Problem Set 3

Applied Stats II

Due: March 28, 2022

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 class on Monday March 28, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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.

QUESTION 1 - MY ANSWER

```

# libraries

pkgTest <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}

lapply(c("tidyverse",
         "stargazer",
         "nnet",
         "ggplot2",
         "MASS"), pkgTest)

setwd("/Users/mark/Documents/ASDS-applied-stats-2-2022/problem_set3")
changeData <- read_csv("./gdpChange.csv")

summary(changeData)
names(changeData)
head(changeData, n=10)
tail(changeData, n=5)

####################
### Question 1 ###
####################

# Question 1 - Part 1. Construct and interpret an unordered multinominal
    logit with GDPWdiff
# as the output and 'no change' as the reference category, including the
    estimated cutoff
# points and coefficients

# keep only the following:
# GDPWdiff which is the Response variable
# REG, and OIL which are the explanatory variables

newchange_data <- changeData[,c("GDPWdiff", "REG", "OIL")]
```

```r
newchange_data

# categories "positive", "negative", "no change"
# no change == 0
# negative < 0
# positive > 0

newchange_data <- within(newchange_data, {
  GDPWdiff1 <- NA
  GDPWdiff1[GDPWdiff == 0] <- "no change"
  GDPWdiff1[GDPWdiff < 0] <- "negative"
  GDPWdiff1[GDPWdiff > 0] <- "positive"
})

newchange_data

newchange_data$GDPWdiff1 <- factor(newchange_data$GDPWdiff1,
                                   levels = c("no change", "positive", "
    negative"))

summary(newchange_data$GDPWdiff1)
# no change  positive  negative
#    16         2600      1105

# run the base multinominal logit
multi_logit <- multinom(GDPWdiff1 ~ REG + OIL, data = newchange_data)
summary(multi_logit)

# the exponentiate coefficients
coef_data <- exp(coef(multi_logit))
coef_data

#             (Intercept)   REG         OIL
# positive    93.10789    5.865024    97.15632
# negative    44.94186    3.972047    119.57794


# the estimated cutoff points
confint_data <- exp(confint(multi_logit))
confint_data

# this gives

# positive

#                 2.5 %       97.5 %
#  (Intercept)  5.493416e+01 1.578085e+02
# REG           1.304269e+00 2.637379e+01
# OIL           1.339263e-04 7.048166e+07

# negative
```

```r

#                   2.5 %       97.5 %
#   (Intercept) 2.643900e+01 7.639360e+01
# REG           8.804391e-01 1.791965e+01
# OIL           1.647467e-04 8.679315e+07

# interpreting the coefficients and cutoff points

# 5.865024 increase suggests a positive growth in the GDP for Democracy (REG: 1=Democracy)?




# Question 1 - Part 2

newchange_data2 <- newchange_data
newchange_data2 <- changeData[,c("GDPWdiff", "REG", "OIL")]
newchange_data2

# categories "positive", "negative", "no change"
# no change == 0
# negative < 0
# positive > 0

summary(newchange_data2)

newchange_data2$GDPWdiff2 <- newchange_data$GDPWdiff


newchange_data2 <- within(newchange_data2, {
  GDPWdiff2 <- NA
  GDPWdiff2[GDPWdiff == 0] <- "no change"
  GDPWdiff2[GDPWdiff < 0] <- "negative"
  GDPWdiff2[GDPWdiff > 0] <- "positive"
})

newchange_data2


# check ordering

is.ordered(newchange_data2$GDPWdiff2)
# FALSE

# ordering

as.ordered(newchange_data2$GDPWdiff2)

# gives Levels: no change < positive < negative

```

```r
# create an ordered multinominal logit


multi_logit2 <- multinom(GDPWdiff2 ~ REG + OIL, data = newchange_data2)
summary(multi_logit2)

multi_logit2

# the exponentiate coefficients
coef_data2 <- exp(coef(multi_logit2))
coef_data2

# this gives

#             (Intercept)       REG          OIL
# no change  0.02234416     0.2587991     0.0003619269
# positive   2.07177984     1.4768404     0.8124904479


# the estimated cutoff points
confint_data2 <- exp(confint(multi_logit2))
confint_data2

# no change

#                 2.5 %         97.5 %
# (Intercept) 1.315877e-02  3.794135e-02
# REG         5.855130e-02  1.143903e+00
# OIL         3.078737e-32  4.254701e+24

# positive

#                 2.5 %       97.5 %
# (Intercept) 1.8861400   2.275691
# REG         1.2736401   1.712460
# OIL         0.6475021   1.019519



# interpreting the coefficients and cutoff points

# the ordered multinominal logit suggest an 1.4768404 increase in the GDP
    for Democracy


```

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.

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

(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).

QUESTION 2 - MY ANSWER

```
1
2  the data
3
4 mex_data <- read.csv("/Users/mark/Documents/ASDS-applied-stats-2-2022/
      problem_set3/MexicoMuniData.csv")
5 str(mex_data)
6 view(mex_data)
7 names(mex_data)
8
9 as.factor(mex_data$PAN.governor.06)
10 as.factor(mex_data$competitive.district)
11 # as.factor(mex_data$cPAN.visits.06)
12
13 # (a) run a POISSON REGRESSION
14
15 mex_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06
      + PAN.governor.06,
16                 data = mex_data, family = poisson(link = log))
17
18 mex_poisson
19
20 # this gives
21
```

```
# Coefficients:
# (Intercept)  competitive.district        marginality.06       PAN.
    governor.06
#     -3.81023           -0.08135              -2.08014
    -0.31158

mex_coeffs <- coefficients(mex_poisson)
mex_coeffs

# this gives

# (Intercept) competitive.district       marginality.06      PAN.governor
    .06
# -3.81023498          -0.08135181          -2.08014361
    -0.31157887
```