Problem Set 3

Ana Sofia Jesus 19327602

24/03/2024

Question 1

1.

After importing the data, I converted the outcome variable (GDPWdiff: Difference in GDP between year t and t-1.) into a non-ordered factor, to ensure that it is appropriately formatted for analysis. Then, I have set the reference category to "no change".

```
gdp_data$gdp_category <- ifelse(gdp_data$GDPWdiff == 0, "no change",
                                ifelse(gdp_data$GDPWdiff > 0, "positive", "negative"))


gdp_data$gdp_category <- relevel(factor(gdp_data$gdp_category), ref = "no change")
```

Finally, I constructed an unordered multinomial logit:

```
model_logit_unordered <- multinom(gdp_category ~ REG + OIL, data = gdp_data)

summary(model_logit_unordered)
```

Obtaninig the following result:

```
Call:
multinom(formula = gdp_category ~ REG + OIL, data = gdp_data)

Coefficients:
         (Intercept)      REG      OIL
negative    3.805370 1.379282 4.783968
positive    4.533759 1.769007 4.576321

Std. Errors:
         (Intercept)      REG      OIL
negative   0.2706832 0.7686958 6.885366
positive   0.2692006 0.7670366 6.885097

Residual Deviance: 4678.77
AIC: 4690.77
```

These results provide insights into how the predictors (REG and OIL) influence the likelihood of transitioning from "no change" to either "negative" or "positive" categories, while considering the effects of other variables in the model.
For the "negative" category: When REG and OIL are both equal to zero, the estimated log odds of transitioning from "no change" to "negative" is approximately 3.8.
For the "positive" category: When REG and OIL are both equal to zero, the estimated log odds of transitioning from "no change" to "positive" is approximately 4.53.

1

For the "negative" category: When REG changes from 0 to 1, there is an estimated change in log odds of transitioning from "no change" to "negative" by approximately 1.38, while holding OIL constant.
For the "positive" category: When REG changes from 0 to 1, there is an estimated change in log odds of transitioning from "no change" to "positive" by approximately 1.77, while holding OIL constant.
Conversely, for the "negative" category: When OIL changes from 0 to 1, there is an estimated change in log odds of transitioning from "no change" to "negative" by approximately 4.78, holding REG constant.
For the "positive" category: When OIL changes from 0 to 1, there is an estimated change in log odds of transitioning from "no change" to "positive" by approximately 4.58, holding REG constant.

2.

First, I specified the levels in the correct order, to ensure that the ordered multinomial logit model is fitted appropriately, making "negative" the reference category.

```
gdp_data$gdp_category <- factor(gdp_data$gdp_category, levels = c("negative", "no change",
"positive"))
```

Next, I constructed the ordered logit model using the polr function.

```
model_ordered <- polr(gdp_category ~ REG + OIL, data = gdp_data, Hess = TRUE)
summary(model_ordered)
```

I have obtained the following output:

```
Call:
polr(formula = gdp_category ~ REG + OIL, data = gdp_data, Hess = TRUE)

Coefficients:
      Value Std. Error t value
REG  0.3985    0.07518   5.300
OIL -0.1987    0.11572  -1.717

Intercepts:
                  Value    Std. Error t value
negative|no change -0.7312   0.0476   -15.3597
no change|positive -0.7105   0.0475   -14.9554

Residual Deviance: 4687.689
AIC: 4695.689
```

These results provide insights into how the predictor variables (REG and OIL) influence the transitions between categories of GDPWdiff in an ordered multinomial logit model, while considering the effects of other variables in the model.
A one-unit increase in REG (moving from non-democracy to democracy) corresponds to an estimated change in the log odds of transitioning from "negative" to "no change" and from "no change" to "positive" by approximately 0.3985, holding OIL constant.
Conversely, a one-unit increase in OIL (presence of significant oil exports) coincides with an estimated change in the log odds of transitioning from "negative" to "no change" and from "no change" to "positive" by approximately -0.1987, holding REG constant.
Additionally, the estimated cutpoint for transitioning from "negative" to "no change" is approximately -0.7312. While the estimated cutpoint for transitioning from "no change" to "positive" is approximately -0.7105.
Overall, comparing both models, the unordered multinomial logit model has lower residual deviance and AIC values, provides interpretable coefficients, and seems to align better with the data's underlying structure, therefore it may be better than the ordered multinomial logit model.

Question 2

a.

After loading the dataset, I started by running the Poisson regression.

```
poisson_model <- glm(PAN.visits.06 ~ competitive.district + marginality.06 + PAN.governor.06,
                     data = mexico_elections,
                     family = poisson)

summary(poisson_model)
```

Obtaining the following output:

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

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
```

Test Statistic= 0.08135/0.17069 = 0.477
P-value: 0.6336
Since the p-value for the coefficient of competitive.district is 0.6336 (higher than 0,05), we fail to reject the null hypothesis, according to which there is no difference in the number of visits by PAN presidential candidates between swing districts and non-swing districts.
This indicates that there is no significant evidence to suggest that PAN presidential candidates visit swing districts more, as the coefficient for competitive.district is not statistically significant at a 0.05 significance level.

b.

The coefficient for marginality.06 is -2.08014.
This coefficient indicates the expected change in the log count of PAN presidential candidate visits for a one-unit increase in marginality.06, holding all other variables constant.
The negative value suggests that as poverty levels increase, the expected log count of PAN presidential candidate visits decreases. In other words, more impoverished areas tend to receive fewer visits from PAN presidential candidates.
The coefficient for PAN.governor.06 is -0.31158.
This coefficient represents the difference in the log count of PAN presidential candidate visits between municipalities in states with PAN-affiliated governors and municipalities in states without such governors, holding all other variables constant. Since the coefficient is negative, it suggests that, on average, municipalities in states with PAN-affiliated governors receive fewer visits from PAN presidential candidates compared to municipalities in states without such governors, after controlling for other factors.

c.

To estimate the mean number of visits from the winning PAN presidential candidate for a hypothetical district with the given characteristics, I have used the coefficients from the Poisson regression model. I have placed them in the formula for the linear predictor and then exponentiated the result to obtain the estimated mean count of visits.

Given:

- $competitive.district = 1$

- $marginality.06 = 0$

- $PAN.governor.06 = 1$

The linear predictor in the Poisson regression model is:

Linear predictor $= \beta_0 + \beta_1 \times competitive.district + \beta_2 \times marginality.06 + \beta_3 \times PAN.governor.06$
$= -3.81023 + (-0.08135 \times 1) + (-2.08014 \times 0) + (-0.31158 \times 1)$
$= -3.81023 - 0.08135 - 0.31158$
$= -4.20316$

I then obtained the estimated mean count of visits by exponentiating the linear predictor:

$$e^{Linear predictor} = e^{-4.20316} = 0.01494827$$

The estimated mean number of visits from the winning PAN presidential candidate for a hypothetical district with the given characteristics is approximately 0.01.