A one unit increase in oil exports is associated with an increase of 0.36 in the log odds of positive growth vs no change in a given country-year The quality of being a democracy (REG = 1) is associated with a 1.01 increase in the log odds of negative growth vs no change in a given country-year. The quality of being a democracy (REG = 1) is associated with a 1.4 increase in the log odds of positive growth vs no change in a given country-year.

```
ord_multi_model <- polr(GDPWdiff_cat ~ REG + OIL, data = data_gdp, Hess
  = TRUE)
summary(ord_multi_model)
exp(coef(ord_multi_model))
```

Interpretation: The quality of being a democracy is associated with an increase of 1.5 in the odds of GDP increasing or decreasing in a given year. A one unit increase in oil export (exports ¿ 50 percent GDP) is associated with 0.82 increase in the odds of a country's GDP increasing or decreasing in a given year.

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.

```
pan_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
PAN.governor.06, data=data_mex, family = poisson)
summary(pan_poisson)
errors <- summary(pan_poisson)$coefficients[, 2]
cfs <- coef(pan_poisson)
pan_poisson_z <- cfs/errors
p <- (1 - pnorm(abs(pan_poisson_z), 0, 1))*2
p
```

No, there is no evidence that PAN candidates visit swing districts more. The estimate for swing districts 0.921 visits on average, but this result is not statistically different from 0.

(b) Interpret the `marginality.06` and `PAN.governor.06` coefficients. Marginality.06 : On average, a one-unit increase in the measure of poverty is associated with a -2.08 decrease in the log odds of Calderon visiting to the district in the run up to the 2009 elections, holding all other variables constant

PAN.governor.06 : On average, the effect of a given district having a sitting PAN governor is associated with a -0.31 decrease in the log odds of Calderon visiting the district in the run up to the 2009 elections, holding all other variables constant. Although these effects are not statistically different from 0.

(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).

```
cfs <- coef(pan_poisson)
exp(cfs[1]*1 + cfs[2]*0 + cfs[3]*1)*100
```

Estimated that Calderon would visit this hypothetical district 0.276 times in the run up to the 2009 election.