ANALYSIS

.....

28/03/2022

Loading the datasets for the two problems and checking the structure

```
odt1=read.csv('odt1.csv',header = T)

odt1 %>% head(10)
```

```
##    MunicipCode pan.vote.09 marginality.06 PAN.governor.06 PAN.visits.06
## 1         1001       0.283         -1.831               0             5
## 2         1002       0.352         -0.620               0             0
## 3         1003       0.359         -0.875               0             0
## 4         1004       0.238         -0.747               0             0
## 5         1005       0.378         -1.234               0             0
## 6         1006       0.145         -1.306               0             0
## 7         1007       0.263         -1.136               0             0
## 8         1008       0.314         -0.810               0             0
## 9         1009       0.306         -0.722               0             0
## 10        1010       0.308         -0.641               0             0
##    competitive.district
## 1                     1
## 2                     1
## 3                     1
## 4                     1
## 5                     1
## 6                     1
## 7                     1
## 8                     1
## 9                     1
## 10                    1
```

```
str(odt1)
```

```
## 'data.frame':    2407 obs. of  6 variables:
##  $ MunicipCode         : int  1001 1002 1003 1004 1005 1006 1007 1008 1009
1010 ...
##  $ pan.vote.09         : num  0.283 0.352 0.359 0.238 0.378 0.145 0.263
0.314 0.306 0.308 ...
##  $ marginality.06      : num  -1.831 -0.62 -0.875 -0.747 -1.234 ...
##  $ PAN.governor.06     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PAN.visits.06       : int  5 0 0 0 0 0 0 0 0 0 ...
##  $ competitive.district: int  1 1 1 1 1 1 1 1 1 1 ...
```

```
odt2=read.csv('odt2.csv',header = T)

odt2 %>% head(10)

##    Column1 COUNTRY CTYNAME YEAR  GDPW OIL REG   EDT GDPWlag GDPWdiff
## 1        1       1 Algeria 1965  6620   1   0  1.45    6502      118
## 2        2       1 Algeria 1966  6612   1   0  1.56    6620       -8
## 3        3       1 Algeria 1967  6982   1   0 1.675    6612      370
## 4        4       1 Algeria 1968  7848   1   0 1.805    6982      866
## 5        5       1 Algeria 1969  8378   1   0  1.95    7848      530
## 6        6       1 Algeria 1970  8536   1   0   2.1    8378      158
## 7        7       1 Algeria 1971  7816   1   0 2.225    8536     -720
## 8        8       1 Algeria 1972  9372   1   0 2.355    7816     1556
## 9        9       1 Algeria 1973  9361   1   0 2.495    9372      -11
## 10      10       1 Algeria 1974 10480   1   0  2.65    9361     1119
##    GDPWdifflag GDPWdifflag2
## 1          419         1071
## 2          118          419
## 3           -8          118
## 4          370           -8
## 5          866          370
## 6          530          866
## 7          158          530
## 8         -720          158
## 9         1556         -720
## 10         -11         1556

str(odt2)

## 'data.frame':    3721 obs. of  12 variables:
##  $ Column1     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ COUNTRY     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CTYNAME     : chr  "Algeria" "Algeria" "Algeria" "Algeria" ...
##  $ YEAR        : int  1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
...
##  $ GDPW        : int  6620 6612 6982 7848 8378 8536 7816 9372 9361 10480
...
##  $ OIL         : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ REG         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ EDT         : chr  "1.45" "1.56" "1.675" "1.805" ...
##  $ GDPWlag     : int  6502 6620 6612 6982 7848 8378 8536 7816 9372 9361
...
##  $ GDPWdiff    : int  118 -8 370 866 530 158 -720 1556 -11 1119 ...
##  $ GDPWdifflag : int  419 118 -8 370 866 530 158 -720 1556 -11 ...
##  $ GDPWdifflag2: int  1071 419 118 -8 370 866 530 158 -720 1556 ...
```

QUESTION ONE

PROBLEM 1.1

```
odt2$OIL=as.factor(odt2$OIL)

odt2$REG=as.factor(odt2$REG)

str(odt2)
```

```
## 'data.frame':    3721 obs. of  12 variables:
##  $ Column1     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ COUNTRY     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CTYNAME     : chr  "Algeria" "Algeria" "Algeria" "Algeria" ...
##  $ YEAR        : int  1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
...
##  $ GDPW        : int  6620 6612 6982 7848 8378 8536 7816 9372 9361 10480
...
##  $ OIL         : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ REG         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ EDT         : chr  "1.45" "1.56" "1.675" "1.805" ...
##  $ GDPWlag     : int  6502 6620 6612 6982 7848 8378 8536 7816 9372 9361
...
##  $ GDPWdiff    : int  118 -8 370 866 530 158 -720 1556 -11 1119 ...
##  $ GDPWdifflag : int  419 118 -8 370 866 530 158 -720 1556 -11 ...
##  $ GDPWdifflag2: int  1071 419 118 -8 370 866 530 158 -720 1556 ...
```

```
# Recode GDPWdiff correctly
odt2 %>% mutate(GDPWdiff = case_when(
  GDPWdiff > 1 ~ 'positive',
  GDPWdiff < -1 ~ 'negative',
  GDPWdiff == 0 ~ 'no_change',
  GDPWdiff == " " ~ 'no_change'
)) -> odt22

odt22 %>% str()
```

```
## 'data.frame':    3721 obs. of  12 variables:
##  $ Column1     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ COUNTRY     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CTYNAME     : chr  "Algeria" "Algeria" "Algeria" "Algeria" ...
##  $ YEAR        : int  1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
...
##  $ GDPW        : int  6620 6612 6982 7848 8378 8536 7816 9372 9361 10480
...
##  $ OIL         : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ REG         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ EDT         : chr  "1.45" "1.56" "1.675" "1.805" ...
##  $ GDPWlag     : int  6502 6620 6612 6982 7848 8378 8536 7816 9372 9361
...
##  $ GDPWdiff    : chr  "positive" "negative" "positive" "positive" ...
```

```
##  $ GDPWdifflag : int  419 118 -8 370 866 530 158 -720 1556 -11 ...
##  $ GDPWdifflag2: int  1071 419 118 -8 370 866 530 158 -720 1556 ...
```

PROBLEM 1.2

Unordered multinomial

```
# Use "no change" as the reference category
#odt22$GDPWdiff <- relevel(odt22$GDPWdiff, ref = "no_change")

model <- multinom(GDPWdiff ~ OIL+REG,data = odt22 )

## # weights:  12 (6 variable)
## initial  value 4071.457142
## iter  10 value 2329.003476
## final  value 2328.948230
## converged

summary(model)

## Call:
## multinom(formula = GDPWdiff ~ OIL + REG, data = odt22)
##
## Coefficients:
##           (Intercept)       OIL1      REG1
## no_change  -3.7948065 -7.8475606 -1.354990
## positive    0.7308643 -0.2112215  0.388635
##
## Std. Errors:
##           (Intercept)       OIL1       REG1
## no_change  0.27015877 31.7861048 0.75823982
## positive   0.04802656  0.1162305 0.07566136
##
## Residual Deviance: 4657.896
## AIC: 4669.896
```

Interpretation

One unit change in gross domestic difference,the odds of resulting to negative deviation in reference to no change,that is gross domestic remains constant is 0.73 times higher while there is 3.79 possiblity odds that changes will be lower or remain the same.

ordered multinom

```
odt22$GDPWdiff=as.factor(odt22$GDPWdiff)

m <- polr(GDPWdiff ~ OIL+REG,data = odt22)
summary(m)

##
## Re-fitting to get Hessian

## Call:
## polr(formula = GDPWdiff ~ OIL + REG, data = odt22)
##
## Coefficients:
##        Value Std. Error t value
## OIL1 -0.2023    0.11614  -1.741
## REG1  0.3973    0.07532   5.275
##
## Intercepts:
##                     Value    Std. Error t value
## negative|no_change  -0.7336   0.0477   -15.3696
## no_change|positive  -0.7128   0.0476   -14.9647
##
## Residual Deviance: 4666.869
## AIC: 4674.869
## (15 observations deleted due to missingness)

odt22$GDPWdiff=as.factor(odt22$GDPWdiff)
```

Interpretation

One unit change in gross domestic difference, the odds that is gross domestic doesn't change is 0.7128 times lower while there is 0.7336 possibility odds that changes will be lower or there will be no change.

QUESTION TWO

```
str(odt1)

## 'data.frame':    2407 obs. of  6 variables:
##  $ MunicipCode         : int  1001 1002 1003 1004 1005 1006 1007 1008 1009
1010 ...
##  $ pan.vote.09         : num  0.283 0.352 0.359 0.238 0.378 0.145 0.263
0.314 0.306 0.308 ...
##  $ marginality.06      : num  -1.831 -0.62 -0.875 -0.747 -1.234 ...
##  $ PAN.governor.06     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PAN.visits.06       : int  5 0 0 0 0 0 0 0 0 0 ...
##  $ competitive.district: int  1 1 1 1 1 1 1 1 1 1 ...

odt1 %>% head()
```

```
##   MunicipCode pan.vote.09 marginality.06 PAN.governor.06 PAN.visits.06
## 1        1001       0.283         -1.831               0             5
## 2        1002       0.352         -0.620               0             0
## 3        1003       0.359         -0.875               0             0
## 4        1004       0.238         -0.747               0             0
## 5        1005       0.378         -1.234               0             0
## 6        1006       0.145         -1.306               0             0
##   competitive.district
## 1                    1
## 2                    1
## 3                    1
## 4                    1
## 5                    1
## 6                    1
```

Problem 2.a

```
odt1$competitive.district=as.factor(odt1$competitive.district)

odt1$PAN.governor.06=as.factor(odt1$PAN.governor.06)


odt1 %>% str()

## 'data.frame':    2407 obs. of  6 variables:
##  $ MunicipCode         : int  1001 1002 1003 1004 1005 1006 1007 1008 1009
1010 ...
##  $ pan.vote.09         : num  0.283 0.352 0.359 0.238 0.378 0.145 0.263
0.314 0.306 0.308 ...
##  $ marginality.06      : num  -1.831 -0.62 -0.875 -0.747 -1.234 ...
##  $ PAN.governor.06     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1
...
##  $ PAN.visits.06       : int  5 0 0 0 0 0 0 0 0 0 ...
##  $ competitive.district: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2
...

output <-glm(formula = PAN.visits.06 ~ competitive.district + marginality.06
+ PAN.governor.06,
             data = odt1, family = poisson)

print(summary(output))

##
## Call:
## glm(formula = PAN.visits.06 ~ competitive.district + marginality.06 +
##     PAN.governor.06, family = poisson, data = odt1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2309  -0.3748  -0.1804  -0.0804  15.2669
##
```

```
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -3.81023    0.22209 -17.156   <2e-16 ***
## competitive.district1 -0.08135    0.17069  -0.477   0.6336
## marginality.06        -2.08014    0.11734 -17.728   <2e-16 ***
## PAN.governor.061      -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

output

##
## Call:  glm(formula = PAN.visits.06 ~ competitive.district + marginality.06
+
##     PAN.governor.06, family = poisson, data = odt1)
##
## Coefficients:
##           (Intercept)  competitive.district1         marginality.06
##              -3.81023               -0.08135                -2.08014
##      PAN.governor.061
##              -0.31158
##
## Degrees of Freedom: 2406 Total (i.e. Null);  2403 Residual
## Null Deviance:       1474
## Residual Deviance: 991.3     AIC: 1299
```

Problem 2.b

A unit change in competive districts would effect rates of visit by 0.08 times lower while failure to visit would influence the governorship re election by 0.31.

Problem 2.c

The general equation is:

panelectionvisits=-3.81023-0.08135competitive_district-2.08014marginality06-0.31158pangovernor061

Probability of being elected is; -3.81023-0.08135(1)-2.08014(0)-0.31158(1) =-4.20316

This implies for a candidate who has visited the given points only,there is odds of 4.20 times less of being selected