Emily Tyger

SODA 496: Logistic Regression Assignment

Due: April 29

The purpose of this logistic regression analysis is to replicate Howard and Roessler’s and Donno’s analysis on the continuous dependent variable democracy. Utilizing the Varieties of Democracy, WDI, DPI, and Mass Mobilization datasets, I merged these tables with my own table measuring coalition to complete the dataset for this analysis and made corrections to the binary feature for democracy.

The dependent variable for this analysis is the binary measure of democracy from the datasets I previously merged for the last assignment. This assignment reports the findings of two logistic regressions, one with the binary measure of democracy at t+1, and another at t+4. The independent variables include the binary value for coalition I scraped from Wikipedia, government vote share which are the primary independent variables of this analysis. The remaining independent variables are represented in the visualization below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comparison of Logistic Regression and Multiple Regression Results** | | | | |
|  | | | | |
|  | Dependent variable: | | | |
|  |  | | | |
|  | t | democracy\_bin | democracy\_cont | |
|  | Logistic | Logistic | Multiple Multiple | |
|  | t+1 | t+4 | t+1 | t+4 |
|  | | | | |
| coalition | 3.802\*\*\* | 1.806\* | 0.080\*\* | 0.080\*\* |
|  | (1.308) | (1.034) | (0.034) | (0.034) |
|  |  |  |  |  |
| misconduct | -1.757 | -2.134\* | -0.075\*\*\* | -0.075\*\*\* |
|  | (1.102) | (1.125) | (0.025) | (0.025) |
|  |  |  |  |  |
| regime\_openness | 0.685 | 1.923\*\* | 0.057\*\*\* | 0.057\*\*\* |
|  | (0.647) | (0.899) | (0.021) | (0.021) |
|  |  |  |  |  |
| numberprev | 0.271\*\* | 0.299\*\* | 0.003 | 0.003 |
|  | (0.134) | (0.129) | (0.003) | (0.003) |
|  |  |  |  |  |
| prior\_lib\_t | -0.583 | -0.225 | -0.002 | -0.002 |
|  | (0.637) | (0.579) | (0.017) | (0.017) |
|  |  |  |  |  |
| gdp\_perCAP | -0.0002 | -0.0003 | -0.00000 | -0.00000 |
|  | (0.0003) | (0.0003) | (0.00001) | (0.00001) |
|  |  |  |  |  |
| FDI | 0.136 | -0.308 | -0.002 | -0.002 |
|  | (0.108) | (0.193) | (0.003) | (0.003) |
|  |  |  |  |  |
| foreign\_aid | 0.000 | 0.000 | 0.000 | 0.000 |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
|  |  |  |  |  |
| gdp\_growth | 0.002 | -0.0004 | 0.00003 | 0.00003 |
|  | (0.002) | (0.001) | (0.00004) | (0.00004) |
|  |  |  |  |  |
| incumbent\_running |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| alternation\_in\_power |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| region | 0.325 | 0.246 | -0.009 | -0.009 |
|  | (0.325) | (0.299) | (0.010) | (0.010) |
|  |  |  |  |  |
| opposition\_mobilization | -0.072 | -0.174 | -0.002 | -0.002 |
|  | (0.212) | (0.195) | (0.007) | (0.007) |
|  |  |  |  |  |
| Constant | -7.047 | -9.972\* | 0.382\*\* | 0.382\*\* |
|  | (4.938) | (5.658) | (0.147) | (0.147) |
|  |  |  |  |  |
|  | | | | |
| Observations | 50 | 51 | 51 | 51 |
| R2 |  |  | 0.566 | 0.566 |
| Adjusted R2 |  |  | 0.444 | 0.444 |
| Log Likelihood | -17.952 | -19.811 |  |  |
| Akaike Inf. Crit. | 59.904 | 63.623 |  |  |
| Residual Std. Error (df = 39) |  |  | 0.097 | 0.097 |
| F Statistic (df = 11; 39) |  |  | 4.625\*\*\* | 4.625\*\*\* |
|  | | | | |
| Note: | \*p\*\*p\*\*\*p<0.01 | | | |

We see at t+1, coalition has a smaller p-value than at t+4 which indicates the resulting magnitude of coalitions is more significant on the dependent variable at t+1; therefore, we can determine that coalition at t+1 is statistically significant. Although coalition results in nearly 2 units increase in democracy at t+4, the resulting magnitude is not statistically significant because we predict a value of 0, and the p-value is too large to fit this prediction. Additionally, as coalition increases at t+1 we see three times the increase per unit in the dependent variable of democracy than at t+4.

Government vote share has a smaller impact on the dependent variable at both values of t than coalition, however it is statistically significant at both values of t where coalition was not. With a hypothesized p-value of 0, the magnitude of government vote share is less than 0.01, which confirms the hypothesis. Regime openness is the only other independent variable in which an increase resulted in over 1-unit change in the dependent variable, and this occurred only at t+4. However, at the same value of t, this result was statistically significant because the magnitude of the p-value was small enough to our hypothesized size of 0.

|  |  |
| --- | --- |
|  |  |

The top visualizations report the predicted probability of democracy on the y-axis at unique values of coalition on the x-axis, holding government vote shares at its mean in order to measure the impact coalition has on the output of the model. The results of both predicted probabilities indicate that when coalition equals 1, where coalition returned true in the Wikipedia assignment, the expected probability of democracy than coalition values of 0. For t+1 the difference between predicted probability with coalition at 0 and 1 while government vote share is held constant is 0.60 and 0.85 respectively. The results of predicted probability at 0 and 1 for t+4 are 0.25 and 0.45.

The visualizations at the bottom are predicted probability plots at t+1 and t+4 respectively, with point estimates where coalition meets the predicted probability while grouping pieces of the graph based on unique values of coalition. The graphs are nearly identical; therefore, we cannot determine anything about the model at t+1 compared to t+4 based on the features of this graph.

The log likelihood of democracy at t+1 is higher than at t+4, but only by 2; their values are -17.952 and -19.811. This indicates that the similarity of t+4 is slightly greater than t+1, but the difference is not large enough to be statistically significant because the models are very similar.

The odds ratio of each logistic regression interprets how much one feature effects the odds of another feature. The features I included in this ratio are coalition and government vote share. The odds ratio at t+1 is 0.9832604 and the odds ratio of the logistic regression at t+4 is 0.891387. Both ratios are below 1 which indicates lower odds; although t+1 is very close to 1 this value still indicates exposure to coalition does not affect the odds of government vote share.

We observe a greater impact of the independent variables on the binary measure of democracy for the logistic regression than ordinary least squares regression results. Additionally, I believe the logistic regression is preforming better than the ols model because our primary independent variables of interest are coalition and government vote shares. For my ols model, there is no statistically significant p-value for government vote share, which indicates to me that something is not correct. Regime openness is no longer statistically significant for the logistic regression as it was for the ols model which indicates to me that the logit model is performing better. The independent values of interest, coalition and government vote share, are the only statistically significant variables in the table for logistic regression; for the ordinary least squares regression, regime openness was more statistically significant than government vote share.