

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

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 any(gdp_data$GDPWdiff == 0)
2 gdp_data$GDP_cat <- ifelse(gdp_data$GDPWdiff > 0, "positive",
3   ifelse(gdp_data$GDPWdiff < 0, "negative", "no change"))
4 gdp_data$GDP_cat <- factor(gdp_data$GDP_cat, ordered = TRUE, levels = c
5   ("no change", "positive", "negative"))
6 gdp_model <- multinom(GDP_cat ~ REG + OIL, data = gdp_data)
7 summary(gdp_model)
8 # output
9 Coefficients:
10      (Intercept)      REG      OIL
11 positive      4.533759  1.769007  4.576321
12 negative      3.805370  1.379282  4.783968
13
14 Std. Errors:
15      (Intercept)      REG      OIL
16 positive      0.2692006  0.7670366  6.885097
17 negative      0.2706832  0.7686958  6.885366
18
19 Residual Deviance: 4678.728
  AIC: 4690.728

```

When a country is not a democracy ($REG = 0$) and the fuel export in 84-86 did not exceed 50 per cent ($OIL = 0$) the log odds of GDP moving from "no changes" to "positive" is 4.53 and the log odds of the difference in GDP moving from "no change" to "negative" is 3.80.

Being a democracy is associated with, on average increasing by 1.76 the expected log odds of GDP difference moving from "no change" to "positive", while it increases the log odds of GDP difference moving from "no change" to "negative" by 1.37, holding OIL constant.

The average ratio of fuel export exceeding 50 per cent in those years, is associated on average with increasing the log odds of GDP difference going from "no change" to "positive" by 4.57, while increasing the log odds of GDP difference going from "no change" to "negative" by 4.78, holding REG constant.

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

```

1 gdp_model2 <- polr(GDP_cat ~ REG + OIL, data = gdp_data, Hess = TRUE)
2 summary(gdp_model2)
3
4 # output

```

```

5
6 Coefficients:
7      Value Std. Error t value
8 REG -0.3566  0.07485  -4.764
9 OIL  0.2306  0.11510   2.003
10
11 Intercepts:
12      Value Std. Error t value
13 no change|positive -5.5846  0.2534  -22.0376
14 positive|negative  0.7491  0.0479   15.6475
15
16 Residual Deviance: 4692.109
17 AIC: 4700.109

```

The coefficients for democracy and fuel exports indicate the change in log odds of moving into a higher category, from "no change" to "positive" or to "negative", when the variables have an one unit increase or in this case when going from 0 to 1. So being a democracy is associated on average with a decrease of 0.35 in the log odds of GDP difference moving to the next category. Fuel export exceeding 50 per cent in 84- 86 is associated on average with an increase of 0.23 in the log odds of GDP difference moving to the next category.

The intercepts indicate the log odds of being in a category in relation to "no change". The intercept for "no change"—"positive" shows the log odds of - 5.58 of GDP difference being positive compared to having no change. The intercept for "positive"—"negative" shows the log odds of 0.74 of GDP difference being negative compared to being positive.

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 pan_poimod <- glm(PAN.visits.06 ~ competitive.district + marginality.06
  + PAN.governor.06 ,

```

```

2 data = mexico_elections , family = poisson )
3 summary(pan_poimod)
4 # output
5 Coefficients:
6             Estimate Std. Error z value Pr(>|z|)
7 (Intercept)    -3.81023    0.22209  -17.156  <2e-16 ***
8 competitive.district -0.08135    0.17069   -0.477    0.6336
9 marginality.06   -2.08014    0.11734  -17.728  <2e-16 ***
10 PAN.governor.06  -0.31158    0.16673   -1.869    0.0617 .
11
12 (Dispersion parameter for poisson family taken to be 1)
13 Null deviance: 1473.87 on 2406 degrees of freedom Residual deviance:
14 991.25 on 2403 degrees of freedom
15 AIC: 1299.2
16 Number of Fisher Scoring iterations: 7

```

Being a swing state is associated with a decrease in the expected number of candidate visits by a multiplicative factor of $e^{-0.08}$, which (calculated in a calculator) is approximately 0.92. But the p-value is 0.63, which is a bit high (as in above 0.05), suggesting that the visits variable is not related to the competitive.district variable. So there is not enough evidence to suggest that there were more visits to swing states than to safe ones, or that the difference in the visits is statistically significant.

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

A one unit increase in marginality.06, holding all other variables constant, decreases the expected number of visits by a multiplicative factor of $e^{-2.08}$, which (calculated in a calculator) is approximately 0.13.

When the state has a PAN-affiliated governor, PAN.governor.06 is 1, the expected number of visits decreases by a multiplicative factor of $e^{-0.31}$, which (calculated in a calculator) is approximately 0.73, compared to when there is none.

- (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).

```

1 cfs <- coef(pan_poimod)
2 hyp_district <- exp(cfs[1] + cfs[2]*1 + cfs[3]*0 + cfs[4]*1)
3 # output = 0.0149

```

The estimated mean number of visits is 0.0149