# Problem Set 3
**Philip Kruger**
**18328699**
**26/03/23**

## Question 1
**part 1**

The data set was loaded, the GdpWdiff column was binned into 5 categories and the unordered multinomial logit was performed with the following code:

```
gdp <- read.csv(file = "gdpChange.csv") #importing data

gdp[, 'GDPWdiff'] <- as.numeric(gdp[, 'GDPWdiff']) #setting column to nemeric
gdp$GDPWdiff2 <- "NA" #creating new column to classify the growth
print(gdp[,13])

#this for loop enters the classification of the growth into the new column
for( i in 1:nrow(gdp)){
if(gdp[i,10] < -500){ print(i)
gdp[i,13] <- "Strong Negative"}
else if(gdp[i,10] > -500 & gdp[i,10] <= -100 ){ gdp[i,13] <- "Negative"}
else if(gdp[i,10] > -100 & gdp[i,10] <= 100 ){ gdp[i,13] <- "Neutral"}
else if(gdp[i,10] > 100 & gdp[i,10] <= 300 ){ gdp[i,13] <- "Positive"}
else if(gdp[i,10] > 300){ gdp[i,13] <- "Strong Positive"}
}

#converting all the necessary columns to factors
gdp[, 'GDPWdiff2'] <- as.factor(gdp[, 'GDPWdiff2'])
gdp[, 'OIL'] <- as.factor(gdp[, 'OIL'])
gdp[, 'REG'] <- as.factor(gdp[, 'REG'])

#Running the multinomial regreassion
multimom_model1 <- multinom(GDPWdiff2 ~ OIL + REG, data = gdp, family = multinomial)
summary(multimom_model1)
```

This output the following coefficients:

```
Coefficients:
(Intercept)       OIL1      REG1
2Negative        -1.3899254 0.5875391 0.9132869
3Strong Negative -2.6183659 1.8939917 1.4514208
4Positive        -0.7194954 0.2731629 0.6437965
5Strong Positive -0.7876322 0.8347651 1.7625156

Std. Errors:
(Intercept)       OIL1      REG1
2Negative         0.07560571 0.1967929 0.12314178
3Strong Negative  0.12373863 0.2023064 0.16557257
4Positive         0.05929290 0.1762443 0.10588540
5Strong Positive  0.06014052 0.1517588 0.09284262

Residual Deviance: 10279.67
AIC: 10303.67
```

The estimated cutoff points were -500, -100, 100 and 300. These were chosen since they split the sample roughly into 5 equal sections.

In a country if they have oil, all else being equal, on average, there is an increase in the log odds ratio (over the baseline odds ratio) -1.3899254 that the growth the next year will be negative, -2.6183659 that the growth will be very negative, -0.7194954 that it will be positive and -0.7876322 that it will be very positive. This in results in a change in the basline odds ratio of 0.249 times for negative, 0.073 times for negative, 0.487 times for positive and 0.455 times for strongly positive. This means that all esle being equal on average having oil decreases the odds of growth or shrinking of the economy, with growwth being more probable than shrinking.

In a country if they are a democracy, all else being equal, on average, there is an increase in the log odds ratio (over the baseline odds ratio) 0.5875391 that the growth the next year will be negative, 1.8939917 that the growth will be very negative, 0.2731629 that it will be positive and 0.8347651 that it will be very positive. This in results in a change in the basline odds ratio of 1.8 times for negative, 6.646 times for negative, 1.19 times for positive and 1.164 times for strongly positive. This means that all elee being equal on average being democratic increases the odds of growth or shrinking of the economy, with shrinking being more probable than growth.

### part 2
The ordered multinomial logit was gotten with the following code:

```
running the polr regression
npol_model1 <- polr(GDPWdiff2 ~ OIL + REG, data = gdp)
summary(pol_model1)
```

It is worth noting that these categories (Strongly Negative, etc.) have been reordered such that they go from smallest to biggest before running this code. which gives output:

```
Coefficients:
Value Std. Error t value
OIL1 -0.05225    0.10649 -0.4907
REG1  0.84888    0.06365 13.3369

Intercepts:
Value     Std. Error t value
1Strong Negative|2Negative  -2.6372    0.0765    -34.4516
2Negative|3Neutral          -1.3766    0.0491    -28.0452
3Neutral|4Positive           0.2354    0.0413      5.7029
4Positive|5Strong Positive   1.0854    0.0458     23.6916

Residual Deviance: 10579.60
AIC: 10591.60
```

The cutoffs are the same as for the previous part. In the intercepts section we can see where the latent variable cutpoints are. from strong negative to negative is -2.6372, form negative to neutral is -1.3766, form neutral to positive is 0.2354 and from positive to strong positive is 1.0854.

The coefficient for Oil is -0.05225, this means that if a country has more than 50 percent of it's exports as oil, all else being equal, on average the log odds that the economy grows relative to the baseline odds ratio is -0.05225. If the country is democratic the coefficient is 0.84888.zz

**2 a)**

The Poisson regression is carried out with the following code:

```
mex <- read.csv(file = "MexicoMuniData.csv") #importing data

#performing the poisson regression
pos_reg <- glm(PAN.visits.06 ~ competitive.district +marginality.06+PAN.governor.06, data = mex, family

summary(pos_reg)
```

This gives us an output of:

```
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.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
```

The Z score for competitive.district is -0.477, on a one sided test, this corresponds to a p-value of .316681 (although the negative coefficient indicates that being a competitive district decreases the probability of a presidential visit, and the p-value the other way is smaller). Thus the result for competitive district is not statistically significant, since the p-value is greater than 0.05. There is not sufficient evidence to dismiss the null hypothesis that presidential candidates do not visit swing districts more than safe districts.

**2. b)**

The marginality.06 coefficient is -2.08014. This means that all else being equal, on average, there is a decrease in the log odds ratio over the baseline odds ratio of -2.08014. This leads to the odds of a presidential candidate visit decreasing by 0.125 times.

The PAN.governor.06 coefficient is -0.31158. This means that all else being equal, on average, there is a decrease in the log odds ratio over the baseline odds ratio of -0.31158. This leads to the odds of a presidential candidate visit decreasing by 0.732 times.

**2. c)**

()Taking all the values as statistically significant) The baseline log odds ratio is -3.81023, for competitive.district = 1 this increases by -0.08135 and for PAN.governor.06 =1 increases by -0.31158. thus:

$$e^{-3.81023-0.08135-0.31158} = 0.0149$$

This means on average the winning presidential candidate would visit this hypothetical district 0.0149 times.