Problem Set 3

Applied Stats II

Due: March 24, 2024

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 24, 2024. 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.

```
1
2     # Reorganise factors
3
4     gdp_data$GDPWdiff_New <- factor(ifelse(gdp_data$GDPWdiff > 0, "
      positive", ifelse(gdp_data$GDPWdiff < 0, "negative", "no change")),
5     levels = c("positive", "no change", "negative"),
6     labels = c("positive", "no change", "negative"))
7
8     gdp_data$REG <- factor(gdp_data$REG, levels = c(1, 0),
9     labels = c("Democracy", "Non-Democracy"))
10
11    gdp_data$OIL <- factor(gdp_data$OIL, levels = c(1, 0), labels = c("
      Exceed 50%", "Otherwise"))
12
13    # Fit the multinomial logistic regression model on the  data
14
15    gdp_data$REG <- relevel(gdp_data$REG, ref = "Non-Democracy")
16    gdp_data$OIL <- relevel(gdp_data$OIL, ref = "Otherwise")
17    gdp_data$GDPWdiff_New <- relevel(gdp_data$GDPWdiff_New , ref = "no
      change")
18
19    multinom_model1 <- multinom(GDPWdiff_New ~ REG + OIL + COUNTRY, data
      = gdp_data)
20    summary(multinom_model1)
21    exp(coef(multinom_model1))
22
23
```

The results are as follows:

```
1
2         > summary(multinom_model1)
3         Call:
4         multinom(formula = GDPWdiff_New ~ REG + OIL + COUNTRY, data = gdp_
      data)
5
6         Coefficients:
7                     (Intercept) REGDemocracy OILExceed 50%  COUNTRY
8         positive      3.015581    0.3550310     8.339428    0.03550780
9         negative      2.900080    0.3539634     8.384195    0.02447895
10
11        Std. Errors:
12                    (Intercept) REGDemocracy OILExceed 50%  COUNTRY
13        positive    0.4045963    0.8686107     0.05903403   0.01064706
14        negative    0.4056846    0.8702220     0.05903378   0.01066312
15
```

```
16        Residual  Deviance :  4568.055
17        AIC:  4584.055
18        >  exp ( coef ( multinom_model1 ) )
19        ( Intercept )  REGDemocracy  OILExceed 50%   COUNTRY
20        positive       20.40094      1.426225        4185.695  1.036146
21        negative       18.17560      1.424703        4377.332  1.024781
22
23
```

Interpretation of coefficients:

Intercept: This is the log odds of no change in GDP, when all predictors are zero. The intercept changes for the shift from no change to positive exp(3) and no change to negative exp(2.9).

REGDemocracy: This is the log odds of GDP change going from no change to positive or no change to negative for a Democracy (1) or a non-Democracy (0).

OILExceed 50 per cent: This is the log odds of GDP change going from no change to positive or no change to negative for a country with fuel exports exceeding 50 per cent (1) or without (0).

Country: log odds representing affects of the country variable.

2. Construct and interpret an ordered multinomial logit with `GDPWdiff` as the outcome variable, including the estimated cutoff points and coefficients.

```
1
2       # ORDERED − PROPORTIONAL ODDS
3
4       ordered_model <− polr (GDPWdiff_New ~ REG + OIL + COUNTRY, data = gdp_
        data ,  Hess=TRUE)
5
6       # Print  the  summary  of  the  model
7       summary ( ordered_model )
8
9
```

The results are as follows:

```
1
2  Coefficients :
3                Value       Std.  Error     t value
4  REGDemocracy    0.004128     0.085548     0.04825
5  OILExceed 50%   0.077470     0.117196     0.66103
6  COUNTRY        −0.010252     0.001123    −9.13108
7
```

```
 8 Intercepts:
 9                        Value     Std. Error   t value
10 no change|positive    −6.2318     0.2640     −23.6012
11 positive|negative      0.1769     0.0770       2.2972
12
13 Residual Deviance: 4606.003
14 AIC: 4616.003
15
16
```

Interpretation: Both being a democracy and having over 50 per cent oil exports, appear to have a moderate positive affect on GDP change.

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.

```
 1
 2  # load data
 3  mexico_elections <- read.csv("https://raw.githubusercontent.com/ASDS-
     TCD/StatsII_Spring2024/main/datasets/MexicoMuniData.csv")
 4
 5  # Poisson regression
 6  mod.ps <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
     PAN.governor.06, data = mexico_elections, family = poisson)
 7  summary(mod.ps)
 8
 9  # interpreting outputs
10  cfs <- coef(mod.ps)
11  cfs
12
13
```

The results are as follows:

```
 1
 2
 3    Coefficients:
 4    Estimate Std. Error z value Pr(>|z|)
 5    (Intercept)            -3.81023      0.22209 -17.156    <2e-16 ***
 6    competitive.district   -0.08135      0.17069  -0.477    0.6336
 7    marginality.06         -2.08014      0.11734 -17.728    <2e-16 ***
 8    PAN.governor.06        -0.31158      0.16673  -1.869    0.0617 .
 9    ---
10    (Dispersion parameter for poisson family taken to be 1)
11
12    Null deviance: 1473.87  on 2406  degrees of freedom
13    Residual deviance:  991.25  on 2403  degrees of freedom
14    AIC: 1299.2
15
16    Number of Fisher Scoring iterations: 7
17
18    >
19    > # interpreting outputs
20    > cfs <- coef(mod.ps)
21    > cfs
22    (Intercept) competitive.district         marginality.06
23    -3.81023498          -0.08135181            -2.08014361
24    PAN.governor.06
25    -0.31157887
26
27
```

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

Interpretation:

For a one unit change in the predictor `marginality.06` coefficient, the difference in the logs of expected counts for the number of times the winning PAN presidential candidate visits is expected to change by -2.08014, given the other predictor variables in the model are held constant.

For a one unit change in the predictor `PAN.governor.06` coefficient, the difference in the logs of expected counts for the number of times the winning PAN presidential candidate visits is expected to change by -0.31158, given the other predictor variables in the model are held constant.

The `marginality.06` coefficient appears to be statistically significant, whereas the `PAN.governor.06` coefficient does not appear to be so.

(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`).

$\lambda = \exp(\beta_0 + \beta_1 \times competitive.district + \beta_2 \times marginality.06 + \beta_3 \times PAN.governor.06)$

$\lambda = \exp(\beta_0 + \beta_1 \times 1 + \beta_2 \times 0 + \beta_3 \times 1)$

$\lambda = \exp(-3.81023 - 0.08135 \times 1 - 2.08014 \times 0 - 0.31158 \times 1)$

$\lambda = \exp(-3.81023 - 0.08135 - 0.31158)$

$\lambda = \exp(-4.20316)$

Estimated Mean Visits $= \exp(-4.20316)$

Estimated Mean Visits ≈ 0.015