Political Science 406 Lab 6: Missing Data

2024-05-06

Due Date: May 10, 2024

Submit your lab here.

Problem 1: Democracy and Inequality

Return to the inequality data set from Lab 1. Repeat the following analysis using multiple imputation to handle missing data. How, if at all, do the results differ? What can we conclude?

```
ineqlm3 <- lm(Gini ~ Polity + log(GDP) + Polity:log(GDP), data=inequality)
```

Try to put together a Tobit-2 selection model instead, positing an explanation for missingness the best that you can with the available data. Describe your assumptions and explain your results.

Problem 2: Colonial Mortality, Expropriation Risk, and Economic Development

In class, we discussed the Acemoglu et al. instrumental-variables natural experiment looking at colonial mortality, institutions, and economic development. You can find the original article. Replication data are available as well.

Start out by making sure you have the data working and that you understand what's going on by replicating the analysis in column 1 of either Table 4 or Table 5 (your choice). You will likely not get exactly the published results, as the replication data are the result of a long debate on data quality that has resulted in some changes in scoring. So the most available (and credible) indicators are no longer exactly the same ones Acemoglu et al used in the 1990s. However, you should get results that are quite similar.

One aspect of the published debate involved arguments about whether the colonial mortality data were credible. The dataset you are using reports alternetive ways of measuring settler mortality. Replicate your results using one of those, replacing the original mortality indicator (logmort0) with an alternative (logmortnaval1). Do the results change? What about the pattern of missingness?

Can anything be done to address missingness here? Try a solution, report the results, and discuss how (if at all) it affects the analysis.