Problem Set 3

Melissa Campbell - 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.

```
 lapply(c("MASS",
+            "nnet",
+            "ggplot2",
+           "car"),  pkgTest)
Loading required package: car
Loading required package: carData
 [[1]]
MASS
TRUE

 [[2]]
nnet
TRUE

 [[3]]
ggplot2
   TRUE

 [[4]]
 car
TRUE

> setwd("/Users/user/Documents/GitHub/StatsII_Spring2023/problemSets/PS03")
> data <- read.csv("gdpChange.csv")
> summary(data)
       X             COUNTRY          CTYNAME                YEAR
 Min.   :   1    Min.   :  1.00   Length:3721         Min.   :1954
 1st Qu.: 931    1st Qu.: 39.00   Class :character    1st Qu.:1967
 Median :1861    Median : 71.00   Mode  :character    Median :1976
 Mean   :1861    Mean   : 70.42                       Mean   :1975
 3rd Qu.:2791    3rd Qu.:103.00                       3rd Qu.:1983
 Max.   :3721    Max.   :135.00                       Max.   :1990
      GDPW            OIL              REG              EDT
 Min.   :  509   Min.   :0.0000   Min.   :0.0000   Length:3721
 1st Qu.: 2566   1st Qu.:0.0000   1st Qu.:0.0000   Class :character
 Median : 6425   Median :0.0000   Median :0.0000   Mode  :character
 Mean   : 9276   Mean   :0.1005   Mean   :0.4015
 3rd Qu.:13470   3rd Qu.:0.0000   3rd Qu.:1.0000
 Max.   :37903   Max.   :1.0000   Max.   :1.0000
    GDPWlag         GDPWdiff        GDPWdifflag        GDPWdifflag2
 Min.   :  509   Min.   :-9257   Min.   :-9257.0   Min.   :-9257.0
 1st Qu.: 2533   1st Qu.:  -24   1st Qu.:  -20.0   1st Qu.:  -19.0
 Median : 6245   Median :  111   Median :  117.0   Median :  116.0
 Mean   : 9090   Mean   :  186   Mean   :  189.7   Mean   :  189.9
 3rd Qu.:13167   3rd Qu.:  415   3rd Qu.:  415.0   3rd Qu.:  405.0
 Max.   :37089   Max.   : 7867   Max.   : 7867.0   Max.   : 7867.0

> data$OIL <- as.factor(data$OIL)
> data$REG <- as.factor(data$REG)
> data$YEAR <- as.factor(data$YEAR)
> data$GDPWdiff_Categories <- as.factor(data$GDPWdiff_Categories)
> #Visualise
> ggplot(data, aes(GDPWdiff_Categories, REG)) +
+   geom_boxplot() +
+   geom_jitter(alpha = 0.3) +
+   scale_x_discrete(labels=function(x){sub("\\s", "\n", x)}) +
+   theme(axis.text.x = element_text(angle = 45)) +
+   facet_grid(GDPWdiff_Categories ~ YEAR)
```

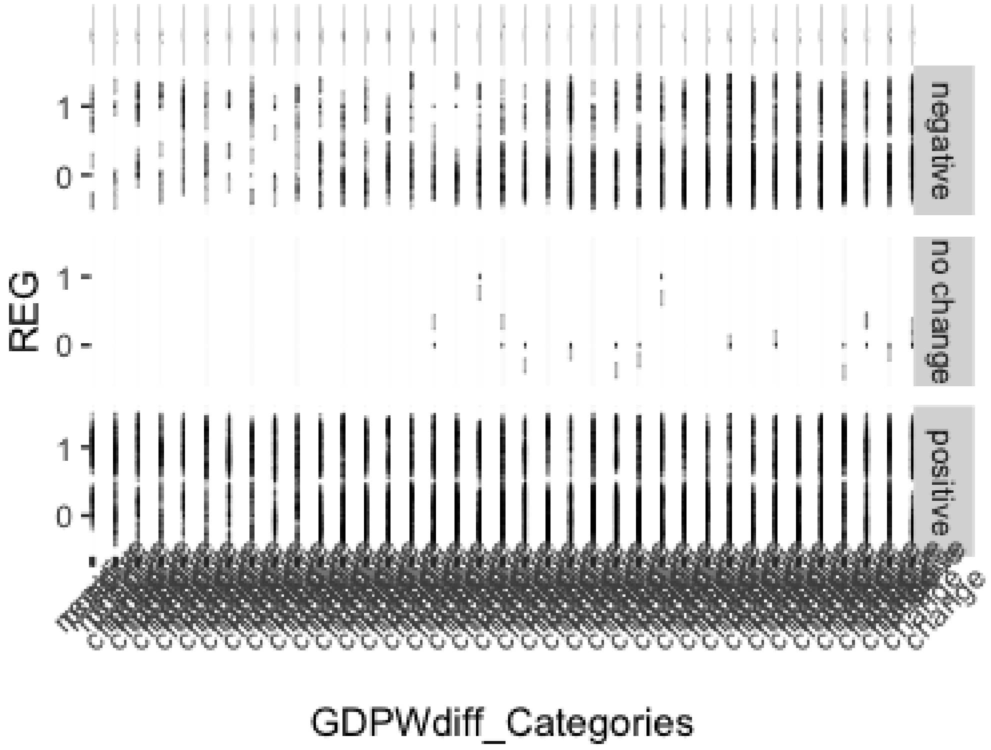

Figure 1: Marginality and PAN Governor in 2006

```
> # 1. Build the unordered multinomial logit model
> umnl_model <- multinom(GDPWdiff_Categories ~ REG + OIL, data = data)
# weights:  12 (6 variable)
initial  value 4087.936326
iter  10 value 2339.387349
```

```
final  value 2339.363928
converged
> # Print the coefficients and estimated cutoff points for the unordered multinomial logit model
> umnl_coef <- coefficients(umnl_model)
> umnl_summary <- summary(umnl_model)$coefficients
> cat("\nThe estimated coefficient values:\n")

The estimated coefficient values:
> print(umnl_coef)
          (Intercept)      REG1       OIL1
no change  -3.8011902 -1.351703 -7.9240683
positive    0.7284081  0.389905 -0.2076511
> cat("\nThe estimated cutoff points:\n")

The estimated cutoff points:
> print(umnl_summary)
          (Intercept)      REG1       OIL1
no change  -3.8011902 -1.351703 -7.9240683
positive    0.7284081  0.389905 -0.2076511
> # 2. Build the ordered multinomial logit model
> omnl_model <- polr(GDPWdiff_Categories ~ REG + OIL, data = data, method = 'logistic')
> # Print the coefficients and estimated cutoff points for the ordered multinomial logit model
> omnl_coef <- coef(omnl_model)
> omnl_summary <- summary(omnl_model)$coef

Re-fitting to get Hessian

> cat("\nThe estimated coefficient values:\n")

The estimated coefficient values:
> print(omnl_coef)
      REG1       OIL1
 0.3984834 -0.1987177
> cat("\nThe estimated cutoff points:\n")

The estimated cutoff points:
> print(omnl_summary)
                         Value Std. Error    t value
REG1                 0.3984834 0.07518467   5.300062
OIL1                -0.1987177 0.11571696  -1.717274
negative|no change -0.7311784 0.04760373 -15.359688
no change|positive -0.7104851 0.04750677 -14.955450

#For the ordered multinomial logit, the coefficients show the impact of
# each explanatory variable on the change of GDPWdiff categories from no
# change to positive, and from no change to negative. For example, for the
# Democracy variable, an increase in one unit increases the log odds of moving
# from no change to positive categories by 0.555, and increases the odds ratio
# of moving from no change to positive categories by exp(0.555) = 1.743 times,
# holding other variables constant. The cutoff points estimate the threshold
# values where the probability of moving from one category to another changes.
```

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.

```
> #Load the dataset
> mexico_muni_data <- read_csv("MexicoMuniData.csv")
Rows: 2407 Columns: 6
      Column specification


Delimiter: ","
dbl (6): MunicipCode, pan.vote.09, marginality.06, PAN.governor.06, PAN.
    vis...

    Use `spec()` to retrieve the full column specification for this data.
    Specify the column types or set `show_col_types = FALSE` to quiet
   this message.
> # Show the first few rows of the dataset
> head(mexico_muni_data)
# A tibble: 6    6
  MunicipCode pan.vote.09 marginality.06 PAN.governor.06 PAN.visits.06
   compet
        <dbl>       <dbl>          <dbl>           <dbl>         <dbl>
    <dbl>
1        1001       0.283          -1.83               0             5
          1
2        1002       0.352          -0.62               0             0
          1
3        1003       0.359          -0.875              0             0
          1
4        1004       0.238          -0.747              0             0
          1
```

```
5         1005        0.378          -1.23                   0               0
           1
6         1006        0.145          -1.31                   0               0
           1
#      with abbreviated variable name    competitive.district
> # Show the column names of the dataset
> colnames(mexico_muni_data)
[1] "MunicipCode"          "pan.vote.09"          "marginality.06"
[4] "PAN.governor.06"      "PAN.visits.06"        "competitive.district"
> # Display the summary of the dataset
> summary(mexico_muni_data)
  MunicipCode      pan.vote.09        marginality.06       PAN.governor.06
 Min.   : 1001    Min.    : 0.0050    Min.    :-2.270000    Min.    :0.0000
 1st Qu.:14108    1st Qu.: 0.1350    1st Qu.:-0.746000    1st Qu.:0.0000
 Median :20246    Median : 0.2370    Median :-0.051000    Median :0.0000
 Mean   :19505    Mean   : 0.2718    Mean   :-0.001373    Mean   :0.2152
 3rd Qu.:24040    3rd Qu.: 0.3600    3rd Qu.: 0.628500    3rd Qu.:0.0000
 Max.   :32057    Max.   :17.0000    Max.    : 3.355000    Max.   :1.0000
 PAN.visits.06      competitive.district
 Min.    : 0.00000   Min.    :0.0000
 1st Qu.: 0.00000   1st Qu.:1.0000
 Median : 0.00000   Median :1.0000
 Mean    : 0.09182   Mean    :0.8214
 3rd Qu.: 0.00000   3rd Qu.:1.0000
 Max.    :35.00000   Max.    :1.0000
> ggplot(mexico_muni_data, aes(x = marginality.06, y = pan.vote.09)) +
+   geom_point() +
+   labs(title = "Marginality and PAN Vote in 2009")
> ggplot(mexico_muni_data, aes(x = PAN.visits.06, y = pan.vote.09)) +
+   geom_point() +
+   labs(title = "PAN visitsin 2006 and PAN Vote in 2009")
> ggplot(mexico_muni_data, aes(x = marginality.06, y = PAN.governor.06))
  +
+   geom_point() +
+   labs(title = "Marginality and PAN Governor in 2006")
> ggplot(mexico_muni_data, aes(x = PAN.visits.06, y = pan.vote.09)) +
+   geom_point() +
+   labs(title = "PAN visits in 2006 and PAN Vote in 2009")
```

```

> model <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
  PAN.governor.06, data = mexico_muni_data, family = "poisson")
> summary(model)

Call:
glm(formula = PAN.visits.06 ~ competitive.district + marginality.06 +
    PAN.governor.06, family = "poisson", data = mexico_muni_data)

Deviance Residuals:
    Min        1Q    Median        3Q      Max
-2.2309   -0.3748   -0.1804   -0.0804   15.2669

```

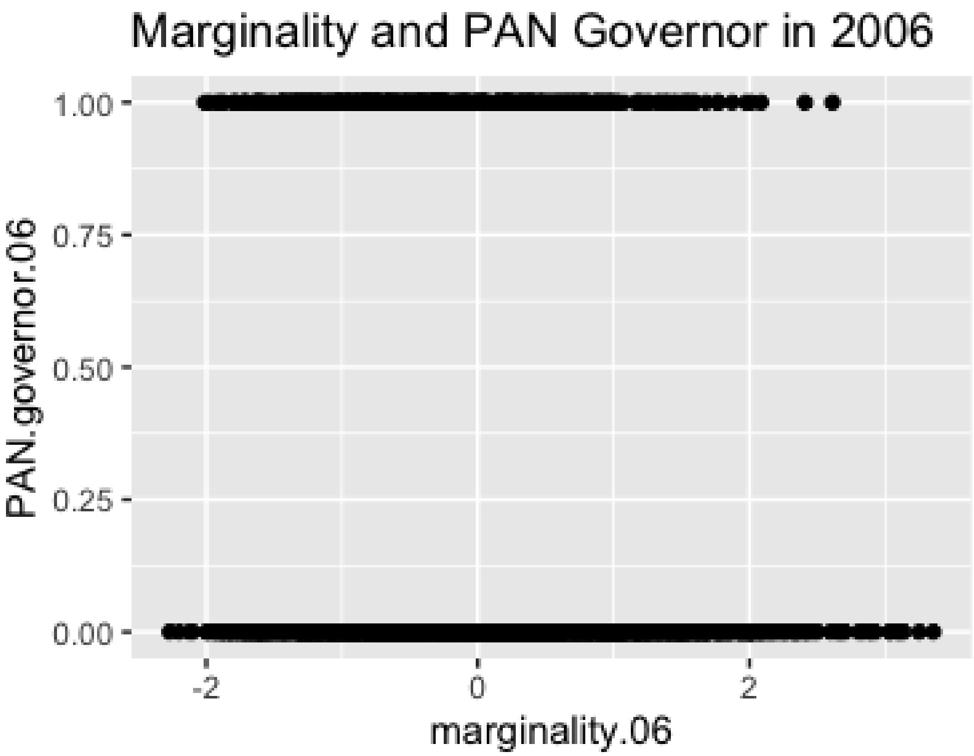

Figure 2: Marginality and PAN Vote in 2009

```
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

Number of Fisher Scoring iterations: 7

> p_value_comp_district <- summary(model)$coefficients[2,4]
> p_value_comp_district
[1] 0.6336394
```

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

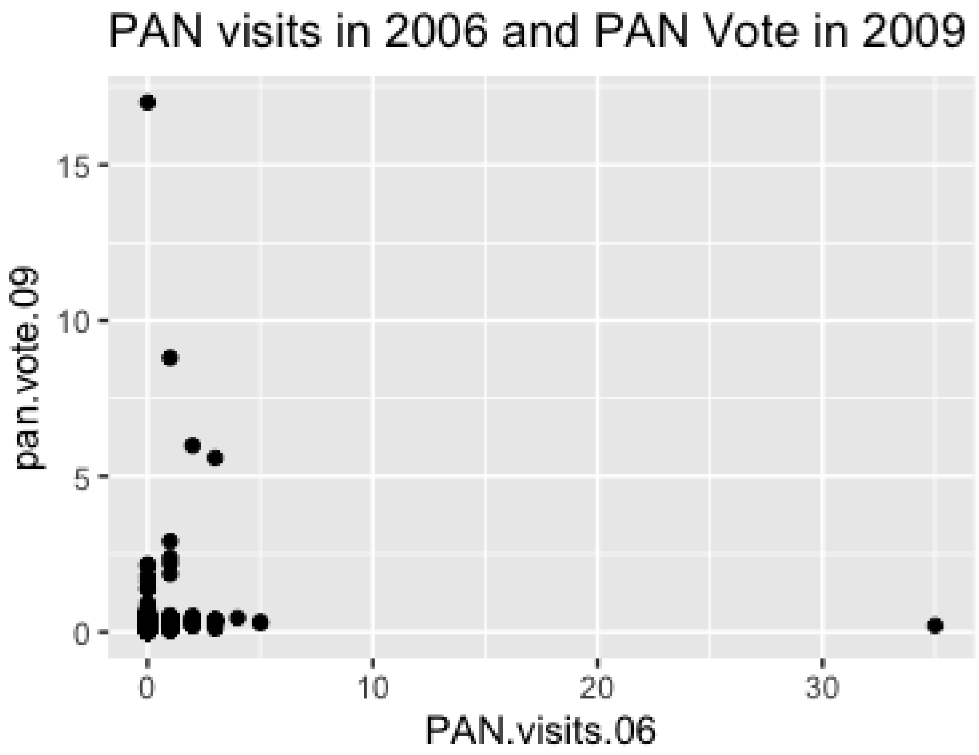

Figure 3: PAN visits in 2006 and PAN Vote in 2009

```

> coef_marginality <- summary(model)$coefficients[3,1]
> coef_marginality
[1] -2.080144
> coef_PAN_governor <- summary(model)$coefficients[4,1]
> coef_PAN_governor
[1] -0.3115789
```

(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).

```
> new_data <- data.frame(competitive.district = 1, marginality.06 = 0, PAN.governor.06 = 1)
> predicted_mean_visits <- predict(model, newdata = new_data, type = "response")
> predicted_mean_visits
         1
0.01494818
```

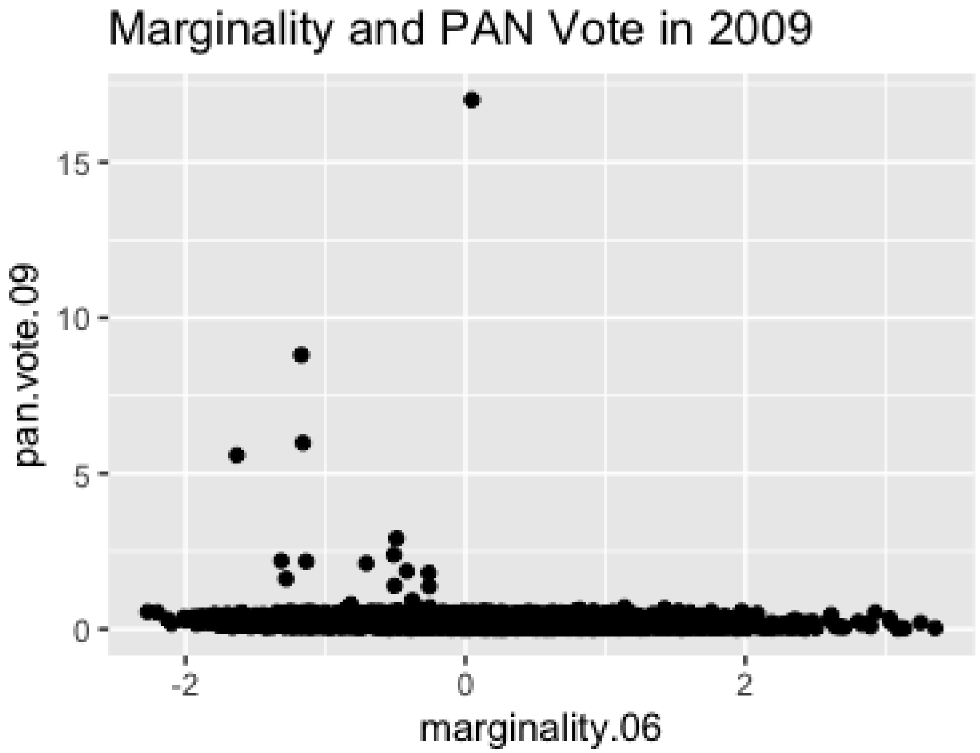

Figure 4: Marginality and PAN Governor in 2006