

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 for which 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

```

1 # Load data
2 gdp_data <- read.csv("gdpChange.csv", stringsAsFactors = F)
3
4 # If necessary, correct any inconsistencies in the levels
5 gdp_data$GDPWdiff_cat <- cut(gdp_data$GDPWdiff,
6                               breaks = c(-Inf, -1, 0, Inf),
7                               labels = c("negative", "no change", "
      positive"))
8
9 # Verify the conversion
10 table(gdp_data$GDPWdiff_cat)
11
12 # Construct the unordered multinomial logit model with "no change" as the
    reference category
13 model_unordered <- multinom(GDPWdiff_cat ~ REG + OIL, data = gdp_data,
    ref = "no change")
14
15 # Display the model summary to interpret coefficients
16 summary(model_unordered)
17
18 gdp_data$GDPWdiff_ord <- factor(gdp_data$GDPWdiff_cat, levels = c("
    negative", "no change", "positive"), ordered = TRUE)
19
20 # Construct the ordered multinomial logit model
21 model_ordered <- polr(GDPWdiff_ord ~ REG + OIL, data = gdp_data)
22
23 # Display the model summary
24 summary(model_ordered)

```

```

Call:
multinom(formula = GDPWdiff_cat ~ REG + OIL, data = gdp_data,
  ref = "no change")

```

```

Coefficients:
      (Intercept)          REG          OIL
no change  -3.8011902 -1.351703 -7.9240683
positive    0.7284081  0.389905 -0.2076511

```

```

Std. Errors:
      (Intercept)          REG          OIL
no change  0.27014596 0.75825317 32.9772055
positive   0.04789662 0.07552484  0.1158094

```

```

Residual Deviance: 4678.728
AIC: 4690.728

```

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.

For the unordered multinomial logit model with `GDPWdiff_cat` as the outcome variable and "no change" as the reference category, the coefficients are interpreted relative to this reference category. Here's a breakdown:

- *Coefficients for "no change" (relative to itself): These coefficients serve as the baseline comparison for other categories and are thus not directly interpretable in the context of other predictors.*
- *Coefficients for "positive" (relative to "no change"): Intercept (0.7284081): This value suggests that, all else being equal, the log odds of observing a "positive" `GDPWdiff` versus "no change" are 0.73 higher when `REG` and `OIL` are at their reference levels (typically 0). `REG` (0.389905): For each one-unit increase in `REG`, the log odds of observing a "positive" `GDPWdiff` versus "no change" increase by 0.39. `OIL` (-0.2076511): For each one-unit increase in `OIL`, the log odds of observing a "positive" `GDPWdiff` versus "no change" decrease by 0.21.*

The model's AIC and residual deviance indicate its fit to the data, with lower values generally indicating a better fit.

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.

```
1 # Load data
2 mexico_elections <- read.csv("MexicoMuniData.csv")
3
4 # Run Poisson regression
5 model_poisson <- glm(PAN.visits.06 ~ competitive.district + marginality.06 +
6   PAN.governor.06, family = poisson(link = "log"),
```

```

7         data = mexico_elections)
8
9 # Display the model summary to examine statistical tests and p-values
10 summary(model_poisson)

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

Number of Fisher Scoring iterations: 7

1 # Create a new data frame for given conditions
2 new_data <- data.frame(competitive.district = 1, marginality.06 = 0, PAN.
    governor.06 = 1)
3
4 # Use the model to predict the average number of visits
5 predicted_visits <- predict(model_poisson, newdata = new_data, type = "
    response")
6
7 # Print the predicted number of visits
8 predicted_visits

      1
0.01494818
>

```

- (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.

The Poisson regression analysis with PAN.visits.06 as the outcome variable and competitive.district, marginality.06, and PAN.governor.06 as predictors does not provide strong evidence that PAN presidential candidates visit swing districts more often. The coefficient for competitive.district is -0.08135 with a standard error of 0.17069, resulting in a z value of -0.477 and a p-value of 0.6336. This suggests that the effect of a district being competitive on the number of visits by PAN presidential candidates is not statistically significant at conventional levels.

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

The coefficient for marginality.06 is -2.08014 with a standard error of 0.11734, leading to a z value of -17.728 and a highly significant p-value of less than 0.0001. This implies that higher levels of poverty (marginality) are associated with fewer visits from the winning PAN presidential candidates, holding other factors constant. The coefficient for PAN.governor.06 is -0.31158 with a standard error of 0.16673, resulting in a z

value of -1.869 and a p-value of 0.0617. This indicates that districts with a PAN-affiliated governor are likely to receive fewer visits from the winning PAN presidential candidates, though this result is at the borderline of conventional significance levels ($p < 0.05$).

- (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`). 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`), the estimated mean number of visits from the winning PAN presidential candidate, based on the Poisson regression model, is approximately 0.01494818 visits. This predicted value is derived using the model coefficients and the specified levels of the predictors. It represents the expected number of visits to a district with these characteristics, based on the model.