

# 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 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.
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

```
1
2 # libraries
3
4 pkgTest <- function(pkg){
5   new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
6   if (length(new.pkg))
7     install.packages(new.pkg, dependencies = TRUE)
8   sapply(pkg, require, character.only = TRUE)
9 }
10
11 lapply(c("tidyverse",
12         "stargazer",
13         "nnet",
14         "ggplot2",
15         "MASS"), pkgTest)
16
17 setwd("/Users/mark/Documents/ASDS-applied-stats-2-2022/problem_set3")
18 changeData <- read_csv("./gdpChange.csv")
19
20 summary(changeData)
21 names(changeData)
22 head(changeData, n=10)
23 tail(changeData, n=5)
24
25 #####
26 ### Question 1 ###
27 #####
28
29 # Question 1 – Part 1. Construct and interpret an unordered multinomial
    logit with GDPWdiff
30 # as the output and 'no change' as the reference category, including the
    estimated cutoff
31 # points and coefficients
32
33 # keep only the following:
34 # GDPWdiff which is the Response variable
35 # REG, and OIL which are the explanatory variables
36
37 newchange_data <- changeData[,c("GDPWdiff", "REG", "OIL")]
```

```

38 newchange_data
39
40 # categories "positive", "negative", "no change"
41 # no change == 0
42 # negative < 0
43 # positive > 0
44
45 newchange_data <- within(newchange_data, {
46   GDPWdiff1 <- NA
47   GDPWdiff1[GDPWdiff == 0] <- "no change"
48   GDPWdiff1[GDPWdiff < 0] <- "negative"
49   GDPWdiff1[GDPWdiff > 0] <- "positive"
50 })
51
52 newchange_data
53
54 newchange_data$GDPWdiff1 <- factor(newchange_data$GDPWdiff1,
55                                   levels = c("no change", "positive", "
56                                   negative"))
57
58 summary(newchange_data$GDPWdiff1)
59 # no change    positive    negative
60 #      16          2600        1105
61
62 # run the base multinomial logit
63 multi_logit <- multinom(GDPWdiff1 ~ REG + OIL, data = newchange_data)
64
65 summary(multi_logit)
66
67 # the exponentiate coefficients
68
69 coef_data <- exp(coef(multi_logit))
70
71 coef_data
72
73
74 #
75 #           (Intercept)      REG      OIL
76 # positive    93.10789    5.865024    97.15632
77 # negative    44.94186    3.972047   119.57794
78
79
80 # the estimated cutoff points
81
82 confint_data <- exp(confint(multi_logit))
83
84 confint_data
85
86
87 # this gives
88
89 # positive
90
91
92 #
93 #           2.5 %           97.5 %
94 # (Intercept) 5.493416e+01 1.578085e+02
95 # REG         1.304269e+00 2.637379e+01
96 # OIL         1.339263e-04 7.048166e+07
97
98 # negative

```

```

88
89 #           2.5 %           97.5 %
90 # (Intercept) 2.643900e+01 7.639360e+01
91 # REG         8.804391e-01 1.791965e+01
92 # OIL         1.647467e-04 8.679315e+07
93
94 # interpreting the coefficients and cutoff points
95
96 # 5.865024 increase suggests a positive growth in the GDP for Democracy (
    REG: 1=Democracy)?
97
98
99
100
101 # Question 1 – Part 2
102
103 newchange_data2 <- newchange_data
104 newchange_data2 <- changeData[,c("GDPWdiff", "REG", "OIL")]
105 newchange_data2
106
107 # categories "positive", "negative", "no change"
108 # no change == 0
109 # negative < 0
110 # positive > 0
111
112 summary(newchange_data2)
113
114 newchange_data2$GDPWdiff2 <- newchange_data$GDPWdiff
115
116
117 newchange_data2 <- within(newchange_data2, {
118   GDPWdiff2 <- NA
119   GDPWdiff2[GDPWdiff == 0] <- "no change"
120   GDPWdiff2[GDPWdiff < 0] <- "negative"
121   GDPWdiff2[GDPWdiff > 0] <- "positive"
122 })
123
124 newchange_data2
125
126
127 # check ordering
128
129 is.ordered(newchange_data2$GDPWdiff2)
130 # FALSE
131
132 # ordering
133
134 as.ordered(newchange_data2$GDPWdiff2)
135
136 # gives Levels: no change < positive < negative
137

```

```

138 # create an ordered multinomial logit
139
140
141 multi_logit2 <- multinom(GDPWdiff2 ~ REG + OIL, data = newchange_data2)
142 summary(multi_logit2)
143
144 multi_logit2
145
146 # the exponentiate coefficients
147 coef_data2 <- exp(coef(multi_logit2))
148 coef_data2
149
150 # this gives
151
152 #           (Intercept)           REG           OIL
153 # no change  0.02234416      0.2587991    0.0003619269
154 # positive   2.07177984      1.4768404    0.8124904479
155
156
157 # the estimated cutoff points
158 confint_data2 <- exp(confint(multi_logit2))
159 confint_data2
160
161 # no change
162
163 #           2.5 %           97.5 %
164 # (Intercept) 1.315877e-02  3.794135e-02
165 # REG         5.855130e-02  1.143903e+00
166 # OIL         3.078737e-32  4.254701e+24
167
168 # positive
169
170 #           2.5 %           97.5 %
171 # (Intercept) 1.8861400    2.275691
172 # REG         1.2736401    1.712460
173 # OIL         0.6475021    1.019519
174
175
176
177 # interpreting the coefficients and cutoff points
178
179 # the ordered multinomial logit suggest an 1.4768404 increase in the GDP
    for Democracy
180
181
182

```

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

```

22 # Coefficients:
23 # (Intercept) competitive.district      marginality.06      PAN.
      governor.06
24 #      -3.81023      -0.08135      -2.08014
      -0.31158
25
26 mex_coefs <- coefficients(mex_poisson)
27 mex_coefs
28
29 # this gives
30
31 # (Intercept) competitive.district      marginality.06      PAN.governor
      .06
32 # -3.81023498      -0.08135181      -2.08014361
      -0.31157887
33
34
35
36
37
38

```