

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 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.

```
gdpChange$GDPWdiffMOD <- cut(gdpChange$GDPWdiff,
                             breaks = c(-100000, -1, 1, 10000),
                             labels = c("positive",
                                         "negative",
                                         "no change"))

model <- multinom(gdpChange$GDPWdiffMOD ~ ., data = data)
summary(model) # model runs for around 20 seconds then an error says that there are
too many weights and im not sure how to fix this. I tried googling it and I couldnt
find an answer.
```

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

```
model2 <- polr(gdpChange$GDPWdiffMOD ~ ., data = data, method = "logistic")
summary(model2) #I have tried running this
a few times and R has crashed each time, im not sure if its an issue with my
laptop or if i've done something wrong but it will not run a summary.
It runs, the program freezes and then crashes.
```

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.

```
mod.ps <- glm(data2$competitive.district ~ ., data = data2, family = poisson)
summary(mod.ps)
```

```
test_statistic <- (778.15 - 542.72)/ 2401
test_statistic # [1] 0.09805498
```

```
p_value <- pchisq(test_statistic * 2401,5, lower.tail = FALSE)
p_value # [1] 7.330925e-49
```

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

```
# Marginality 06 - coefficient is -0.05708,
# this means that for a one unit increase on marginality 06 there is
# an expected decrease in the response variable of 0.05708.
# PAN.governor.06 - coefficient is 0.03499,
# this means that for a one unit increase on marginality 06 there is
# an expected increase in the response variable of 0.03499.
```

- (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`).

```
input_data <- data.frame(
  competitive.district = 1,
  marginality.06 = 0,
  PAN.governor.06 = 1,
  pan.vote.09 = mean(data$pan.vote.09),
  PAN.visits.06 = mean(data$PAN.visits.06),
  MunicipCode = mean(data$MunicipCode)
)
```

```
mean_visits <- predict(mod.ps, newdata = input_data, type = "response")
mean_visits
```

I could not get this code to work, I was receiving NA as the response which is obviously not correct. I included the code because I am not sure where I have gone wrong here